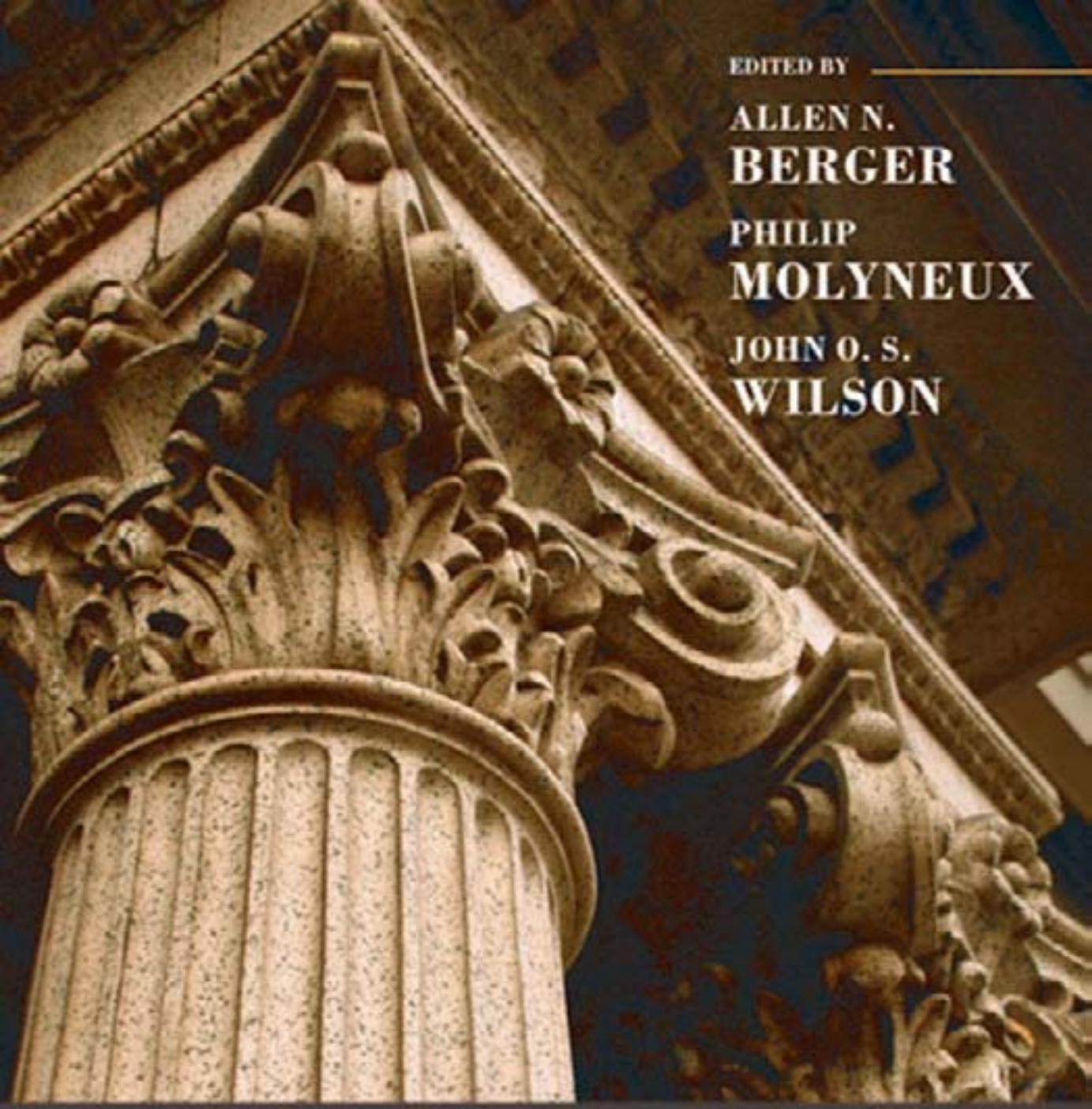

EDITED BY

ALLEN N. BERGER

PHILIP MOLYNEUX

JOHN O. S. WILSON

The Oxford Handbook *of* BANKING

THE OXFORD HANDBOOK OF

BANKING

This page intentionally left blank

THE OXFORD HANDBOOK OF

BANKING

Edited by

ALLEN N. BERGER

PHILIP MOLYNEUX

and

JOHN O. S. WILSON

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto
With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam
Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries
Published in the United States
by Oxford University Press Inc., New York

First published 2010

British Library Cataloguing in Publication Data
Data available
Library of Congress Cataloging in Publication Data
Library of Congress Control Number: 2009934149
Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
CPI Antony Rowe, Chippenham, Wiltshire
ISBN 978–0–19–923661–9
1 3 5 7 9 10 8 6 4 2

For Mindy
(Allen N. Berger)

For Delyth, Alun, Gareth, Gethin,
Catrin, Lois, and Rhiannon
(Philip Molyneux)

For Alison, Kathryn, Elizabeth, and Jean
(John O. S. Wilson)

This page intentionally left blank

Preface

When we commenced this project at the start of 2007, banks around the world were posting record profits and major risks appeared to have abated. The general macroeconomic environment embodied in rising stock markets and buoyant economic growth provided the bedrock for the strong performance we observed at that time.

Since then things have changed completely! During the first half of 2007, following increases in interest rates, the rate of US subprime mortgage delinquencies increased, prices of mortgage-backed securities were reduced, and the cost of insuring these securities against default increased. The turmoil that then hit the global financial system has been unprecedented. There have been widespread government bank bailouts, recapitalization plans, liquidity injections, and credit guarantee schemes affecting many countries. The cost of all this activity has been enormous. In April 2009, the IMF stated that US financial institutions were likely to incur $2.7 trillion of losses from the global crisis, part of a worldwide total expected to top $4 trillion. Global banking sector instability and gridlock in many other financial markets have raised profound concerns about the stability of the financial system and the business models used by banks within the system.

As we completed the handbook in early 2009, banking systems in many countries (particularly in the US, UK, Germany, Spain, Ireland, Iceland, and throughout central and eastern Europe) remained in a state of crisis, threatening capacity to perform effective intermediation functions for many years to come. Government intervention in banking continues via the purchases of impaired assets, recapitalizations of troubled banks, and injections of liquidity into the system. Academics and policymakers continue to debate proposals to fix the financial system. Reforms that have been or are likely to be adopted in the coming months include: extending the coverage of bank regulation; increasing capital requirements; designing countercyclical capital requirements; enhancing regulation and supervision of bank liquidity; enhanced supervision of credit rating agencies; codes covering executive pay and benefits; improving arrangements for regulation of the activities of cross-border banks; and a shift in focus from micro- to macro-prudential supervision. If lessons are learned from the ongoing financial crisis, there is little doubt that the banking industry which emerges in the years to come will be very different from the one we observed at the beginning of 2007.

This handbook provides the reader with a comprehensive overview and analysis of banking. The authors of the following 36 chapters comprise a collection of

leading academics and policymakers in the field. These authors emanate from universities in the United States, Europe, South America, and Asia; the US Federal Reserve System; the Office of the Comptroller of Currency; the European Central Bank; the Bank of Thailand; the World Bank; the International Monetary Fund; and the World Trade Organization. The book strikes a balance among abstract theory, empirical research, practitioner analysis, and policy-related material. Different chapters in the handbook have different emphases on these four ingredients. We hope that the contributions contained in this handbook set the stage for future research and policy debate for many years to come.

Acknowledgments

Firstly, and most importantly we wish to thank the contributors to the handbook. We are delighted to have brought together such an outstanding set of research experts from academic and policy arenas across Europe, North America, South America, and Asia. These experts have shown a high level of commitment and perseverance to the project from beginning to end. Without their expertise, dedication, and efficiency in producing scholarly banking chapters this handbook would never have been possible.

The production of this handbook has also relied heavily on the exceptional enthusiasm and commitment of Oxford University Press, most notably Sarah Caro, publisher for Economics and Finance who was crucial in helping us kick-start the project. Oxford University Press delegates and a number of anonymous referees also played an important role in advising on the shape of the handbook at its early stages. We would also like to acknowledge the help and advice of Aimee Wright and Jen Wilkinson, Assistant Commissioning Editors who worked closely with us throughout the entire production process. Harriet Ayles, Emma Hawes, Lianne Slavin, and Lucy Gostwick also played a crucial role toward the end of the project. Our proofreader, Michael Janes, also provided invaluable service toward the final stages of the project.

We would also like to acknowledge the support of our home institutions: the Moore School of Business at University of South Carolina; the Business School at Bangor University; and the Management School at University of St Andrews.

A number of individuals provided comments to us on the content and style on parts of the handbook, including Barbara Casu, Mark Flannery, Claudia Girardone, John Goddard, Dick Herring, Ed Kane, Donal McKillop, Joe Peek, and Larry White. A big thank you goes to them.

Finally, we would like to thank our families and friends for their encouragement and patience over the last three years while completing this handbook. Their support is much appreciated.

This page intentionally left blank

Contents

II. REGULATORY AND POLICY PERSPECTIVES

III. BANK PERFORMANCE

IV. MACROECONOMIC PERSPECTIVES IN BANKING

V. INTERNATIONAL DIFFERENCES IN BANKING STRUCTURES AND ENVIRONMENTS

List of Figures

List of Tables

List of Abbreviations

ABCP	asset-backed commercial paper
ABS	asset-backed securities
ACH	automated clearing house
ADB	Asian Development Bank
ADEBA	Asociación de Bancos Privados de Capital Argentino
AEI	American Enterprise Institute for Public Policy Research
AMA	advanced measurement approach
AMLF	asset-backed [commercial paper] money-market-fund liquidity facility
APR	annual percentage rate
ARM	adjustable rate mortgage
ASRF	asymptotic single risk factor
ATM	automated teller machine
BBA	British Bankers' Association
BCBS	Basel Committee on Bank Supervision
BCCI	Bank of Credit and Commerce International
BCRA	Banco Central de la República Argentina
BHC	bank holding company
BIS	Bank for International Settlements
bps	basis points
C and I	commercial and industrial
CAD	Capital Adequacy Directive
CAPM	capital asset pricing model
CAR	cumulative abnormal return
CB	central bank
CBC	commercial bank clearinghouse
CBO	collateralized bond obligation
CBOT	Chicago Board of Trade
CCF	credit conversion factor

CD	certificate of deposit
CDF	cumulative distribution function
CDO	collateralized debt obligation
CDS	credit default swap
CEE	Central Eastern Europe
CEPR	Centre for Economic Policy Research
CESifo	Center for Economic Studies Institute for Economic Research
CFS	Center for Financial Studies
CFTC	Commodity Futures Trading Commission
CGFS	Committee on the Global Financial System
CHIPS	Clearing House Interbank Payments System
CLO	collateralized loan obligation
CLS	continuous linked settlement
CLTV	combined loan to value
CME	Chicago Mercantile Exchange
CMO	collateralized mortgage obligation
CORI	Contracting and Organizations Research Institute
CPDO	constant proportion debt obligation
CPI	consumer price index
CRA	credit rating agency
CRO	credit rating organization
DOJ	Department of Justice
DVP	delivery versus payment
EAD	exposure at default
EAIEL	East Asian International Economic Law and Policy Program
EBRD	European Bank for Reconstruction and Development
ECB	European Central Bank
ECLAC	Economic Commission for Latin America and the Caribbean
EFA	Employers Forum on Age
EL	expected loss
ELA	emergency liquidity assistance
EME	emerging market economies
ESC	économies, sociétés, civilisations
ESRC	Economic and Social Research Council
ETA	electronic transfer account
ETF	Exchange-traded fund

EVT	extreme-value theory
FASB	Financial Accounting Standards Board
FASIT	financial asset securitization investment trust
FDI	foreign direct investment
FDIC	Federal Deposit Insurance Corporation
FDICIA	Federal Deposit Insurance Corporation Improvement Act (1991)
FEEM	Fondazione Eni Enrico Mattei
FHA	Federal Housing Authority
FHLB	Federal Home Loan Bank
FICO	Fair Isaac Corporation
FIRREA	Financial Institutions Reform, Recovery and Enforcement Act (1989)
FRBNY	Federal Reserve Bank of New York
FRBSF	Federal Reserve Bank of San Francisco
FSA	Financial Services Authority
FSAP	Financial Services Action Plan
FSCS	Financial Services Compensation Scheme
FSLIC	Federal Savings and Loan Insurance Corporation
FSU	former Soviet Union
FTC	Federal Trade Commission
GATS	General Agreement on Trade in Services
GDP	gross domestic product
GLBA	Gramm-Leach-Bliley Act
GNMA	Government National Mortgage Association
GSE	government-sponsored enterprise
HEL	home equity loan
HHI	Herfindahl-Hirschman Index
HMDA	Home Mortgage Disclosure Act
HMSO	Her Majesty's Stationery Office
IADB	Inter-American Development Bank
IBBEA	Interstate Banking and Branching Efficiency Act
ICBA	International Credit Brokers Alliance
ICBC	Industrial and Commercial Bank of China
IFRS	International Financial Reporting Standards
IMF	International Monetary Fund
IPO	initial public offering

IRB	internal ratings based
ISDA	International Swap Dealers Association
JPMC	JPMorgan Chase & Co.
KMV	Kealhofer, McQuown, and Vasicek
LBO	leveraged buyout
LCBO	large and complex banking organization
LCFIs	large complex financial institutions
LGD	loss given default
LIBOR	London Inter-Bank Offered Rate
LOLR	lender of last resort
LSE	London School of Economics
LTCM	Long-Term Capital Management
LVPS	large-value payment system
M&A	mergers and acquisitions
MBS	mortgage-backed securities
MFN	most favored nation
MMMF	money market mutual fund
MSA	Metropolitan Statistical Area
NASDAQ	National Association of Securities Dealers Automated Quotations
NBER	National Bureau of Economic Research
NCAER	National Council for Applied Economic Research
NEIO	New Empirical Industrial Organization
NRB	net regulatory benefit (or burden)
NRSRO	Nationally Recognized Statistical Rating Organization
NYSE	New York Stock Exchange
OCC	Office of the Comptroller of the Currency
OECD	Organization for Economic Cooperation and Development
OFHEO	Office of Federal Housing Enterprise Oversight
OIS	overnight index swap
OMO	open-market operations
OTC	over the counter
OTS	Office of Thrift Supervision
PCA	prompt corrective action
PD	probability of default
PE	private equity
PMI	Private Mortgage Insurance

PVP	payment versus payment
QSPE	qualifying special purpose entity
QSPV	qualifying special purpose vehicle
RAROC	risk-adjusted return on capital
RBCR	risk-based capital ratio
REIT	real estate investment trust
REMIC	real estate mortgage conduit
RFC	Reconstruction Finance Corporation
RIC	regulated investment company
RIETI	Research Institute of Economy, Trade, and Industry
RMBS	residential-mortgage-backed securities
ROA	return on assets
ROE	return on equity
RTGS	real-time gross settlement
RWA	risk-weighted assets
SBCS	small business credit scoring
SCF	Survey of Consumer Finance
SCP	Structure–Conduct–Performance
SEC	Securities and Exchange Commission
SEE	South Eastern Europe
SEIR	structured early intervention and resolution
SIFMA	Securities Industry and Financial Markets Association
SIPC	Securities Investor Protection Corporation
SIV	structured investment vehicle
S&L	savings and loans
SME	small and medium sized enterprise
SOCB	state-owned commercial bank
SOE	state-owned enterprise
SPD	stress probability of default
SPE	special purpose entity
SPV	special purpose vehicle
SSBF	Survey of Small Business Finance
SUERF	Société Universitaire Européenne de Recherches Financières
TAF	term auction facility
Tankan	kigyō tanki keizai kansoku chōsa

TARGET	Trans-European automated real-time gross settlement express transfer
TARP	Troubled Asset Relief Program
T-Bill	Treasury Bill
TBTF	too big to fail
TIIE	twenty-eight-day interbank equilibrium interest rate
UL	unexpected loss
VAR	value-at-risk
VIE	variable interest entity

List of Contributors

Michel Aglietta is Professor at the University of Paris-X Nanterre, Scientific Advisor for the Research Center in International Economics (CEPII), consultant at Groupama-am, and former member of the French Council of Economic Advisors attached to the Prime Minister. His most influential writings include *Régulation et crise du capitalisme*, *La Violence de la Monnaie* (with A. Orléan), and *Globalisation financière: l'aventure obligée* (with A. Brender and Virginie Coudert). Recent books include *Corporate Governance adrift* (with A Rebérioux), 2005; *Désordres dans le capitalisme mondial* (with A. Berrebi), 2007; *La crise*, 2008; and *Crise et rénovation de la finance*, 2009.

Franklin Allen is the Nippon Life Professor of Finance and has been on the faculty at Wharton since 1980. He is Co-Director of the Financial Institutions Center. He is a past President of the American Finance Association. He received his doctorate from Oxford University.

Linda Allen is the Presidential Professor of Finance at the Zicklin School of Business at Baruch College, City University of New York, and Adjunct Professor of Finance at the Stern School of Business at New York University. She has been a member of the Standard & Poor's Academic Council since its formation in 2004.

Olivier de Bandt holds a PhD from the Department of Economics of the University of Chicago. He is currently Director of Business Conditions and Macroeconomic Forecasting at the Banque de France and Associate Professor at the University of Paris X.

James R. Barth is the Lowder Eminent Scholar in Finance at Auburn University and a Senior Finance Fellow at the Milken Institute. His research has focused on financial institutions and capital markets, both domestic and global, with special emphasis on regulatory issues. Most recently, he served as leader of an international team advising the People's Bank of China on banking reform.

Allen N. Berger is the H. Montague Osteen, Jr., Professor in Banking and Finance, Moore School of Business, University of South Carolina; Senior Fellow, Wharton Financial Institutions Center; Extramural Fellow, CentER, Tilburg University; and Secretary/Treasurer, Financial Intermediation Research Society. He has published more than a hundred articles, including papers in *Journal of Political Economy, American Economic Review, Journal of Monetary Economics, Journal of Finance, Journal of Financial Economics, and Review of Financial Studies.*

John P. Bonin is the Chester D. Hubbard Professor of Economics and Social Science at Wesleyan University in Connecticut. Bonin was the Editor of the *Journal of Comparative Economics* from 1996 to 2006. He has worked as a consultant for the World Bank, the United Nations, the US Treasury, the Institute for EastWest Studies, the Institute for International Economics (WIIW) in Vienna, and the 1990 Institute.

Arnoud W. A. Boot is Professor of Corporate Finance and Financial Markets at the University of Amsterdam, and research fellow at CEPR. He is a member of the Bank Council of the Dutch Central Bank (DNB) and advisor to the Riksbank (Central Bank of Sweden). His research focuses on corporate finance and financial institutions. Prior to his current positions, he was on the faculty of the J. L. Kellogg Graduate School of Management at Northwestern University.

Claudia M. Buch holds the chair for International Finance and Macroeconomics University of Tübingen. She has been Scientific Director of the Institute for Applied Economic Research since January 2005 and a member of the Council of Economic Advisers to the German Ministry of Economics and Labor since 2004. She was affiliated with the Kiel Institute for World Economics (Germany) from 1992 to 2003, where she headed the research area 'Financial Markets'. She graduated from the University of Bonn with a Master of Economics in 1991 and from the University of Wisconsin with a Master of Business Administration in 1989. From the University of Kiel, she received a PhD in Economics in 1996 and a Habilitation (German post-doc. degree) in 2002. She has been a visiting scholar at the Salomon Center for the Study of Financial Institutions (New York University), at the University of Michigan (Ann Arbor), at the National Bureau of Economic Research (Cambridge, Mass.), and at the Research Centre of the Deutsche Bundesbank (Frankfurt).

Charles W. Calomiris is the Henry Kaufman Professor of Financial Institutions at the Columbia University Graduate School of Business and a Professor at Columbia's School of International and Public Affairs. Professor Calomiris co-directs the Project on Financial Deregulation at the American Enterprise Institute, is a member of the Shadow Financial Regulatory Committee and the Financial Economists Roundtable, and is a Research Associate of the National Bureau of Economic Research. He received a BA in economics from Yale University in 1979 and a PhD in Economics from Stanford University in 1985.

Gerard Caprio is Professor of Economics and Chair of the Center for Development Economics, Williams College; previously: Director for Financial Sector Policy and Manager, Financial Sector Research, World Bank; Vice President, JPMorgan; economist at the Federal Reserve and IMF; Adjunct Professor, George Washington University. Jerry has researched and written extensively on financial sector issues and banking crises in developing countries, including *Rethinking Bank Regulation: Till Angels Govern* (Cambridge, 2006), with James Barth and Ross Levine. He is a co-editor of the *Journal of Financial Stability.*

Elena Carletti is a Professor at the European University Institute. Her research interests are in the areas of banking, financial stability, corporate governance industrial organization, and competition policy. She holds a PhD from the London School of Economics and a Habilitation from the University of Mannheim.

Fernando J. Cardim de Carvalho is Professor of Economics at the Institute of Economics, Federal University of Rio de Janeiro (Brazil) and Director of the Money and Finance Study Group headquartered at the same university. A former chairman of the Brazilian Association of Graduate Schools in Economics (ANPEC), his research interests are Keynesian macroeconomics, international monetary economics, and financial systems.

Jacopo Carmassi is a researcher at Assonime, the Association of Joint Stock Companies incorporated in Italy, and a Fellow of the Wharton Financial Institutions Center. He holds a PhD in Law and Economics from the University LUISS Guido Carli of Rome and is a former visiting scholar at the Wharton Financial Institutions Center. His research and work activities have focused on banking and financial markets regulation and supervision.

Nicola Cetorelli is a Senior Economist in the Research Department of the Federal Reserve Bank of New York and a Fellow of the Wharton Financial Institutions Center. His Research work has focused on the many ways banks affect the real economy, with a specific interest more recently in topics of international banking and financial globalization. He has published in the *Journal of Finance*, the *Journal of Money, Credit and Banking, the Journal of Economic Dynamics and Control*, and various policy journals. He received his PhD in Economics from Brown University and a BA from the University of Rome, Italy.

Gayle L. DeLong is an Associate Professor in the Economics and Finance Department at Baruch College. She received her Masters in International Business Studies from the University of South Carolina in 1984 and her PhD in International Business and Finance from New York University in 1998. Her research interests include bank mergers, international finance, and the economics of health care. She received research grants from the Zicklin School of Business, TransCoop Program of the Alexander Humboldt Foundation, and the Federal Deposit Insurance Corporation's Center for Financial Research, where she is a Fellow. She received the Eugene M. Lang Fellowship for Junior Faculty as well as doctoral fellowships at New York University. Since 2007, she has served on the Research Committee of Sensible Action for Ending Mercury-induced Neurological Disorders.

Asli Demirgüç-Kunt is the Senior Research Manager in Finance and Private Sector, in the World Bank's Development Economics Research Group. After joining the Bank in 1989 as a Young Economist, she has been in different divisions of the

Research Group, working on financial sector issues. Her recent research has focused on banking crises, financial regulation, and the links between finance and economic development.

Robert DeYoung is Capitol Federal Professor in Financial Institutions and Markets at the University of Kansas. He is also an editor of the *Journal of Money, Credit, and Banking,* a research program coordinator at the Federal Deposit Insurance Corporation, and a Visiting Scholar at the Federal Reserve Bank of Kansas City.

Astrid A. Dick is Assistant Professor of Economics at INSEAD. Previously, she spent five years as an economist at the Federal Reserve Bank of New York and at the Federal Reserve Board in Washington. She received her PhD in Economics from the Massachusetts Institute of Technology in 2002.

Thomas A. Durkin is Senior Economist, Board of Governors of the Federal Reserve System, Washington, DC. He has specialized in the economics and regulation of consumer financial services in the federal government as Senior Economist at the Federal Reserve Board, in the academic area as Associate Professor of Finance at the Pennsylvania State University, and in the private sector as Chief Economist and Director of Research of the American Financial Services Association. He holds an AB degree from Georgetown University and a PhD degree from Columbia University.

Robert A. Eisenbeis is currently Chief Monetary Economist at Cumberland Advisors. He recently retired as Executive Vice President and Director of Research at the Federal Reserve Bank of Atlanta, and had previously been at the Board of Governors and FDIC. He is now a member of the US Shadow Financial Regulatory Committee.

Gregory Elliehausen holds a Research Associate Professor position with the Financial Services Research Program at The George Washington University, Washington, DC. He has held a similar position with the McDonough School of Business at Georgetown University and has been a member of the economics staff at the Federal Reserve Board. He holds BS and PhD degrees from the Pennsylvania State University.

Mark J. Flannery earned economics degrees from Princeton and Yale Universities. Since 1989, he has held an Eminent Scholar chair in the University of Florida's Department of Finance. He has consulted with various regulatory agencies and helped found the FDIC's Center for Financial Research.

W. Scott Frame is a financial economist and policy advisor at the Federal Reserve Bank of Atlanta. His research broadly relates to the effect of technological and regulatory change on retail credit markets. Prior to joining the Atlanta Fed, Scott was a senior financial economist at the US Treasury Department.

Xavier Freixas is Dean of the Undergraduate School of Economics and Business Administration, Professor at the Universitat Pompeu Fabra in Barcelona (Spain), and

Research Fellow at CEPR. He is also Chairman of the Risk-Based Regulation Program of the Global Association of Risk Professionals (GARP). He has been a consultant for the European Investment Bank, the New York Fed, the European Central Bank, the World Bank, the Interamerican Development Bank, and the European Investment Bank. He received his PhD from the University of Toulouse in 1978.

John Goddard is Professor of Financial Economics at Bangor Business School, University of Wales, Bangor. Before joining Bangor University in 2005, he was Professor of Economics at Swansea University. His research interests include the economics of banking and financial services, industrial organization, and the economics of professional soccer.

Michael B. Gordy is a Senior Economist at the Federal Reserve Board. His current research focuses on credit risk pricing, and on the design, calibration, and computation of models of portfolio risk. Michael is a recipient of Risk's 2004 Quant of the Year and GARP's 2003 Financial Risk Manager of the Year, and serves as an associate editor of *Journal of Banking & Finance*, *Journal of Credit Risk*, and *International Journal of Central Banking*. Michael received his PhD in Economics from the Massachusetts Institute of Technology in 1994.

Timothy H. Hannan is Senior Economist at the Board of Governors of the Federal Reserve System. He received his PhD in Economics from the University of Wisconsin in 1974. His area of interest is bank behavior and competition.

Philipp Hartmann is Head of the Financial Research Division at the European Central Bank, a Fellow of the Centre for Economic Policy Research, and the Vice-President of SUERF—the European Money and Finance Forum. He works in banking, financial stability, European capital markets, and international monetary affairs. His research is published in academic journals, including *Review of Economics and Statistics, Journal of Money, Credit, and Banking, Economic Policy, Journal of International Money and Finance*, and *Journal of Banking & Finance*, and several books. His policy work has been discussed in fora such as the ECOFIN Council, the ECB Governing Council, European Commission fora, and the Basel Committee on Banking Supervision and published in numerous reports.

Iftekhar Hasan is the Cary L. Wellington Professor of Finance at the Rensselaer Polytechnic Institute in Troy, New York. He is also scientific advisor of the Bank of Finland in Helsinki. He is an associate editor of *Journal of Money, Credit, and Banking*, *Journal of Banking & Finance*, and *Journal of International Money and Finance* and serves as the editor of *Journal of Financial Stability*.

Erik A. Heitfield is a senior economist at the Federal Reserve Board where he conducts research on the measurement and management of credit risk in complex financial institutions. In 2004 and 2005 Erik co-chaired the Basel Committee's Loss Given Default Working Group and in 2006 and 2007 Erik was a senior staff

economist for the President's Council of Economic Advisors. Erik received his PhD in Economics from the University of California at Berkeley in 1998.

Richard Herring is Jacob Safra Professor of International Banking and Professor of Finance at The Wharton School, University of Pennsylvania, where he is also Co-Director of The Wharton Financial Institutions Center. Outside the University he serves as Co-Chair of the US Shadow Financial Regulatory Committee. His area of special interest is financial regulation and international finance.

Patrick Honohan is Professor of International Financial Economics and Development at Trinity College Dublin and a research fellow of CEPR. Previously he was a Senior Advisor at the World Bank and his career has also included periods on the staffs of the IMF, the Central Bank of Ireland, the Economic and Social Research Institute, Dublin, and as Economic Advisor to the Taoiseach. In 2009, he was appointed Governor of the Central Bank of Ireland.

Joseph P. Hughes is Professor of Economics at Rutgers University. His research has been published in such journals as *American Economic Review*, *Journal of Banking & Finance*, *Journal of Economic Theory*, *Journal of Financial Services Research*, *Journal of Money, Credit, and Banking*, and *Review of Economics and Statistics*. He received his PhD from the University of North Carolina at Chapel Hill.

David Humphrey is F. W. Smith Eminent Scholar in Banking at Florida State University. He received his PhD in Economics from the University of California (Berkeley) and was previously at the Federal Reserve Board and Federal Reserve Bank of Richmond for sixteen years working on banking and payment system issues.

Edward J. Kane is the James F. Cleary Professor in Finance at Boston College. He is a member of the US Shadow Financial Regulatory Committee, a visiting scholar at the Federal Reserve Bank of San Francisco, and a senior fellow in the Federal Deposit Insurance Corporation's Center for Financial Research. He is a past president and fellow of the American Finance Association, a former Guggenheim fellow, and a long-time research associate of the National Bureau of Economic Research. He served for twelve years on the finance committee of Teachers Insurance. His BS is from Georgetown University and his PhD from the Massachusetts Institute for Technology.

George G. Kaufman is the John F. Smith Professor of Economics and Finance at Loyola University Chicago and consultant to the Federal Reserve Bank of Chicago. Previously he was an economist at the Federal Reserve Bank of Chicago and a Professor of Finance at the University of Oregon. Kaufman is co-editor of *Journal of Financial Stability* and a founding editor of *Journal of Financial Services Research*. He is former president of the Western Finance Association, the Midwest Finance Association, and the North American Economic and Finance Association. He serves as co-chair of the Shadow Financial Regulatory Committee. He holds a PhD in Economics from the University of Iowa.

Leora Klapper is a Senior Economist in the Finance and Private Sector Team of the Development Economics Research Group at the World Bank in Washington, DC. Her published research focuses on entrepreneurship, SME and corporate finance, corporate governance, and corporate risk management. She is the also the author of the 'World Bank Group Entrepreneurship Database'. Her operational work includes promoting the development of credit information bureaus in South Asia; factoring in Eastern Europe; and the use of technology for the delivery of financial services in India and Africa. Prior to joining the World Bank, she worked at the Board of Governors of the Federal Reserve System, the Bank of Israel, and Salomon Smith Barney. She holds a PhD in Financial Economics from the Stern School of Business at New York University.

Andreas Lehnert is Chief of the Household and Real Estate Finance section in the Division of Research and Statistics at the Federal Reserve Board in Washington, DC. He monitors and analyzes developments in the household sector, especially in areas of mortgages and housing markets. His research in these areas has focused on both individual households as well as institutions. With regard to individual household behavior, his research has addressed such topics as the choice of mortgage type, the effect of consumer bankruptcy and foreclosure laws, the relationship of housing wealth and consumption, and the effect of consumer protection regulations such as the Community Reinvestment Act. With regard to institutions, his research has addressed topics such as the effect of the housing-related Government-sponsored enterprises on mortgage rates, the role of bank capital regulation on mortgage markets, and the role of banking competition on credit supply to riskier borrowers.

Juan A. Marchetti is an economist at the World Trade Organization (Trade in Services Division). He is the secretary of the WTO Committee on Trade in Financial Services. He holds a degree in Economics from the University of Buenos Aires.

James McAndrews is Senior Vice President and Head of the Money and Payments Studies Function at the Federal Reserve Bank of New York. He received his PhD in Economics from the University of Iowa, was previously Research Advisor at the Federal Reserve Bank of Philadelphia, and works on payment issues.

David Marqués Ibáñez is an economist at the Financial Research Division of the European Central Bank. He previously worked as a policy economist, in charge of analysing financing conditions in the euro area at the Monetary Policy Directorate of the ECB since the inception of the euro. He holds undergraduates degrees from the University of Granada and the University of Florence, and a Master's degree and a PhD from the University of Wales, Bangor. His research interests are securitization, empirical banking, corporate finance, and microeconometrics.

Maria Soledad Martinez-Peria is a Senior Economist in the Finance and Private Sector Development Team of the Development Economics Research Group at the World Bank. Her published work has focused on currency and banking crises, depositor market discipline, foreign bank participation in developing countries, and access to finance. Currently, she is conducting research on financial sector outreach, bank financing to SMEs, and on the impact of remittances on financial development. Prior to joining the World Bank, she worked at the Brookings Institution, the Central Bank of Argentina, the Federal Reserve Board, and the International Monetary Fund. She holds a PhD in Economics from the University of California, Berkeley and a BA from Stanford University.

Loretta J. Mester is a senior Vice President and director of research at the Federal Reserve Bank of Philadelphia; an Adjunct Professor of Finance at the Wharton School, University of Pennsylvania; and a Fellow at the Wharton Financial Institutions Center. Her research has been published in *Journal of Finance*, *American Economic Review*, *Review of Financial Studies*, and *Review of Economics and Statistics*, among others. She received her PhD in Economics from Princeton University.

Benoît Mojon is the Head of the Monetary Policy Research Division at the Banque de France. He previously worked in the research departments of the Federal Reserve Bank of Chicago (2007–2008) and the European Central Bank (1998–2006). He was Associated Professor at the University of Aix en Provence from 2004 to 2006. His research on monetary policy transmission and on inflation dynamics was published in leading academic journals such as the *Journal of Monetary Economics*, the *Review of Economics and Statistics*, and the NBER Macroannual.

Philip Molyneux is currently Professor in Banking and Finance and Head of Bangor Business School at Bangor University. His research interests include competition, risk, and performance in European banking and developments in global wealth management. In 2001 he was the Visiting Bertill Daniellson Research Fellow at the Stockholm School of Economics and University of Gothenburg. Between 2002 and 2005 he acted as a member of the ECON Financial Services expert panel for the European Parliament. He has recently held visiting Professorships at Bocconi University, Erasmus University, and Bolzano Free University, Italy.

Alan D. Morrison is Professor of Finance at the Saïd Business School of the University of Oxford, a Research Affiliate of the Centre for Economic Policy Research, and a Research Associate of the European Corporate Governance Institute. His research concerns the regulation and industrial organization of commercial and investment banks.

Daniel E. Nolle is a Senior Financial Economist with the US Office of the Comptroller of the Currency. He is responsible for policy analysis on banking structure and performance issues, and he has contributed to the work of the Basel Committee

on Banking Supervision. Daniel Nolle's research interests are in international banking, the US banking system structure, and Internet banking.

Bruno M. Parigi is Professor of Economics, University of Padova, Italy. He has held visiting positions at the European Central Bank, University of Munich, University of Toulouse, and the European University Institute. His research interests are in banking, central banking, and corporate governance. He received his PhD from Rutgers University.

Luiz Fernando de Paula is Associate Professor of Economics at the State University of Rio de Janeiro (Brazil) and CNPq Researcher. His research interests include banking efficiency, financial systems, Keynesian macroeconomics, and economic policies related to emerging countries.

Joe Peek holds the Gatton Endowed Chair in International Banking and Financial Economics at the University of Kentucky. Joe Peek received a BS in Mathematics from Oklahoma State University, and a PhD in Economics from Northwestern University. His research topics include international banking, bank regulation, monetary policy, and credit availability.

José Luis Peydró has been an economist at the European Central Bank since 2005. His specialization is financial intermediation and monetary policy. He holds a PhD in Finance from INSEAD, a Masters in Economics and Finance from CEMFI, and a BA in Economics from the University of Barcelona. He won the First Prize for the best Economics BA student in Spain in 1997.

Eric S. Rosengren is the President and Chief Executive Officer of the Federal Reserve Bank of Boston. In his research, Rosengren has written extensively on macroeconomics, international banking, bank supervision, and risk management. Rosengren holds a BA from Colby College and a PhD from the University of Wisconsin.

Anthony Saunders is the John M. Schiff Professor of Finance and former Chair of the Department of Finance at the Stern School of Business at New York University. He holds positions on the Board of Academic Consultants of the Federal Reserve Board of Governors, as well as the Council of Research Advisors for the Federal National Mortgage Association, and has been a visiting scholar at the Comptroller of the Currency and at the International Monetary Fund.

Wanvimol Sawangngoenyuang is an economist at the Research Division of the Bank of Thailand. She completed her PhD in Economics at Claremont Graduate University. She has taught Economics at California State Polytechnic University and Loyola Marymount University. Her research has focused on financial institutions and capital markets, both domestic and global, with special emphasis on banking crises.

Martin Scheicher is Senior Economist at the Financial Research Division of the European Central Bank. Before joining the ECB he worked at the Austrian Central

Bank and the University of Vienna. His research interests are risk management and empirical finance. His most recent work has focused on the market pricing of credit derivatives. He was educated at the University of Vienna and the London School of Economics.

Kevin J. Stiroh is a Vice President at the Federal Reserve Bank of New York. Kevin Stiroh received a BA in Economics and Psychology from Swarthmore College and a PhD in Economics from Harvard University.

Philip E. Strahan is a Professor of Finance at the Carroll School of Management at Boston College, a Faculty Research Fellow at the National Bureau of Economic Research, and a Fellow at the Wharton Financial Institutions Center. Previously, Professor Strahan spent seven years in the Research and Market Analysis Group of the Federal Reserve Bank of New York. He received a PhD in Economics from the University of Chicago in 1993. Professor Strahan's research interests include the effects of deregulation on the structure and efficiency of the banking industry, the impact of consolidation on bank lending, the political economy of banking, deregulation, and risk management practices in banking.

Anjan V. Thakor is the John E. Simon Professor of Finance and Senior Associate Dean at the Olin School of Business at Washington University in St Louis and a Research Associate of the European Corporate Governance Institute. He was previously the Edward J. Frey Professor of Banking and Finance at the University of Michigan Business School, and the NBD Professor of Finance at the Indiana University Business School. He received his PhD from Northwestern University. He has served as editor of *Journal of Financial Intermediation* and associate editor of many other journals, and is past President of *The Financial Intermediation Research Society.*

Hirofumi Uchida is Associate Professor of Banking and Finance at the Graduate School of Business Administration at Kobe. His research interests focus on financial institutions and financial system architecture. Hirofumi Uchida was a visiting scholar at the Kelley School of Business at Indiana University as a 2003 Fulbright Scholar.

Gregory Udell is the Chase Chair of Banking and Finance at Indiana University and Associate Editor of six academic journals. He has been a consultant to the Federal Reserve Board, the Bank of Japan, the International Financial Corporation, the OECD, the Riksbank, the World Bank, and the Federal Reserve Banks of Chicago and San Francisco.

Paul Wachtel is Professor of Economics at the Stern School of Business at New York University where he has been on the faculty for thirty-five years and has served as Department Chair and Vice Dean. Wachtel is an associate editor of *Journal of Banking & Finance* and of *Journal of Financial Stability.* He has been a consultant

for the World Bank, the European Bank for Reconstruction and Development, the Croatian National Bank, and the Bank of Finland.

Lawrence J. White is a Professor of Economics at the Stern School of Business at New York University. His research encompasses financial services regulation and antitrust and regulatory policy, more broadly. During 1986–1989 he took leave from NYU to serve as a board member on the Federal Home Loan Bank Board.

Jonathan Williams is a Professor in Banking and Finance at Bangor University, where he is affiliated to the Centre of Banking and Finance and Institute of European Finance. His research interests focus on the effects that financial liberalization has had on the banking industry in both industrialized countries and emerging markets.

John O. S. Wilson is Professor of Banking and Finance at the University of St Andrews, Scotland. His research interests focus on the areas of European banking, UK and US credit unions, and financial exclusion. He has been involved in work which examines the profitability and growth of European banks; competition, risk, and performance in European banking; growth, development, diversification, technology adoption, mergers, and performance of credit unions, and their role in tackling financial exclusion.

Bilal Zia is an economist in the Finance Team of the Development Research Group at the World Bank. He joined in July 2006 after completing his PhD in Economics from the Massachusetts Institute of Technology. His research is focused on firms, banks, and access to finance. In recent work, he has studied financial constraints and the allocation of credit subsidies using unique loan-level data from Pakistan; the impact of financial shocks on banks and bank lending; and the effects of financial shocks on industrial composition and corporate groups. He is also involved in studying household level banking services using randomized control trials and household surveys in Indonesia. He holds an MCP and a PhD in Economics from the Massachusetts Institute of Technology, and a BSc (Hons) from the London School of Economics.

This page intentionally left blank

CHAPTER 1

BANKING

AN OVERVIEW

ALLEN N. BERGER
PHILIP MOLYNEUX
JOHN O. S. WILSON[1]

Introduction

Banks play critical roles in every economy. They operate the payments system, are the major source of credit for large swathes of the economy, and (usually) act as a safe haven for depositors' funds. The banking system aids in allocating resources from those in surplus (depositors) to those in deficit (borrowers) by transforming relatively small liquid deposits into larger illiquid loans. This intermediation process helps match deposit and loan supply and provides liquidity to an economy. If intermediation is undertaken in an efficient manner, then deposit and credit demands can be met at low cost, benefiting the parties concerned as well as the economy overall. In addition to these on-balance sheet activities, banking organizations have long engaged in traditional off-balance sheet operations, providing loan commitments, letters of credit, and other guarantees that help counterparties plan for future investments, and in some cases gain access to alternative sources of external finance

[1] Thanks to Mark Flannery and Ed Kane for insightful comments on this chapter.

(e.g., commercial paper market). They also provide an expansive range of various derivative contracts that allow counterparties to hedge their market risks.[2]

In recent years, this simple conceptualization of banking business has radically changed. The largest banks in many countries have transformed themselves, typically via merger and acquisition (M&A), into multi-product financial service conglomerates with offerings including: retail banking, asset management, brokerage, insurance, investment banking, and wealth management. These major developments on the product side have also been matched by the emergence of a diverse array of new funding sources. Driven by securitization, particularly of residential mortgages, banks have become less constrained by their deposit bases for lending. On-balance sheet assets have increasingly been bundled and sold into the market to release capital to finance expansion. Off-balance sheet vehicles (such as Structured Investment Vehicles (SIVs), SIV-lites, and conduits) have been created to enable banks to collateralize assets funded by the issue of short-term paper, not only generating trading profits, but also enabling them to raise resources to finance growing funding gaps (loans minus deposits). Small and medium-sized institutions have also actively participated in diversifying their product and funding features.

The phenomenal growth in structured credit products has been a major recent feature of modern banking business. The issuance of such products in the US and Europe grew from around $500 billion in 2000 to $2.6 trillion in 2007, while global issuance of collateralized debt obligations (CDOs) grew from about $150 billion in 2000 to about $1.2 trillion in 2007 (IMF 2008: 56). Banks actively created off-balance sheet vehicles that packaged various market and credit risks by pooling assets (such as bonds, loans, or mortgage-backed securities) and then divided the resulting cash flows into various tranches (according to their risk/credit rating) that were then sold to investors. The investors could choose to hold different tranches reflecting their risk–return preferences. Up until mid-2007, the demand for structured credit products boomed. Investors were attracted to these securities because they typically appeared to offer higher returns than equivalently rated company bonds. Banks were also attracted to the business as it allowed them to reduce their regulatory capital charges by transferring credit risk to other parties. In general, the view was that the new structured credit products were beneficial as they allowed greater risk to be shared across a broader spectrum of investors, or, to put this another way, banks no longer had to be the major holders of credit risk.

At the peak of the credit cycle in 2006, around one-fifth of US mortgage originations were of the subprime variety, and 75 percent of these were securitized of which around 80 percent were funded by AAA-rated paper (IMF, 2008: 59).

[2] Numerous theories have sought to explain why banking is necessary. These theories primarily relate to: delegated monitoring; information production; liquidity transformation; consumption smoothing; and the role of banks as commitment mechanisms. Notable contributions include: Leland and Pyle, 1977; Diamond and Dybvig, 1983; Diamond, 1984; Fama, 1985; Boyd and Prescott, 1986; Holmstrom and Tirole, 1998; Diamond and Rajan, 2001; and Kashyap, Rajan, and Stein, 2002.

When foreclosures and defaults on US subprime mortgages accelerated from late 2006 onwards, the value of the securities backed by such assets rapidly declined, particularly because the complex nature of the tranching and the lack of transparency of the bank's off-balance sheet vehicles made it nearly impossible to value such assets. Holders of investments backed by subprime mortgages did not know what they were worth, and banks became wary of lending to each other because they also did not know the extent of losses held in structured credit vehicles. In addition, real estate prices tumbled in the US adversely affecting prime and subprime borrowers alike who increasingly defaulted, further putting downward pressure on the value of securitized mortgage products and bank loan books (Foote, et al., 2008). All in all, this culminated in a liquidity freeze in interbank markets and the subsequent credit crunch (Crouhy, Jarrow, and Turnbull, 2008; and Hellwig, 2008).

As the meltdown in credit markets continued, banking sector traumas have been experienced around the globe. The first high-profile casualty surprisingly was not in the US but in the UK, Northern Rock, one of the country's largest mortgage lenders experienced a run on its deposits (the first since Victorian times) and had to be rescued (nationalized) by the government in September 2007. The main cause of failure was cited as a reckless business model, overdependence on short-term wholesale funds, as well as failings in regulatory oversight (HM Treasury, 2008). There then followed further bank collapses. On 16 March 2008, Bear Stearns became the largest casualty of the credit crunch to that date when the failing investment bank was purchased by J.P. Morgan Chase for a nominal amount ($2 per share or $236 million) following the provisions of earlier liquidity support (a revised offer of $10 per share was made on 24 March enabling J. P. Morgan Chase to acquire 39.5 percent of Bear Stearns). In addition, the Federal Reserve extended safety net arrangements to ensure that J.P. Morgan Chase would not suffer significant losses on loans extended to Bear Stearns. On 11 June 2008, the FDIC took over IndyMac Bank, a large alt-A mortgage lender that suffered large losses on these mortgages.[3] The bank had $32 billion in assets, making it the second largest bank failure in US history. The estimated cost of the failure at the time of this writing is $8.9 billion. The takeover followed a slow run or 'walk' on the bank of $1.3 billion in deposits withdrawn between 27 June and 10 July. This followed a public warning about the bank from Senator Charles Schumer. At the same time, Fannie Mae and Freddie Mac, who hold or guarantee over $5 trillion in US mortgages (about half of the total), were having their own problems of a 'walk' on their outstanding stock and shares, both of which declined by more than 80 percent in value from a year earlier. On Sunday, 13 July 2008, Treasury Secretary Henry Paulson announced a plan to insure that both organizations would continue to support the housing market. This consisted of a proposal that the Treasury would

[3] An Alt-A mortgage is a type of US mortgage that, for various reasons, is considered riskier than 'prime' and less risky than 'subprime', and often does not require income verification of the borrower.

temporarily increase its credit lines to the organizations, that they may borrow from the Federal Reserve under certain circumstances, and that the Treasury would get temporary authority to buy their shares should that be necessary. In early September 2008, Freddie Mac and Fannie Mae were placed into conservatorship of the Federal Housing Finance Agency (FHFA). September witnessed further turmoil by the demise of Lehman Brothers and the sale of Merrill Lynch to Bank of America. The two remaining large investment banks, Goldman Sachs and Morgan Stanley, converted to bank holding companies. AIG, the world's largest insurance company was rescued by the Federal Reserve courtesy of an $85 billion emergency loan and in exchange, the Federal government acquired a 79.9 percent equity stake. Washington Mutual (WaMu) was acquired by the US Office of Thrift Supervision (OTS) and the bulk of its untroubled assets sold to J.P. Morgan Chase. In addition to the aforementioned problems, various UK banks also were experiencing severe financing difficulties. HBOS agreed on 17 September 2008 to an emergency acquisition by its UK rival Lloyds TSB after a major fall in its stock price originating from growing fears about its exposure to British and American mortgage-backed securities (MBS). The UK government waived its competition rules making the deal possible. On 29 September, Bradford and Bingley Bank was nationalized. The government assumed control of the bank's £50 billion mortgage and loan portfolio, while its deposit and branch network was sold to Spain's Grupo Santander.

A major feature of the crisis or turmoil that has engulfed banks from September 2008 onwards has been growing market concerns about their capital strength, particularly in the US and Europe. This is despite the fact that many banks have sought to boost solvency by a variety of means. Table 1.1 highlights the amount of capital raised from July 2007 to December 2008 and level of write-downs that have occurred in major banks from the start of 2007 to December 2008. It can be seen that, overall, capital injections exceed write-downs by about $60 billion.

Growing worries about the capital strength of banks led to a collapse in stock prices and widespread bailouts. The highest profile bank bailout being the November rescue of Citigroup. In a complex deal, the US government announced it was purchasing $20 billion of preferred stock in Citigroup and warrants on 4.5 percent of its common stock. The preferred stock carried an 8 percent dividend. This acquisition followed an earlier purchase of $25 billion of the same preferred stock using Troubled Asset Relief Program (TARP) Funds.[4] Under the agreement, Citigroup and regulators will support up to $306 billion of largely residential and commercial real estate loans and certain other assets, which will remain on the bank's balance sheet. Citigroup will shoulder losses on the first $29 billion of that

[4] The TARP is a plan under which the US Treasury would acquire up to $700 billion worth of mortgage-backed securities. After various revisions the plan was introduced on 20 Sept. 2008 by US Treasury Secretary Hank Paulson. At the time of writing, about half of the TARP funds have been allocated, mostly to capital injections in US banks.

Table 1.1. Bank write-downs and capital raised up to December 2008

Bank	Write-down and loss	Capital raised
Wachovia Corporation	96.5	11.0
Citigroup Inc.	67.2	113.8
Merrill Lynch and Co.	55.9	29.9
UBS AG	48.6	34.0
Washington Mutual Inc.	45.6	12.1
HSBC Holdings Plc	33.1	4.9
Bank of America Corp.	27.4	58.5
National City Corp.	26.2	8.9
Morgan Stanley	21.5	24.6
JPMorgan Chase & Co.	20.5	44.7
Lehman Brothers Holdings Inc.	16.2	13.9
Royal Bank of Scotland Group Plc	15.6	50.0
Wells Fargo and Company	14.6	41.8
Credit Suisse Group AG	14.5	12.2
Bayerische Landesbank	14.5	8.8
IKB Deutsche Industriebank AG	14.1	11.6
Deutsche Bank AG	12.7	6.0
ING Groep NV	10.2	18.1
HBOS Plc	9.5	23.6
Crédit Agricole SA	9.4	12.2
Fortis	9.0	22.0
Société Générale	8.2	11.3
Mizuho Financial Group Inc	7.4	6.6
Goldman Sachs Group Inc.	7.1	20.5
Canadian Imperial Bank of Commerce	7.0	2.5
Barclays Plc	6.7	28.0
BNP Paribas	5.8	3.5
Hypo Real Estate Holding AG	5.5	0
KBC Groep NV	5.0	4.8
Dresdner Bank AG	5.0	0
Indymac Bancorp	4.9	0
Natixis	4.7	7.9
Landesbank Baden-Wurttemberg	4.5	0
UniCredit SpA	4.3	10.3
Nomura Holdings Inc.	4.2	5.9
E*TRADE Financial Corp.	4.1	2.2
HSH Nordbank AG	4.0	1.7
Lloyds TSB Group Plc	3.8	13.4
Bank of China Ltd	3.7	0
Rabobank	3.3	0
WestLB AG	3.3	6.9
Bear Stearns Companies Inc.	3.2	0
Commerzbank AG	3.0	11.3
Royal Bank of Canada	2.8	0.3
Fifth Third Bancorp	2.7	6.0

(cont.)

Table 1.1. (*Continued*)

Bank	Write-down and loss	Capital raised
DZ Bank AG	2.5	0
Landesbank Sachsen AG	2.4	0
Sovereign Bancorp Inc.	2.4	1.9
US Bancorp	2.2	6.6
ABN AMRO Holding NV	2	0
Mitsubishi UFJ Financial Group	1.8	10.5
Industrial and Commercial Bank of Chin	1.7	0
KeyCorp	1.6	4.2
Dexia SA	1.6	8.8
Bank Hapoalim B.M.	1.6	2.2
Marshall and Ilsley Corp.	1.5	1.7
Sumitomo Mitsui Financial Group	1.2	12.7
Bank of Montreal	1.1	0
Alliance and Leicester Plc	1.1	0
Groupe Caisse d'Epargne	1.1	0
Bank of Novia Scotia	1.1	0.2
Sumitomo Trust and Banking Co.	1.0	2.0
Gulf International Bank	1	1
National Bank of Canada	0.7	1
DBS Group Holdings Limited	0.2	3.7
American Express	0.0	3.4
Other European Banks (not listed above)	8.5	4.0
Other Asian Banks (not listed above)	5.2	11.4
Other US Banks (not listed above)	3.6	23.0
Other Canadian Banks (not listed above)	0.5	1.4
Total	744.6	805.3

Notes: All the charges stem from the collapse of the US subprime mortgage market and reflect credit losses or write-downs of mortgage assets that are not subprime, as well as charges taken on leveraged-loan commitments since the beginning of 2007. They are net of financial hedges the firms used to mitigate losses and pre-tax figures unless the bank only provided after-tax numbers. Credit losses include the increase in the provisions for bad loans, impacted by the rising defaults in mortgage payments. Capital raised includes common stock, preferred shares, subordinated debt, and hybrid securities, which count as Tier 1 or Tier 2 capital, as well as equity stakes or subsidiaries, sold for capital strengthening. Capital data begins with funds raised in July 2007. All numbers are in billions of US dollars, converted at the October 2008 exchange rate if reported in another currency.

Source: Bloomberg.

portfolio and any remaining losses will be split between Citigroup and the government, with the bank absorbing 10 percent and the government absorbing 90 percent. The Citigroup deal was in certain respects similar to an effort orchestrated by Swiss financial regulators for UBS, another large global bank. In October, the Swiss central bank and UBS reached an agreement to transfer as much as $60 billion of troubled securities and other assets from UBS's balance sheet to a

separate entity. Other major European banks that have sought substantial government support include Royal Bank of Scotland (by late November, the Bank was nearly 60 percent government owned) and Lloyds TSB (that acquired HBOS—already 40 percent state owned).

As the above events clearly indicate, this has been a momentous time for banking and the global economy has not faced such serious financial turmoil since the 1930s. These recent events have shockingly reminded us that the new style of intermediation activity is not without its risks. Banks are among the most leveraged of any type of firm. In the course of business, they rely on scale and various risk management mechanisms to ensure that deposit withdrawals, loan supply, and off-balance sheet obligations can be met. They use their own internal systems and are obliged by regulators, as well as the market, to maintain sufficient levels of capital and liquidity in order to back their business. Regulators also provide safety nets, such as deposit insurance and emergency lending facilities, in order to bolster confidence in the system. History, of course, tells us that, irrespective of what checks and balances are put in place, any inkling of a lack of confidence in an individual bank, a number of banks, or the markets on which banks depend, can signal potential disaster.

By the beginning of 2009 (the time of this writing), concerns about the stability of the global banking system had continued to mount. Between the summer of 2007 and December 2008, the Federal Reserve, the European Central Bank, and the Bank of England all undertook major refinancing operations aimed at injecting liquidity into gridlocked interbank markets. The turmoil has also brought about a range of measures initiated by individual countries relating to the offer of bank guarantees and various rescue plans the most important of which are summarized below :

> Bank deposit guarantees schemes have been strengthened. The governments of Austria, Denmark, Germany, Hungary, Ireland, Slovakia and Slovenia are among those that have announced unlimited guarantees. In addition, the European commission has proposed rules to increase the minimum deposit insurance from €50,000 to €100,000. The US also raised its deposit insurance caps from $100,000 to $250,000 for most accounts. (*Financial Times*, 21 Nov. 2008)

Bank and financial firm rescue plans have been enacted by several governments. In the US, the government bailed out Citigroup and AIG and provided support to Bear Stearns (in relation to the sale to J.P. Morgan Chase, which was conditional on the Federal Reserve lending Bear Stearns $29 billion on a nonrecourse basis). The Federal Reserve has also used the Term Auction Facility (TAF)[5] to provide liquidity to banks and the monthly amount of these auctions increased throughout 2008 to

[5] The Term Auction Facility (TAF) is a temporary program managed by the Federal Reserve aimed to 'address elevated pressures in short-term funding markets'. Under the TAF, the authorities auction collateralized loans with terms of 28 and 84 days to depository institutions that are 'in generally sound financial condition' and 'are expected to remain so over the terms of TAF loans'. Eligible collateral includes a variety of financial assets. The program was introduced in December 2007.

$300 billion by November 2008 (compared with $20 billion when the TAF was introduced in December 2007). A total of $1.6 trillion in loans to banks were made for various types of collateral by November 2008. In October 2008, the Fed announced that it was to expand the collateral it will lend against to include commercial paper, to help address ongoing liquidity concerns. By November 2008, the Federal Reserve had acquired $271 billion of such paper, out of a program limit of $1.4 trillion. Also, in November, the Fed announced the $200 billion Term Asset-Backed Securities Loan Facility—a program supported the issuance of asset-backed securities (ABS) collateralized by loans related to autos, credit cards, education, and small businesses. In the same month, the Federal Reserve also announced a $600 billion program to purchase MBS of Government-Sponsored Enterprises (such as Freddie Mac and Fannie Mae) in a move aimed at reducing mortgage rates.

Countries around the globe have enacted similar measures. In the UK, by the end of November 2008, the government had injected over £37 billion into three banks (RBS, HBOS, and Lloyds TSB). The authorities have also agreed to guarantee £250 billion in bank borrowing. The Bank of England is to lend at least £200 billion to banks via auctions so as to inject liquidity into the system. In France, the state pledged up to €40 billion to recapitalize banks and up to €320 billion to guarantee bank lending. In Germany, the government has agreed to inject a maximum of up to €80 billion for bank recapitalization plus an additional €400 billion in interbank lending guarantees. (See also Chapter 32 and Goddard, Molyneux, and Wilson (forthcoming), for details of other European country bank rescue plans).

In addition to the above, during 2008, many governments have introduced fiscal stimulus packages aimed at boosting demand (for instance, the US government announced a $168 billion package in February) followed by similar announcements later in the year in the UK (£20 billion), Germany (€50 billion), Italy (€80 billion), and Spain (€40 billion). At the time of writing, ramifications of the the impact of the aforementioned bailouts and fiscal stimulus rumble on with a strong possibility that the global economy will experience a severe economic slowdown. The seriousness of the ongoing funding difficulties faced by banks and the potential for significant macroeconomic disruption cannot be understated as identified in the G20 meeting held in Washington, DC on 15 November.[6] The current turmoil will have ramifications for the structure of the banking industry, the strategies which banks follow, how they perform, and how they are regulated and supervised (Baily,

[6] The G20 meeting identified the causes of the financial crises in a formal declaration stating: 'During a period of strong global growth, growing capital flows, and prolonged stability earlier this decade, market participants sought higher yields without an adequate appreciation of the risks and failed to exercise proper due diligence. At the same time, weak underwriting standards, unsound risk management practices, increasingly complex and opaque financial products, and consequent excessive leverage combined to create vulnerabilities in the system. Policymakers, regulators and supervisors, in some advanced countries, did not adequately appreciate and address the risks building up in financial markets, keep pace with financial innovation, or take into account the systemic ramifications of domestic regulatory actions'. See < http://www.whitehouse.gov/news/releases/2008/11/20081115-1.html>.

Elmendorf, and Litan, 2008; Buiter, 2008; Caprio, Demirgüç-Kunt, and Kane, 2008; Yellen, 2008; and Udell, 2009). Acharya and Richardson (2009) produce an edited collection of contributions by prominent academics that offers financial policy recommendations and actions to restore the global financial system.

Back to the future: Drivers of change

Now let us turn the clock back to the start of 2007—how things looked so different then! Banks in the US and Europe had been posting record profits, major risks appeared to have abated, and banking systems had been deregulated, allowing for more competition and innovation. The general operating environment over the preceding decade or so had favored banks: declining interest rates, the stock market bubble of 1991–2001, and relatively buoyant economic growth (fueled to a major extent by booming real estate values and the bank credit that funded this) provided the bedrock for the strong performance of banking systems.

In addition to the generally favorable economic climate, banking business has been transformed by deregulation that removed barriers to competition in traditional and new (non-banking) product areas as well as geographically. In the US, the Gramm-Leach-Bliley (GLB) Act of 1999 effectively repealed the Glass-Steagall Act (1933) and granted broad-based securities and insurance powers to commercial banking organizations. Also, the Interstate Banking and Branching Efficiency Act of 1994 reduced geographical barriers to competition by permitting almost nationwide branch banking as of 1997 (although there was a national deposit cap of 10 percent). In Europe, the European Union's Single Market Programme had legislated for the possibility of a universal banking based system and a single banking license in 1992, and the introduction of the euro in 1999 further removed barriers to cross-border trade in banking and financial services (Goddard, Molyneux, and Wilson 2001; and Goddard, et al., 2007). Likewise, in Japan, the 'Big Bang' reforms introduced between 1998 and 2001 established a universal banking model, also enabling greater access to foreign banks wishing to enter the domestic financial services sector.

Inextricably linked to the deregulation trend have been the moves toward the harmonizing of regulations—across countries and different financial service sectors. In general, there has been a strong policy move to create more uniform regulatory structures so that no jurisdiction or sector of the financial services industry has an unfair competitive advantage. This is best reflected in European Union harmonization of financial services regulation under the Single Market Program as well as capital adequacy regulation under the Bank for International Settlements Basel I (1988), and the more recent updated Basel II (2006), that

establishes minimum capital adequacy guidelines for internationally active banks. Virtually all developed countries' banking systems, and most others, currently adhere to Basel capital standards. In the European Union, Basel II is to be incorporated into European Union law under the Capital Adequacy Directive 3 (CAD 3) and in the US transition to Basel II began in 2008. The Bank for International Settlements (BIS) has been instrumental in helping to establish minimum international standards in the regulation of banks (particularly in emerging and less-developed countries) via its guidance and oversight on the 'Core Principles for Effective Banking Supervision' (1997; 2006). Other noteworthy initiatives include the OECD's anti-money-laundering/anti-terrorist financing initiatives under the aegis of the Financial Action Task Force.

Technology has been another important element in transforming the banking industry. Banks are major users of IT and other financial technologies. Technological advances have revolutionized both back-office processing and analysis of financial data, as well as front-office delivery systems. Evidence suggests that the former has led to significant improvements in bank costs and lending capacity whereas the latter has improved the quality and variety of banking services available to customers (Berger, 2003). Possibly the most substantial impact of technology on the banking system has been on the payments system, where electronic payments technologies and funds transfers have replaced paper-based payments (cash and checks) and paper record-keeping. The reduction in costs from such changes has been significant (Humphrey, et al., 2006).

Developments in financial technologies, including new tools of financial engineering and risk management, coupled with the growth of new and broader derivatives markets, were also believed (up until recently at least) to have improved banks' risk management capabilities.[7] The wider use of interest rate swaps and other derivative products facilitated by advances in trading and risk management technologies meant that banks could manage interest rate risk and other market risks more effectively. During the 1990s, banks shifted more of their activity toward non-interest income as a source of revenue, including trading revenue. As such, they increasingly adopted new (e.g., value-at-risk or VAR) technologies to manage their market risk. Similar risk measurement financial technologies, which link capital requirements to credit risk have been incorporated into Basel II.

Advances in new technologies have been inextricably linked to financial innovation (Molyneux and Shamrouk, 1999; and Frame and White, 2004). The meteoric growth in asset securitization stands out as a key example. Here the financial innovation element relates to the creation of synthetic liquid, tradable securities

[7] The current turmoil in credit markets begs the question as to why so many practitioners, policymakers, and other commentators did not see the limitations of the risk management practices being implemented. Arguments relating to disaster myopia and the difficulty in pricing various risks ('Knightian uncertainty') are possible explanations.

created from pools of illiquid, non-tradeable assets (for instance, individual residential mortgages, credit card receivables, and so on) where usually the payoff features of the traded securities differ significantly from those of the underlying assets. Advances in technology enable the efficient monitoring and analysis of information related to the performance and operation of the asset pools. A key aspect of this process relates to advances in credit scoring technology. This enables banks to transform quantitative information about individual borrowers (such as income, employment history, past payment record, and so on) into a 'single numerical credit score, which lenders can use when screening and approving loan applications; securitizers can use (this information) to group loans of similar risk into pools, and investors can also use this (together with other information) to evaluate the risk of the resulting asset-backed securities' (DeYoung, 2007: 46).

In addition to asset securitization/structured products mentioned earlier, there has also been an explosion in the types of financial products and services available. The emergence of alternative assets including hedge funds, private equity, REITs (real estate investment trusts), commodities, and their respective indices means that both retail and professional investors now have access to an extensive range of investments that seek to hunt out absolute (alpha) returns. These are complemented by the rapid recent growth in index trackers—such as Exchange-Traded Funds (ETFs)—that provide beta returns tracking a myriad of indexes—either long or short. The latter are increasingly competing with actively managed mutual funds (*Economist*, 1 Mar. 2008). As banks have sought to diversify their revenues, many have sought to build substantial private banking/wealth management franchises offering such services.

It is widely recognized that factors including the buoyant economic environment, deregulation, and technological and financial innovation have transformed the banking landscape. However, less has been said about the changing strategic focus of banks. The aforementioned forces have resulted in generally more competitive banking systems and this also has meant that banks now have to compete more aggressively than ever before for their key resource—capital. It is uncertain as to why banks wish to maintain capital resources well in excess of their regulatory minimum. According to Berger, et al. (2008), this could be for a number of reasons—high earnings retention, the perceived advantages associated with high economic capital (e.g., protection of a valuable charter), acquisition plans, and/or external pressure from regulators or the financial markets. Some banking organizations may also hold excess capital in anticipation of a crisis in order to cover a significant portion of losses, and to allow more lending and off-balance sheet activities than would otherwise be the case under such conditions, gaining market share on their less-capitalized competitors (Berger and Bouwman, 2009b). Whatever the motives, it is a fact that many banks have held capital resources well in excess of their regulatory minimums and that higher capitalized banks also tend to be better performers. The motives noted above, of course, are not mutually exclusive and it has been primarily the strategic desire of banks to manage their

most costly resource—economic and regulatory capital—more efficiently in order to boost value creation for their owners. This simply means that banks have increasingly focused on strategies that seek to generate risk-adjusted returns in excess of the opportunity cost of capital (Fiordelisi and Molyneux, 2006). Corporate restructuring practices commonplace in the non-financial sector have become widespread in the banking industry. Business process re-engineering, outsourcing, open architecture (providing third-party products and services), joint ventures with high tech firms, third-party processing, and the drive for mega-scale in key sectors (credit cards, global custody, treasury activities) have primarily been motivated by the desire to generate risk-adjusted returns (in excess of the cost of capital) in order to boost returns and value for shareholders. Markets are not only more important for the business that banks do, but also for gauging their performance, especially as they all seek to raise costly capital from an ever-widening group of investors. The desire to generate returns sufficient to obtain capital resources at the appropriate cost has encouraged banks into many new areas of business—particularly those areas where capital requirements are less onerous compared with traditional on-balance sheet credits. The securitization phenomenon has been a major (spiraling) outcome of this trend.

Transformation of the Banking Landscape

The banking industry has been transforming for decades now. A study by Berger, et al. (1995) documented significant changes in the US banking industry from 1979 (prior to major deregulation of the early 1980s) to 1994 (prior to the effects of the Interstate Banking and Branching Efficiency Act of 1994, which permitted almost nationwide branch banking). The authors found that virtually all aspects of the US banking industry had changed dramatically over these fifteen years. Over one-third of all independent banking organizations (top-tier bank holding companies or unaffiliated banks) disappeared over the 1979–94 period, even while the industry was growing.[8] On the asset side of the balance sheet, the industry lost market power over many of its large borrowers, who were able to choose among many alternative sources of finance. On the liability side, the industry evolved from a position of protected monopsony in which banks purchased deposit funds at regulated, below-market interest rates toward a market setting in which banks paid closer to competitive prices to raise funds. With respect to individual consumers, electronic

[8] The number of FDIC insured commercial banks fell by 27% over the same period.

interfaces such as automated teller machines and online banking altered the way many customers interact with their banks.

Over the last decade or so since that study, the structural features of global banking systems have radically altered. Developed banking markets have all experienced significant declines in the number of banks and industry concentration has generally increased at both the national and regional levels. In the US, for example, the number of FDIC-insured commercial banks fell from 10,359 in 1994 to 7,283 by December of 2007. Substantial declines have also been witnessed in Europe (Goddard, et al., 2007) and Japan. The decline in the number of banks, particularly in developed countries, however, has not been matched by a fall in the number of branches, quite the opposite. For example, the number of bank branches in the US increased 27 percent between 1994 and 2006, although average branch size (measured by the number of employees) has fallen (Hannan and Hanweck, 2008). Evidence from Europe also illustrates the current trend to increased branch numbers: they grew by around 5 percent between 2002 and 2006 (European Central Bank, 2007a). Of course, within Europe there are substantial differences across countries: Germany and the UK experienced declines, while France, Italy, and Spain had substantial increases over the aforementioned period.

The decline in the number of banks has mainly been a consequence of the M&A trend. The US stands out in this respect, as between 1980 and 2005, 11,500 bank mergers took place amounting to 440 mergers annually (Mester, 2007). Europe has also experienced substantial consolidation both domestically and increasingly on a cross-border basis. The latter has been motivated by the attractions of Europe's single market as well as the limited growth prospects available in increasingly congested domestic banking systems. Western European banks have also been major acquirers in the transition economies of Eastern Europe, where their financial systems are now dominated by foreign institutions. Spanish banks have a major presence throughout Latin America. The larger US banks have focused on building substantial regional (if not national) franchises as well as acquiring banks particularly in Mexico and also in Latin America. Many large banking organizations have acquired wholly or partially a wide range of banks in Asia, with a particular focus on China, India, and, most recently, Vietnam.

While the consolidation trend has had the overall impact of reducing the number of banks operating in many large developed markets, this trend is not universal. In many countries there has been an increase in foreign institutions. For example, between 2002 and 2006, out of twenty-seven European Union member states, bank numbers increased in Denmark, Estonia, Greece, Latvia, Lithuania, Malta, and Slovakia. All of these countries, apart from Denmark and Greece, are new members to the European Union that experienced significant foreign bank entry (European Central Bank, 2007a). The increase in foreign bank presence in different parts of the world is determined by a number of factors. As noted by Berger (2007), the high proportion of foreign banks in Eastern Europe is mainly a result of state

privatization and a lack of local banking experience, whereas in Latin America it is more a consequence of liberalization programs post-crisis. 'It is also noteworthy that three significant present and future economic powerhouses in Asia—Japan, China, and India—have relatively low foreign bank penetration' (Berger, 2007: 1969). Evidence suggests that foreign banks tend to have a larger presence in systems where entry barriers are low and typically where they tend to be more efficient compared to their domestic counterparts in developing/emerging markets (Claessens, et al., 2001). In contrast, foreign banks tend to be less efficient than their domestic competitors in developed countries (Berger, et al., 2000), and this is often forwarded as an explanation for the relatively modest (albeit increasing) foreign bank shares in many developed financial systems.

Another interesting recent development has been for banks and sovereign wealth funds from emerging economies to make acquisitions/or to acquire stakes in international banks (Paulson, 2009). In March 2008, for example, it was confirmed that China's biggest bank, the Industrial and Commercial Bank of China (ICBC), had acquired a 20 percent stake in South Africa's largest bank, Standard Bank. The second half of 2007 and early 2008 also witnessed unprecedented injections of capital by sovereign wealth funds into major banks so as to shore up their eroding capital bases. Major examples include: Barclays (China Development Bank and Temasek, Singapore), Citigroup (Abu Dhabi Investment Authority), Credit Suisse (Qatar Investment Authority), Morgan Stanley (Chinese Investment Corporation), Merrill Lynch (Temasek, Singapore), and UBS (Government of Singapore Investment Corporation). This trend reflected the strength of emerging markets and the desire of banks and investors from these countries to pursue international diversification strategies by gaining ownership presence in major international banks.

In the light of current market turmoil, the general consolidation trend and the growing presence of foreign banks is likely to continue—perhaps seeing an increasing number of major Western banks being acquired by emerging market institutions. What seems inevitable, however, is that banking systems in developed countries at least will continue to remain concentrated with a handful of banks dominating domestic systems. Projections about the future structure of US banking, for instance, envisage a system, 'characterized by several thousand very small to medium-size community bank organizations, a less numerous group of midsize regional organizations, and a handful of extremely large multinational banking organizations. . . . the US banking industry is not likely to resemble the banking industries in countries such as Germany, which have only a handful of universal banks' (Jones and Critchfield, 2005: 48). One should add, however, that a key difference between banking in the US and in other developed countries relates to the high level of new bank entrants. DeYoung (2007) notes that in the 1980s, 1990s, and early 2000s, around 3,000 new banking charters were granted by state and federal banking supervisory authorities—and there is evidence that many of these de novo banks were established in markets where established banks had been acquired (Berger, et al., 2004). As far as we are aware, no other developed banking system shows anywhere near this level of new entry. In part because of this

entry and in part because of rigorous antitrust enforcement which requires divestiture in mergers with significant local market overlap, local banking markets in the US have not become more concentrated over time. From 1994 to 2006, the mean local commercial bank deposit Herfindahl-Hirschman Index (HHI) actually fell slightly from .1976 to .1785 in metropolitan markets and from 0.4208 to 0.3847 in rural markets. In Europe, out of twenty-seven European Union countries, the asset market share of the top five banks has slightly fallen in over half the member countries since 2002. However, national concentration levels remain high—the five bank asset concentration ratio for the monetary union (euro) countries averaged 53.7 percent in 2006, and for the EU27 at 58.9 percent (European Central Bank, 2007a). Levels of concentration are typically much higher for the new member states.

The dominance of a handful of banks in national or/and local banking systems raises questions about competition and market power. Much attention has been paid to the impact of consolidation on small business lending, suggesting that as small banks have an advantage in processing 'soft' information they are more able to build stronger relationships and thus are better placed to provide credit facilities. The argument goes that big banks rely on 'hard' information reflected in transactional banking which somehow cannot deal with informationally opaque customers. Put simply, consolidation would be expected to reduce relationship banking and boost transactional banking leading to a reduction in lending to small firms. To a certain extent, the literature finds that big banks behave differently from small banks. Berger, et al. (2005), for example, find that large banks tend to lend at a greater distance, interact more impersonally with their borrowers, have shorter and less-exclusive relationships, and typically do not alleviate credit constraints as effectively. While the stylized dichotomy between the role of small and large banks in small business lending is widely noted in the literature, in reality the market for small business credit (in the US at least) is much more complex, reflected in a broad array of different lending technologies with markets exhibiting contestable features (Berger and Udell, 2006; and Berger, Rosen, and Udell, 2007). Evidence from the US, in general, does not support the view that consolidation has reduced the quantity or increased the pricing of small business banking services, as other local banks tend to pick up small business credits that are discarded by consolidating banks (Berger, et al., 1998; and Avery and Samolyk, 2004), although there is more mixed evidence from Europe (Bonaccorsi di Patti and Gobbi, 2007) and Japan (Uchida, Udell, and Watanabe, 2007).

The consolidation trend has also spawned an extensive literature looking at a broad array of features (see DeYoung, et al., forthcoming, for a detailed review). A recent snapshot of the sort of related issues considered include:

- Motives for mergers: Hughes, et al. (2003); Campa and Hernando (2006); DeLong and DeYoung (2007)
- Features of acquisition targets: Focarelli, Panetta, and Salleo (2002); Hosono, Sakai, and Tsuru (2006); Valkanov and Kleimeier (2007)

- Diversification benefits: Stiroh and Rumble (2006); Pozzolo and Focarelli (2007)
- Impact of deregulation on M&A activity: Carletti, et al. (2007); Jeon and Miller (2007)
- Multimarket competition effects of M&A activity: Hannan and Prager (2004); Berger and Dick (2007); Berger, et al. (2007)
- Impact of mergers on deposit prices: Focarelli and Panetta (2003); Craig and Dinger (2007)
- Monoline versus universal financial service providers: Yom (2005)
- Exploiting safety net subsidies: Mishkin (2006); Brewer and Jagtiani (2007)
- Impact of mergers on systemic risk: De Nicolo and Kwast (2002); Baele, De Jonghe, and Vander Vennet (2007)
- Efficiency effects of mergers: Carbó and Humphrey (2004); Cornett, McNutt, and Tehranian (2006); Hannan and Pilloff (2006); De Guevara and Maudos (2007).

Clearly the structural landscape of banking in many countries has changed as a result of the consolidation trend as well as the increased presence of foreign institutions. The landscape will inevitably further change in the light of widespread bank bailout and rescue packages. This changing environment poses various challenges to shareholders and managers in the context of appropriate corporate governance structures, as well as to regulators who need to determine the appropriate way to supervise universal banking firms that have global reach (Caprio, Demirgüç-Kunt, and Kane, 2008).

Despite these major structural developments, up until mid-2007 there was a general consensus that the US and many other banking systems were sound, particularly because banks appeared to be holding historically high levels of capital bolstered by Basel I and (more recently) Basel II requirements (Berger, et al., 2008). There was also evidence that for large banks their strong capital positions complemented their liquidity creation, although for small banks capital strength and liquidity creation appear to move in opposite directions (Berger and Bouwman, 2004a). Having said this, however, in Europe there had been a gradual trend for banks to hold lower levels of capital but more liquidity on the balance sheet compared to US banks, and the level of liquid assets systematically declined (albeit from high levels) from the 1990s onwards. High capital levels were required to back the rapid credit expansion that occurred from the mid-1990s onwards. In the US and UK, credit growth far exceeded core deposit gathering, leaving significant funding gaps that had to be financed via the interbank market and from securitization activity. This new intermediation model worked extremely well with an increasing portion of retail lending by banks shifting from portfolio lending that generated interest income to securitized lending that earned non-interest revenue. As competition in traditional deposit and lending business squeezed interest margins, product market deregulation, such as the Gramm-Leach-Bliley Act of 1999 in the US and Japan's 'Big-Bang', enabled banking

organizations to produce or sell a wider range of services including equity and debt underwriting, securities brokerage, and insurance products and consequently increase non-interest income. In developed banking systems, cost levels were generally flat (if not declining) and the increase in total revenues from traditional and (increasingly) non-traditional sources meant that by the mid-2000s, bank profitability was strong in many countries. By 2006, the banking systems of France, Italy, Spain, Sweden, and the UK were posting returns on equity around 20 percent (European Central Bank, 2007b), and in the US 'return on assets edged up to match its highest annual level in recent decades' (Carlson and Weinbach, 2007: A37).

By early 2007, the general consensus appeared to be that high-performing banking systems, supported by excess capital and state-of-the-art risk management capabilities, bolstered by appropriate market-based regulation, would continue to finance growth at recent historical levels. Things have now certainly changed!

At the time of writing, the prospects for the banking industry are almost diametrically opposed to the view held eighteen months earlier. Commercial and residential real estate values continue to fall, avenues for bank financing via the securitization business and interbank markets have dried up, and major banks have suffered large losses of capital. Consequently, many of the largest banking organizations have had to raise additional capital (as illustrated in Table 1.1). Banks' funding gaps (loans minus deposits) remain at historically high levels, with added concerns about the ability of the system to meet substantial off-balance sheet commitments that still may be drawn. Much of this has yet to feed through into the real economy.

This handbook aims to provide further insight into many of the aforementioned developments as well as indications of future prospects in the banking business.

Book structure and chapter summaries

The contents of this Handbook are subdivided into five Parts, as follows: Part I, The Theory of Banking; Part II, Regulatory and Policy Perspectives; Part III, Bank Performance; Part IV, Macroeconomic Perspectives in Banking; and Part V, International Differences in Banking Structures and Environments. This section provides brief summaries of the chapter contents.

The theory of banking

Part I of this Handbook comprises seven chapters and examines why banks exist, how they function, how they are managed, and their legal, organizational, and

governance structures. Particular emphasis is placed on the evolution of banks within the wider financial system. It is noted that the scale, scope, and complexity of banking business have increased as banks have diversified across product and geographic lines. This has led to changes in the techniques used by banks to manage liquidity, credit, and other risks. New complex organizational structures have emerged, including large international financial conglomerates that pose new challenges for regulation and supervision.

In Chapter 2, Franklin Allen and Elena Carletti examine the roles of banks in ameliorating informational asymmetries that may arise between lenders and borrowers; providing inter-temporal smoothing of risk; and contributing to economic growth in Europe, the US, and Asia. In general, euro area countries have small but rapidly developing stock markets. Bank lending relative to GDP is substantial, and bond markets play an important role in the financial system. The UK has a large stock market and a large banking sector, but the UK bond market is relatively small. The US banking sector is small in relation to the size of the US economy, but both the stock market and the bond market are relatively large. Japan has a relatively large banking sector and highly developed capital markets. The chapter compares the role of bank-based and market-based banking systems and discusses aspects relating to relationship banking and the finance and growth debate. An interesting finding is that market-based financial systems like the US tend to be more innovative than bank-based systems.

Competition among banks, and between banks and non-banking financial institutions and financial markets has intensified in recent years. This competition has led to the transformation of banks, and the growing complementarities between banks and capital markets. In Chapter 3, Arnoud Boot and Anjan Thakor examine how banks choose between relationship- and transactions-based lending, and more generally the role of debt versus equity instruments and the economic functions of banks. The arguments presented suggest that banks have a growing dependence on the capital market for sources of revenue, for raising equity capital and for risk management, while capital market participants rely increasingly on banks' skills in financial innovation and portfolio management. The increased integration of banks with financial markets raises domestic and cross-border financial stability concerns, which in turn has implications for the design of domestic and international financial system regulation.

Banks are exposed to market risk, interest rate risk, credit risk, liquidity risk, and operational risk. For any bank, the measurement and management of risk is of the utmost importance. In Chapter 4, Linda Allen and Anthony Saunders describe the widely used VAR method of risk measurement. Accurate risk measurement enables banks to develop a risk management strategy, using derivative instruments such as futures, forwards, options, and swaps. However, the recent subprime crisis demonstrates that the use of derivative instruments does not by itself mitigate the risks of banking.

One of the key functions of the banking sector is maintaining liquidity. Banks use short-term liquid deposits to finance longer-term illiquid lending, and provide liquidity off the balance sheet in the form of loan commitments and other claims on their liquid funds. In Chapter 5, Philip Strahan examines the role of banks in providing funding liquidity (the ability to raise cash on demand) and in maintaining market liquidity (the ability to trade assets at low cost), thereby enhancing the efficiency of financial markets. Banks dominate in the provision of funding liquidity because of the structure of their balance sheets as well as their access to government-guaranteed deposits and central bank liquidity. There is considerable functional overlap between commercial banks and other financial institutions in providing market liquidity through devices such as loan syndication and securitization.

Deregulation and technological innovation have permitted banking organizations such as financial holding companies to capture an increasing share of their revenue stream from non-interest sources. The increase in non-interest income reflects in part diversification into investment banking, venture capital, insurance underwriting, and fee- and commission-paying services linked to traditional retail banking services. In Chapter 6, Kevin Stiroh examines the effects of diversification on the risk and return characteristics of financial institutions. In many cases, risk-adjusted returns have declined following diversification into non-interest earning activities. This phenomenon may be due to a tendency to diversify revenue streams, rather than clients, with the effect that interest and non-interest income are increasingly exposed to the same shocks. Alternatively, managers may have been willing to sacrifice profits to achieve growth through diversification, or the adjustment costs associated with diversification may have been larger than anticipated.

Under a universal banking model, the services of both commercial and investment banks are provided under one roof. Universal banks provide traditional deposit taking, lending, and payments services, as well as asset management, brokerage, insurance, non-financial business (commerce), and securities underwriting services. In Chapter 7, Alan Morrison examines the evolution of universal banking across countries and over time. Universal banking has operated in Germany for many years, but was generally restricted in the US (via Section 20 subsidiaries) until Congress passed the 1999 Gramm-Leach-Bliley Act. Potential conflicts of interests, such as the cross-selling of inappropriate in-house insurance and investment services to bank customers, or the mispricing of internal capital transfers between different parts of financial service groups and so on, are key issues for the universal banking model, which present significant challenges for regulation and for the wider health of the financial system.

As commercial banks have diversified into investment banking, a number of large international conglomerates have emerged. In Chapter 8, Richard Herring and Jacopo Carmassi examine the phenomenon of the international financial conglomerate. Typically, conglomerates have complex organizational structures—in some cases comprising hundreds of majority-owned subsidiaries. A subsidiary-based

model may help ease problems of asymmetric information among shareholders, creditors, and managers; mitigate conflicts of interest; insulate the rest of the group from risk emanating from individual subsidiaries; or reduce taxes. However, the very scale and complexity of the largest international financial conglomerates poses new threats to the stability of the global financial system.

Regulatory and policy perspectives

Part II of this Handbook comprises nine chapters and examines the various roles of central banks, regulatory and supervisory authorities, and other government agencies which impact directly on the banking industry. Central banks execute monetary policy, which operates to a large degree through the banking system; act as a lender of last resort; and perform various other functions such as operating parts of the payments system. Government agencies provide safety net protection—such as explicit or implicit deposit insurance, unconditional payment system guarantees, and takeovers of troubled institutions—to prevent widespread or systemic bank failure. In part to protect against systemic failure, and in part to offset some of the perverse incentive effects of government safety net protection, government authorities also engage in prudential regulation and supervision, and set policies concerning bank closure. Competition and antitrust policy aimed at preventing abuses of market power also impact directly on the banking industry. So too do explicit or implicit government policy concerning foreign entry into domestic markets and foreign ownership of domestic industry.

The historical evolution of central banks has been shaped by successive monetary and financial crises throughout the nineteenth and twentieth centuries. In Chapter 9, Michel Aglietta and Benoît Mojon examine issues related to central banking. Today, the four major tasks of the central bank are: the settlement of interbank payments; bank regulation and supervision; lender of last resort; and the execution of monetary policy. However, not all central banks perform all four tasks; in some countries one or more of these functions is delegated to separate government agencies. In the UK, for example, regulation and supervision is the responsibility of the Financial Services Authority (FSA). In the US, responsibility for regulation and supervision is divided among several agencies, including the Office of the Comptroller of the Currency (OCC) and the Federal Deposit Insurance Corporation (FDIC), as well as the central bank (the Federal Reserve). Future challenges that will influence the further evolution of central banking include securitization, electronic payments, asset price volatility, and the increasing internationalization of the banking industry.

In Chapter 10, Joe Peek and Eric Rosengren examine the role of the central bank in executing monetary policy, and the broader role of the banking sector in monetary policy transmission. Monetary policy is believed to affect real expenditure

through three channels: the traditional interest rate channel, whereby changes in interest rates affect the spending preferences of consumers; the broad credit channel, whereby interest rate changes influence investor behavior and the borrowing preferences of the corporate sector; and the bank lending channel, whereby monetary policy affects the supply of bank credit through its effect on depositor behavior, or changes in the value of bank assets and liabilities. The empirical research reviewed using aggregate data suggests that bank lending contracts when monetary policy becomes tighter. The research also suggests that the effects of monetary policy will be influenced by the characteristics of the banking industry. For example, banks' capital constraints may limit their ability to increase lending in response to expansionary monetary policy. Various forms of financial innovation (such as securitization) may reduce the future importance of the bank lending channel.

Central banks also play an important role as lender of last resort to banks experiencing liquidity problems. Lending of last resort provides insolvent banks with liquidity and allows them to escape market discipline. In Chapter 11, Xavier Freixas and Bruno Parigi examine this lender of last resort function, and its relationship with bank closure policy. The difficulties in distinguishing liquidity and solvency shocks are highlighted. The lender of last resort function is usually handled by the central bank, while scrutiny of bank closure is commonly the responsibility of a separate agency, often a deposit insurer. The current financial crisis highlights the complexity of the lender of last resort function, which encompasses issues relating to monetary policy, bank supervision and regulation, and the operation of the interbank market. The authors posit that the lender of last resort function should be an integral and interdependent part of an overall banking safety net, which encompasses a deposit insurance system, a system of capital regulation, and a set of legal procedures to bailout or liquidate troubled banks.

The design of regulatory arrangements for the banking industry can lead to conflicts of interest that have the potential to undermine the quality of supervision and enforcement. In Chapter 12, Ed Kane explains how in extreme cases, conflicts of interest, combined with intense competition and technological and financial innovation, can give rise to inappropriate behavior on the part of bankers, increasing the probability of a banking crisis. Kane notes that recent technological change and regulatory competition has encouraged banks to securitize their loans in ways that masked credit risks, while supervisors have outsourced much of their responsibility to credit rating agencies.

Deposit insurance is intended to prevent 'runs' on individual banks by depositors. It also limits losses to depositors in the event of bank failure, and reduces the risk that a run on one bank might undermine confidence in others through a contagion effect. However, a flawed deposit insurance system might cause more harm than good, if moral hazard results in excessive risk taking or recklessness on the part of banks. In Chapter 13, Robert Eisenbeis and George Kaufman show a

well-functioning and efficient deposit insurance guarantee system which involved: closing promptly struggling banks when leverage ratios declines to an unacceptably low level; assigning credit losses to uninsured bank claimants promptly; and re-opening closed institutions as soon as possible to allow insured depositors and pre-existing borrowers full access to their funds and credit lines. Such a system, the authors argue, can be designed as part of an efficient financial safety net system.

To reduce moral hazard and systemic risk, regulators require banks to hold capital in order to absorb unforeseen risks. Standards developed by the Basel Committee on Banking Supervision (via Basel I and Basel II) have gone some way to aligning such capital requirements with banks' risk profiles. In Chapter 14, Michael Gordy and Erik Heitfield examine the rationale for capital regulation, and describe the key features of Basel II. The authors focus on the theoretical and empirical underpinnings of Basel II, and the challenges in rating the riskiness of assets contained in bank portfolios. A key issue identified is the extent to which the bank risk rating systems are responsive to changes in borrower default risk over the business cycle. If this is indeed the case, capital requirements under the Internal-Ratings-Based (IRB) approach will increase as an economy moves into recession and decline as an economy moves an expansion. Basel II is likely to make it more difficult for policymakers to maintain macroeconomic stability if banks' lending is procyclical.

As banking business has become increasingly complex, the usefulness of the traditional tools of supervision in monitoring risk taking by banks has been called into question. However, the efficient markets hypothesis suggests that private investors might be able to identify the risks associated with investing in shares of banks and other complex financial institutions, and thus exert market discipline. In Chapter 15, Mark Flannery defines the concept of market discipline and explains the importance of its two main components: market *monitoring* of bank value and market *influence* over a bank's strategic choices. He contends that banks' *ex ante* strategic choices will reflect market conditions if investors can identify and react to the condition of banking firms in a timely and accurate manner. The evidence presented in the chapter suggests that forward-looking market-based information embodied in banks' equity prices, debt instruments, and ratings can act to inform bank supervision, and to indicate the need for supervisory corrective action sooner than might otherwise be the case.

Competition in banking is important, because any form of market failure or anti-competitive behavior on the part of banks has far-reaching implications for productive efficiency, consumer welfare, and economic growth. In Chapter 16, Astrid Dick and Timothy Hannan review the methodologies used by researchers and policymakers to assess the form and intensity of competition in the banking industry. Methods for the measurement of competition are based on the structure–conduct–performance paradigm, and non-structural approaches from the new empirical industrial organization literature. Recently, alternative techniques have been developed, based on structural simulation modelling. Finally, the

authors examine how competition policy in the European and US banking industry has evolved. In recent times, policy in Europe has moved closer to that of the approach advocated by regulators in the US, in which regulators not only deal with the outcomes of potential mergers, but also more widely with the potential barriers to competition in the financial services industry.

Internationalization of banking and other financial services has been promoted by financial deregulation, as well as the globalization trend. To assess the extent of market openness and the potential for cross-border provision of financial services, it is useful to examine the commitments to greater openness made by countries under the auspices of the World Trade Organization. In Chapter 17, James Barth, Juan Marchetti, Daniel Nolle, and Wanvimol Sawangngoenyuang present evidence of some significant and worrying divergences between WTO commitments freely entered into by national governments, and the practices of bankers and national regulators. As a whole, developed countries tend to be more open than developing countries, but developing countries are more likely to meet their WTO obligations than developed countries.

Bank performance

Part III of the book comprises seven chapters dealing with bank performance. A number of issues are assessed, including efficiency, technological change, globalization, and ability to deliver small business, consumer, and mortgage lending services.

In Chapter 18, Joseph Hughes and Loretta Mester outline the different approaches used to examine the efficiency and overall performance of banks. Here the authors outline various structural and non-structural approaches to efficiency measurement. The structural approach requires a choice of the underlying production features of banking (intermediation, production, value-added, or other) and the specification of cost, profit, or revenue functions, from which (using various optimization techniques) one can derive relative performance measures. The authors stress that the role of capital and risk is important in bank's production features and therefore should be included in structural evaluations of bank performance. Non-structural approaches simply relate to the use of accounting/financial ratios to measure bank performance. The chapter highlights the growing interest in using structural approaches to examine corporate governance and ownership issues, whereas non-structural indicators (such as Tobin's Q ratio) are widely used as indicators of the value of a bank's investment opportunities (or charter values). The chapter ends with a brief discussion on how consolidation has impacted bank performance.

Technological advances and financial innovation have led to fundamental changes in the nature of banking over the last twenty-five years. In Chapter 19, Scott Frame and Lawrence White focus on innovations in banking products (subprime mortgages,

retail services including the growth of debit cards, online banking, and the use of pre-paid cards) and processes (automated clearing houses, small businesses credit scoring, asset securitization, and risk management). In addition, various new organizational forms, such as internet-only banks and the establishment of Section 20 securities subsidiaries are discussed. Financial and technological innovations have impacted on bank performance and the wider economy.

International mergers between banks are relatively recent phenomena. In Chapter 20, Claudia Buch and Gayle DeLong examine the causes and effects of international bank mergers. The authors examine the determinants of and barriers to cross-border bank mergers, and their impact on the efficiency, competitiveness, and riskiness of financial institutions and the financial systems. Bank mergers tend to take place mostly between institutions from large and developed countries; between banks based in countries in close regional proximity; and between banks from countries that share a common cultural background.

Banks are the single largest provider of external finance to small businesses. In lending to small business, banks use a number of different lending technologies to overcome a lack of publicly available financial information. In Chapter 21, Allen Berger discusses small business lending. In particular, he covers how some of the technologies used to make lending decisions to small business have evolved over time from relationship-based models relying on soft information to more sophisticated models based on different combinations of hard and soft information. He also examines the effects of banking industry consolidation and technological progress on the use of the lending technologies and their effects on small business credit. He finds that consolidation and technological progress and their interactions appear to have resulted in banks placing a greater reliance on quantitative information to make lending decisions. This, the author notes, is reflected in greater distances between banks and their small business clients.

Consumer lending is an area of banking activity that attracts substantial political interest. Recent years have seen a substantial growth in consumer lending. In Chapter 22, Thomas Durkin and Gregory Elliehausen examine the key features and risks inherent in consumer lending, highlighting approaches to evaluating credit supply default risk, and the inextricable influence of adverse selection and asymmetric information in the consumer credit process. An interesting discussion of the role of consumer credit scoring is provided, noting that this has helped lower the costs of the credit evaluation process, and in addition has the advantage of providing consistent application processing across all loan applicants, which simplifies the management of lending.

Mortgage lending is an important part of the banking industry. In Chapter 23, Andreas Lehnert notes that deregulation and process and product innovations have allowed banks to separate origination, funding, and servicing functions. This has allowed small scale financial institutions to originate mortgages, and sell them to other financial institutions. These financial institutions can package the mortgages

and sell them to investors. Evidenced by the recent turmoil in the US subprime market, this unbundling process has introduced tensions among borrowers, mortgage funders, investors, and regulators.

Securitization and financial innovation have transformed the financial system. The role of banks has been changed from 'originate and hold' to 'originate, repackage, and sell'. In Chapter 24, David Marqués Ibañez and Martin Scheicher examine the securitization process and highlight the main reasons as to why banks undertake such activity. Particular focus is placed on the main types of instruments (mortgage-backed securities, collateralized debt obligations, credit default swaps) and their valuation. The chapter also examines the impact of the securitization trend on bank credit and the transmission of monetary policy. They argue that banks' incentives and ability to lend are now much more linked to financial market conditions compared with when banks primarily funded their lending via deposits.

Macroeconomic perspectives in banking

Part IV of the book comprises six chapters discussing the interactions between banking firms and the macroeconomy. This part of the book includes a discussion of the determinants of bank failures and crises, and the impact on financial stability, institutional development, and economic growth.

In Chapter 25, Olivier DeBandt, Philipp Hartmann, and José Luis Peydró examine elements of systemic risk in banking. The notion of contagion in interbank markets is examined along with the interactions between banks and asset prices in crisis periods. Recent advances in the role of liquidity for banking system stability are also examined. The authors note that the practical identification of specific contagion cases continues to be a challenge for empirical research. Nevertheless, the ongoing credit crisis presents new opportunities for innovative research in this area.

In Chapter 26, Gerard Caprio and Patrick Honohon chart the frequency and severity of banking crises. They note that the early years of the new millennium saw a drop in the frequency of banking crises both in developing and high-income economies. A combination of factors were attributed to such a decline, including low real interest rates; sound macroeconomic policies; the expansion of deposit insurance schemes; accumulation of official foreign exchange reserves; well-capitalized banking systems; and the expansion of derivatives and securitization. However, the authors point out that the recent events in the US subprime mortgage markets illustrate information problems in the financial system, which lead to investors taking excessive risks on products they do not understand.

In Chapter 27, Charles Calomiris reviews the theory and historical evidence related to the prevalence of bank failure, panics, and contagion. He argues that banking system panics are neither random events nor inherent to the function of

banks or the structure of bank balance sheets, but are caused by temporary confusion about the incidence of shocks within the banking system. Drawing on empirical evidence, the author argues that deposit insurance and other policies intended to prevent instability have become the single greatest source of banking instability.

In Chapter 28, David Humphrey and James McAndrews analyze the use of retail payments systems (cash, checks, debit cards, credit cards, automated clearing houses, giro networks) and wholesale payment systems (secure giro and wire transfer networks) across countries. The costs and benefits of different systems to the banking system are highlighted, and policy avenues are explored with respect to both retail (privacy issues, card interchange fees, etc.) and wholesale payment systems (integration of back-office systems and uses of large value payment systems, systemic risk, etc.).

A large body of theoretical and empirical research now suggests that financial institutions and markets aid long run economic growth. Market frictions in the form of information and transaction costs necessitate the need for financial markets and intermediaries to mobilize savings, allocate resources, exert corporate control, facilitate risk management, and ease the exchange of goods and services (Levine, 1997; and Levine, 2005). The level of financial development depends on a number of factors, including the degree of economic freedom, the protection of property rights, and the origin and quality of the legal system (La Porta, R., Lopez-De-Silanes, F., Shleifer, A., and Vishny, R.W., 1998). In Chapter 29, Asli Demirgüç-Kunt examines the links between financial development and economic growth. Particular focus is placed on the role that government can play in promoting access to financial services leading to higher levels of financial development that eventually encourage economic growth and prosperity. The empirical literature appears to suggest that well-developed financial systems play an independent and causal role in promoting long-run economic growth. Such effects are proportionately greater across poorer segments of the population.

In Chapter 30, Nicola Cetorelli continues by examining the links between financial development and real economic activity, focusing on the specific mechanisms, such as competition, which link bank activity to the real economy. Evidence suggests that bank concentration is inversely related to economic growth due to lower credit availability, although this effect varies across industries. For instance, concentration allows for the development of long-lasting lending relationships and this seems to enhance growth in industries where young firms are more dependent on external finance.

International differences in banking structures and environments

Part V of the handbook focuses on the features of banking systems in different parts of the world. Six chapters highlight the main structural and institutional features of

various systems. The chapters cover banking in the US, the EU15 countries, Transition countries, Latin America, Japan, and the developing nations of Asia. The banking systems in each of these geographic areas have evolved and changed over time. These chapters note that even as the financial services industry has become more globalized, major differences exist within regions and across countries. For example, the US operates a dual banking system where there is competition from efficient capital markets. The European Union has a single bank license and monetary union covering most of the countries. Eastern European banking systems have undergone a transition from central planning to privatization and the domination of foreign banks, and many are now undergoing more change following accession to the European Union. Latin America has undergone post-crisis liberalization programs, leading to a reduced role of the state and high foreign ownership. Japan has seen the virtual collapse and restructuring of its banking system. The developing nations of Asia still have high state bank ownership (China, India, and Pakistan), limited foreign entry in some markets (China and India), and have had to cope with the aftermath of the East Asia financial crisis (Indonesia, Malaysia, Thailand, and others).

Having said this, similar trends are apparent in various systems—namely, a decline in the number of banks particularly in developed economies, consolidation and concentration, the increased role of foreign banks, the broadening of banks into other financial services areas, greater disintermediation, and the ongoing and omnipresent role of regulatory change. However, as we noted earlier, the heterogeneity of different banking systems can be highlighted, among other factors, by the strong performance of UK and US banks (up until the start of 2007) compared to the weak returns posted by Japanese banks; the variation in the role of foreign and state ownership across many emerging banking markets; differences in business activity restrictions imposed on banking activity in various countries; and so on.

The US banking system has undergone dramatic changes in recent years. Following the removal of restrictive regulations pertaining to branching, product, and price competition, there has been a systematic decline in the number of banks via both mergers and acquisitions and bank failures. This has been accompanied by a significant increase in the number of new banks being chartered. Changes in regulation (notably the Gramm-Leach-Bliley Act of 1999) have allowed US banking organizations to establish financial holding companies and engage in the full range of financial services. The move toward a universal banking model has led to changes in balance sheet composition, strategy, and performance. In Chapter 31, Robert DeYoung discusses in detail the evolution of the US banking industry over the past twenty-five years. He examines how deregulation, technological change, and financial innovation have affected industry structure and the strategies banks pursue. He presents persuasive evidence that suggests that small and large banks can coexist in long-run equilibrium by pursuing very different strategies. In such an equilibrium, large banks use advantages afforded by scale to pursue a

transaction-based banking business model, which is reliant on technology and hard information, while small banks maintain a geographically focused strategy to build and maintain long-term lending relationships. Large banks can thus produce high-volume standardized products at low cost, while small banks can produce lower volumes of more tailored products at a higher price.

Banking in European Union 15 countries has experienced marked changes in recent years. In Chapter 32, John Goddard, Philip Molyneux, and John Wilson examine the evolving structural and regulatory features of the industry. The number of banks has fallen substantially as a result of a merger and acquisitions (M&As) wave, and over the last few years, within-country industry concentration has reached historically high levels. Banks have increasingly focused on generating revenues through non-interest sources of income and there has been widespread diversification into areas such as insurance, pensions, mutual funds, and various securities-related areas. The regulatory environment has constantly changed with the European Union, progressing with legislation aimed at removing barriers for the creation of a single European financial services marketplace as well as the implementation of the new Basel II capital adequacy rules (under CAD3). However, despite wide-scale financial integration, there remain substantial differences in the features and performance of banks in different countries.

In Chapter 33, John Bonin, Iftekhar Hasan, and Paul Wachtel examine banking in transition countries. They note that given the centralized planning systems adopted by Soviet Bloc countries, banking in the transition countries is a relatively recent phenomenon. Attempts were made to establish a banking industry amidst the economic chaos following the fall of Communism in the late 1980s and early 1990s. Unsurprisingly, things went wrong and banking crises occurred. In response, banks were privatized and regulations enacted. The authors argue that banking structures in transition countries are now for the most part populated by privately owned (mainly foreign), relatively well-capitalized banks overseen by a set of regulations and supervision.

Extensive deregulation has taken place in the banking systems of Latin America in recent years. In Chapter 34, Fernando Carvalho, Luiz de Paula, and Jonathan Williams assess the extent to which interest rate deregulation, bank privatization, and the removal of barriers to foreign banking led to banking crises in Brazil, Chile, Argentina, and Mexico. The authors note that banking crises have not led to reversals in the financial liberalization process. Instead, most countries in the area have invested in building regulatory and supervisory infrastructures to ensure the future stability of the banking system. In the long run, it is hoped that such investments will yield a lower cost of capital and wider access to finance leading to economic growth.

Japan has a financial system in which the banking system has traditionally played a more important role than the stock market. The banking system is also rather complex with a wide array of different types of private, co-operative, and public

banks, all undertaking a range of banking business. The system has undergone dramatic changes over recent years, mainly as a result of the major financial crisis that started in the early 1990s and then culminated in 1997–8. This was caused to some extent by an asset price bubble in real estate which was amplified by excessive bank lending to the sector. This resulted in the failure of a number of banks and a massive build-up of non-performing loans in the banking system. The perilous state of the banking system in the late 1990s resulted in a wide range of reforms aimed at improving banking and financial sector soundness, as well as restructuring of the banking system. In Chapter 35, Hirofumi Uchida and Gregory Udell outline the segmented nature of the Japanese banking system and discuss many of the aforementioned issues. The authors note that the dependence on banking appears to have diminished in recent years.

In Chapter 36, Leora Klapper, Maria Soledad Martinez Peria, and Bilal Zia examine banking in the developing nations of Asia. The authors examine the significant reforms and structural changes that have taken place in the aftermath of the financial crisis in East Asia in 1997. Reforms have included privatization and allowing greater foreign participation in the banking industry. The authors highlight variations across countries based on the extent to which they have been aggressive in the reform agenda. They note that while countries such as Pakistan and Korea have made real progress in reform, India and China have proceeded more slowly. Overall, reform has been slower than in transition countries and Latin America.

References

Acharya, V. and Richardson, M. (eds.) (2009). *Restoring Financial Stability: How to Repair a Failed System,* New York: John Wiley and Sons.

Avery, R. B. and Samolyk, K. A. (2004). 'Bank Consolidation and The Provision of Banking Services: Small Commercial Loans', *Journal of Financial Services Research*, 25: 291–325.

Baele, L., De Jonghe, O., and Vander Vennet, R. (2007). 'Does the Stock Market Value Bank Diversification?', *Journal of Banking & Finance*, 31: 1999–2023.

Baily, M. N., Elmendorf, D. W., and Litan, R. E. (2008). *The Great Credit Squeeze: How it Happened, How to Prevent Another*, Washington: Brookings Institution.

Bank of International Settlements (1997). *Core Principles for Effective Banking Supervision*, Basel: Bank of International Settlements.

—— (2006). *Core Principles for Effective Banking Supervision*, Basel: Bank of International Settlements.

Berger, A. N. (2003). 'The Economic Effects of Technological Progress: Evidence from the Banking Industry', *Journal of Money, Credit, and Banking*, 35: 141–76.

—— (2007). 'Obstacles to a Global Banking System: "Old Europe" versus "New Europe"', *Journal of Banking & Finance*, 31: 1955–1973.

Berger, A. N., Bonime, S. D., Goldberg, L. G., and White, L. J. (2004). 'The Dynamics of Market Entry: The Effects of Mergers and Acquisitions on De Novo Entry and Small Business Lending in the Banking Industry', *Journal of Business* 77: 797–834.

—— and Bouwman, C. H. S. (2009a). 'Bank Liquidity Creation', *Review of Financial Studies*, 22: 3779–837.

—— —— (2009b). 'Bank Capital, Survival, and Performance around Financial Crises', University of South Carolina Working Paper.

—— DeYoung, R., Flannery, M. J., Lee, D., and Oztekin, O. (2008). 'How Do Large Banking Organizations Manage their Capital Ratios?', *Journal of Financial Services Research*, 34: 123–49.

—— —— Genay, H., Udell, G. F. (2000). 'The Globalization of Financial Institutions: Evidence from Cross-Border Banking Performance', Brookings-Wharton Papers on Financial Services, 3: 23–158.

—— and Dick, A. A. (2007). 'Entry into Banking Markets and the Early-Mover Advantage', *Journal of Money, Credit, and Banking*, 39: 775–807.

—— —— Goldberg, L. G., and White, L. J. (2007). 'Competition from Large, Multimarket Firms and the Performance of Small, Single-Market Firms: Evidence from the Banking Industry', *Journal of Money, Credit, and Banking*, 39: 331–68.

—— Kashyap, A. K., and Scalise, J. M. (1995). 'The Transformation of the U.S. Banking Industry: What a Long, Strange Trip It's Been', *Brookings* Papers on Economic Activity, 2: 55–218.

—— Miller, N. H., Petersen, M. A., Rajan, R. G., and Stein, J. C. (2005). 'Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks', *Journal of Financial Economics*, 76: 237–69.

—— Rosen, R. J., and Udell, G. F. (2007). 'Does Market Size Structure Affect Competition? The Case of Small Business Lending', *Journal of Banking & Finance* 31: 11–33.

—— Saunders, A., Scalise, J. M., Udell, G. F. (1998). 'The Effects of Bank Mergers and Acquisitions on Small Business Lending', *Journal of Financial Economics*, 50: 187–229.

—— and Udell, G. F. (2006). 'A More Complete Conceptual Framework for SME Finance', *Journal of Banking & Finance*, 30: 2945–66.

Bonaccorsi di Patti, E. and Gobbi, G. (2007). 'Winners or Losers? The Effects of Banking Consolidation on Corporate Borrowers', *Journal of Finance*, 62: 669–95.

Boyd, J. and Prescott, E. (1986). 'Financial Intermediary Coalitions', *Journal of Economic Theory*, 38: 211–32.

Brewer, E. and Jagtiani, J. (2007). 'How Much Would Banks be Willing to Pay to Become "Too-Big-to-Fail" and to Capture other Benefits?', Federal Reserve Bank of Kansas City Research Working Paper 07–05.

Buiter, W. (2008). 'Lessons from the Global Credit Crisis for Social Democrats', Background paper for the Dr. J. M. Den Uyllezing lecture, presented 15 December in den Rode Hoed, Amsterdam. Available from http://www.nber.org/~wbuiter/uyl.pdf.

Campa, J. M. and Hernando, I. (2006). 'M&As Performance in the European Financial Industry', *Journal of Banking & Finance*, 30: 3367–92.

Caprio, G., Demirgüç-Kunt, A., and Kane, E. J. (2008). 'The 2007 Meltdown in Structured Securitization Searching for Lessons, Not Scapegoats', World Bank Policy Research Working Paper 4756, October.

Carbó Valverde, S. and Humphrey, D. B. (2004). 'Predicted and Actual Costs from Individual Bank Mergers', *Journal of Economics & Business*, 56: 137–57.

Carletti, E., Hartmann, P., and Ongena, S. (2007). 'The Economic Impact of Merger Control: What is Special about Banking?', European Central Bank Working Paper 786, July.
Carlson, M. and Weinbach, G. C. (2007). 'Profits and Balance Sheet Developments at US Commercial Banks in 2006', *Federal Reserve Bulletin*, 93, July: A38–A71.
Claessens, S., Demirguc-Kunt, A., Huizinga, H. (2001). 'How Does Foreign Entry Affect the Domestic Banking Market?', *Journal of Banking & Finance*, 25: 891–911.
Cornett, M. M., McNutt, J. J., and Tehranian, H. (2006). 'Performance Changes Around Bank Mergers: Revenue Enhancements versus Cost Reductions', *Journal of Money, Credit, and Banking*, 38: 1013–50.
Craig, B. R. and Dinger, V. (2007). 'Bank Mergers and the Dynamics of Deposit Interest Rates', Federal Reserve Bank of Cleveland Working Paper.
Crouhy, M. G., Jarrow, R. A., and Turnbull, M. (2008). 'Insights and Analysis of Current Events: The Subprime Credit Crisis of 2007', *Journal of Derivatives*, 16: 81–110.
De Guevara J. F. and Maudos J. (2007). 'Explanatory Factors of Market Power in the Banking System', *Manchester School*, 75: 275–96.
DeLong, G. and DeYoung, R. (2007). 'Learning by Observing: Information Spillovers in the Execution and Valuation of Commercial Bank M&As', *Journal of Finance*, 62: 181–216.
De Nicolo, G. and Kwast, M. L. (2002). 'Systemic Risk and Financial Consolidation: Are They Related?', International Monetary Fund Working Paper 02/55.
DeYoung, R. (2007). 'Safety, Soundness, and the Evolution of the U.S. Banking Industry', *Federal Reserve Bank of Atlanta Economic Review*, First and Second Quarters, 41–66.
—— Evanoff, D., and Molyneux, P. (forthcoming). 'Mergers and Acquisitions of Financial Institutions: A Review of the Post-2000 Literature', *Journal of Financial Services Research.*
Diamond D. W. (1984). 'Financial Intermediation and Delegated Monitoring', *Review of Economics Studies*, 51: 393–414.
—— and Dybvig, P. (1983). 'Bank Runs, Deposit Insurance, and Liquidity', *Journal of Political Economy* 91: 401–19.
—— and Rajan, R. G. (2001). 'Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking', *Journal of Political Economy*, 109: 287–327.
European Central Bank (2007a). *EU Banking Structures*, October, Frankfurt: European Central Bank.
—— (2007b). *EU Banking Sector Stability*, November, Frankfurt: European Central Bank.
Economist (2008). 'Money for Old Hope: A Special Report on Asset Management', 1 March 2008, 1–20.
Fama, E. F. (1985). 'What's Different about Banks?', *Journal of Monetary Economics* 15: 29–39.
Financial Times (2008). 'Stumbling to Stability', 21 November, 11.
Fiordelisi, F. and Molyneux, P. (2006). *Shareholder Value in Banking*. London: Palgrave Macmillan.
Focarelli, D. and Panetta, F. (2003). 'Are Mergers Beneficial to Consumers? Evidence from the Italian Market for Bank Deposits', *American Economic Review*, 93: 1152–72.
—— —— and Salleo, C. (2002). 'Why Do Banks Merge?', *Journal of Money Credit and Banking*, 34: 1047–66.
Foote, C. L., Gerardi, K., Goeete, L., and Willen, P. S. (2008). 'Subprime Facts: What (We Think) We Know about the Subprime Crisis and What We Don't', Federal Reserve Bank of Boston Public Policy Discussion Paper 08–2.
Frame, W. S. and White, L. J. (2004). 'Empirical Studies of Financial Innovation: Lots of Talk, Little Action?', *Journal of Economic Literature*, 42: 116–44.

Goddard, J., Molyneux, P., and Wilson, J. O. S. (2001). *European Banking: Efficiency, Technology and Growth*. Chichester: John Wiley and Sons.

——— ——— ——— (forthcoming). 'The Financial Crisis in Europe: Evolution, Policy Responses and Lessons for the Future', *Journal of Financial Regulation and Compliance*.

——— ——— ——— and Tavakoli, M. (2007). 'European Banking: An Overview', *Journal of Banking & Finance*, 31: 1911–35.

Hannan, T. H. and Prager, R. (2004). 'The Competitive Implications of Multimarket Bank Branching', *Journal of Banking & Finance*, 28: 1889–1914.

——— and Pilloff, S. J. (2006). 'Acquisition Targets and Motives in the Banking Industry', Federal Reserve Board Finance and Economics Discussion Series 40.

——— and Hanweck, G. A. (2008). 'Recent Trends in the Number and Size of Bank Branches: An Examination of Likely Determinants', Federal Reserve Board Finance and Economics Discussion Series 02.

Hellwig, M. (2008). 'Systemic Risk in the Financial Sector: An Analysis of the Subprime Mortgage Financial Crisis', Max Planck Institute for Research on Collective Goods, Bonn 2008/43.

HM Treasury (2008). 'The Run on the Rock', House of Commons Treasury Committee, HC56-1.

Holmstrom, B. and Tirole, J. (1998). 'Public and Private Supply of Liquidity', *Journal of Political Economy*, 106: 1–40.

Hosono, K., Sakai, K., and Tsuru, K. (2006). 'Consolidation of Cooperative Banks (Shinkin) in Japan: Motives and Consequences', Research Institute of Economy, Trade, and Industry Discussion Paper Series 06-E-034.

Hughes, J., Lang, W., Mester, L., Moon Ch., and Pagano, M. (2003). 'Do Bankers Sacrifice Value to Build Empires? Managerial Incentives, Industry Consolidation, and Financial Performance', *Journal of Banking & Finance*, 27: 417–47.

Humphrey, D., Willesson, M., Bergendahl, G., and Lindblom, T. (2006). 'Benefits from a Changing Payment Technology in European Banking', *Journal of Banking & Finance*, 30: 1631–52.

International Monetary Fund (2008). 'Global Financial Stability Report: Containing Systemic Risk and Restoring Financial Soundness', Washington, DC: International Monetary Fund.

Jeon, Y. and Miller, S. M. (2007). 'Births, Deaths, and Marriages in the U.S. Commercial Banking Industry', *Economic Inquiry*, 45: 324–41.

Jones, K. D. and Critchfield, T. S. (2005). 'Consolidation in the U.S. Banking Industry: Is the Long, Strange Trip About to End?', *Federal Deposit Insurance Corporation Banking Review*, 17: 30–61.

Kashyap, A. K., Rajan, R. G., and Stein, J. C. (2002). 'Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking', *Journal of Finance*, 57: 33–73.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R. W. (1998). 'Law and Finance', *Journal of Political Economy*, 106: 1113–55.

Leland, H. and Pyle, D. E. (1977). 'Information Asymmetries, Financial Structure and Financial Intermediation', *Journal of Finance*, 32: 371–87.

Levine, R. (1997). 'Financial Development and Economic Growth: Views and Agenda', *Journal of Economic Literature*, 35: 688–726.

——— (2005). 'Finance and Growth: Theory and Evidence', in P. Aghion, and S. Durlaff, (eds.), *Handbook of Economic Growth*. Netherlands: Elsevier.

MESTER, L. J. (2007). 'Some Thoughts on the Evolution of the Banking System and the Process of Financial Intermediation', *Federal Reserve Bank of Atlanta Economic Review*, First and Second Quarters, 67–75.

MISHKIN, F. S. (2006). 'How Big A Problem Is Too Big To Fail?', *Journal of Economic Literature*, 44: 988–1004.

MOLYNEUX, P. and SHAMROUK, N. (1999). *Financial Innovation*. Chichester: John Wiley.

PAULSON, A. L. (2009). 'Raising Capital: The Role of Sovereign Wealth Funds', *Chicago Fed Letter*, January.

POZZOLO, A. F. and FOCARELLI, D. (2007). 'Cross-Border M&As in the Financial Sector. Is Banking Different from Insurance?', Bank of Italy Working Paper.

STIROH, K. J. and RUMBLE, A. (2006). 'The Dark Side of Diversification: The Case of U.S. Financial Holding Companies', *Journal of Banking & Finance*, 30: 2131–61.

UCHIDA, H., UDELL, G. F., WATANABE, W. (2007). 'Bank Size and Lending Relationships in Japan', National Bureau of Economic Research Working Paper 13005.

UDELL, G. F. (2009). 'Wall Street, Main Street, and a Credit Crunch: Thoughts on the Current Financial Crisis', *Business Horizons*, 2: 117–25.

VALKANOV, E. and KLEIMEIER S. (2007). 'The Role of Regulatory Capital in International Bank Mergers and Acquisitions', *Research in International Business and Finance*, 21: 50–68.

YELLEN, J. (2008). 'The Mortgage Meltdown, Financial Markets, and the Economy', *Federal Reserve Bank of San Francisco Economic Letter*, 35–6.

YOM, R. (2005). 'Limited-Purpose Banks: Their Specialties, Performance, and Prospects', *Federal Deposit Insurance Corporation Banking Review*, 17: 19–36.

This page intentionally left blank

PART I

THE THEORY OF BANKING

This page intentionally left blank

CHAPTER 2

THE ROLES OF BANKS IN FINANCIAL SYSTEMS

FRANKLIN ALLEN
ELENA CARLETTI[1]

Introduction

Understanding the many roles that banks play in the financial system is one of the fundamental issues in theoretical economics and finance. The crisis that started in the summer of 2007 underlines just how important banks are to the economy. The efficiency of the process through which savings are channeled into productive activities is crucial for growth and general welfare. Banks are one part of this process. Figure 2.1 gives an overview of the functioning of a financial system. Lenders of funds are primarily households and firms. These lenders can supply funds to the ultimate borrowers, who are mainly firms, governments, and households, in two ways. The first is through financial markets, which consist of money markets, bond markets, and equity markets. The second is through banks and

[1] We are grateful to the editors for helpful comments.

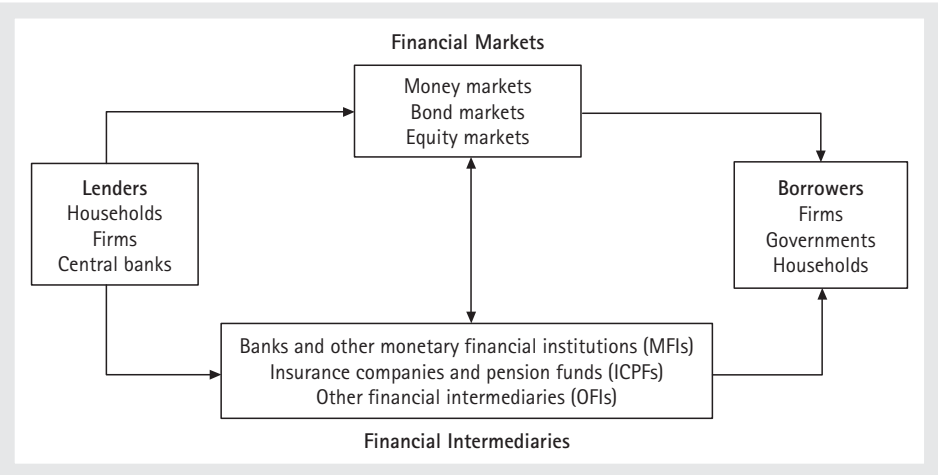

Fig. 2.1. An overview of the financial system

Source: Allen, Chui, and Maddaloni, (2004) p. 491.

other financial intermediaries such as money market funds, mutual funds, insurance companies, and pension funds.

Despite the trend of globalization in recent years, the importance of banks in different economies varies significantly. Figure 2.2 shows a comparison of the long-term financing structure of the Euro area, the UK, the US, Japan, and non-Japan Asia[2] in 1995 and 2003. The figures are given as a percentage of GDP. Bank loans consist of domestic credit to the private sector. The figures in the stock market column are the total market capitalization. The bond market figures are divided into public- and private sector bonds.

It can be seen from Figure 2.2a that in 1995 the euro area had small stock markets but large bank loans and in that sense could be considered as bank-based. However, it also had a significant bond market in terms of both public- and private sector debt. The UK was significantly different, with a large stock market and bank loans but a small bond market, particularly in terms of private sector debt.[3] In some sense it seems to be both market-based and bank-based. The main features of the US financial structure are a small amount of bank loans, a significant stock market, and a much larger bond market than any of the other areas in relative terms. It is the most market-based economy. Japan has significant amounts of finance in all categories. It is very much a bank- and market-based economy. Non-Japan Asia is more similar to the UK: bank loans and the stock market are important but the bond market is not.

[2] This includes Hong Kong, Indonesia, Korea, Malaysia, the Philippines, Singapore, Taiwan, and Thailand.

[3] The UK used to have a significant corporate bond market but this died during the 1970s when inflation was high. It has not revived in recent years despite the reduction in inflation.

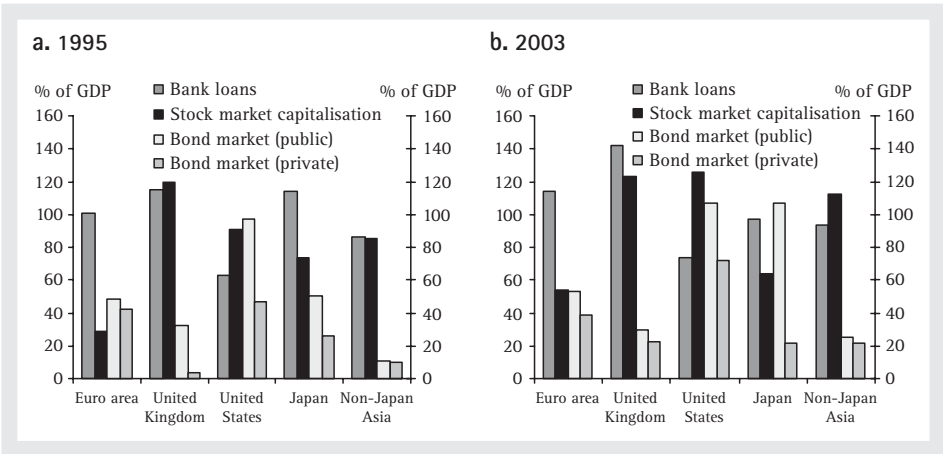

Fig. 2.2. Size of the financial markets by country/region

Source: Allen, Chui, and Maddaloni (2004) p. 492. *Original sources*: CEIC Data Ltd, International Financial Statistics, and national sources.

Figure 2.2b shows the situation in 2003, several years after the Asian crises. It can be seen that the structure is basically the same. The main difference is that Japanese government debt has increased significantly. One interesting feature is that the financial structure in non-Japan Asia has not changed significantly despite the Asian crises.

Figure 2.2 focuses on the claims that are issued by borrowers. Another way of considering the importance of banks is to look at household assets. These are shown in Figure 2.3a. This shows that all the economies are distinctly different. Households in the euro area own significantly fewer financial assets than in the other economies, with a total of 192 percent of GDP compared with 306 percent, 327 percent, and 267 percent for the UK, the US, and Japan, respectively. In terms of the composition of assets there are also large differences. In the euro area, assets held in banks are the most important, insurance and pension funds are next, with direct holdings of shares after that. One striking thing is that household portfolios in the UK are very similar to those in the euro area, with one significant difference: the investment in insurance and pension funds is dramatically higher. This is presumably a result of the difference in public sector pension schemes. In the UK, the basic pension from the state is minimal, while in the euro area, state pensions are usually generous. The US is an outlier in terms of the direct holdings of shares and other equity. Also, households have relatively little in banks. Meanwhile, Japan is an outlier in terms of the amount of assets held in banks where households hold much more in this form than households in other countries. In fact, the Japanese post office bank is the largest deposit taker in the world. Japanese households also have significant amounts in insurance and pension funds. This is to a large extent in insurance companies that offer debt-like contracts. Given the

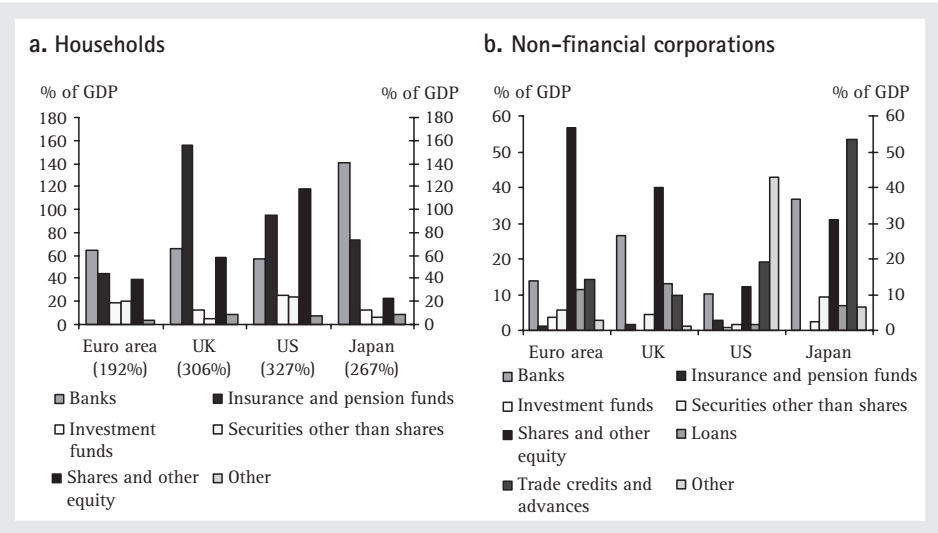

Fig. 2.3. Portfolio allocation (average 1995–2002)

Source: Allen, Chui, and Maddaloni (2004) p. 493. *Original sources*: ECB, Federal Reserve Board and Bank of Japan.

small holdings of shares and other equity, the Japanese bear significantly less financial risk than the households in the US and UK. The US has somewhat less intermediation than the other economies, although the total amount of intermediation is significant in all economies.

Figure 2.3b shows the assets of non-financial corporations. These again underline significant differences across the economies. The euro area and the UK are quite similar except for the amount of shares and other equity held and the amount of trade credits. These are both larger in the euro area than in the UK. The US has much less investment than the other countries except for the 'other' category. This includes holdings of other assets, which are not identified explicitly in the flow of funds data.[4] Japan is perhaps the most different. It has significantly more assets in banks and more trade credit than other countries.

The implication of Figures 2.2 and 2.3 is that the importance of banks and their roles are significantly different in different economies. We start by considering the basic rationales for the existence of banks. The following sections consider the monitoring role of banks and their risk sharing role. The bearing of risks by banks can have important implications for financial stability. The next sections consider banking crises and the contagion between banks that can occur in a crisis. Then we

[4] The column representing 'other' assets is 'unidentified miscellaneous' assets. It is a residual item, arising after accounting for all asset or liability items reported by classified flow of funds' sectors. (In other words, accounting items that do not represent claims on another party are all classified as 'other'.) One example would be the accounting value of goodwill after M&A activities.

consider the role of banks in spurring growth, the corporate governance role of banks, and relationship banking.

Delegated monitoring and banks

An argument that is often put forward in favor of bank-based systems is that banks allow various informational problems to be solved. One important problem is if borrowers must take some action to make proper use of the funds they have borrowed. This action could be the level of effort or choice of project from among various different risky alternatives. The borrower can always claim that a low outcome is due to bad luck rather than from not taking the correct action. Lenders cannot observe the borrower's action unless they pay a fixed cost to monitor the borrower. In a financial market with many lenders, there is a free-rider problem. Each lender is small, so it is not worth paying the fixed cost. Everybody would like to free-ride, leaving it to someone else to bear the monitoring cost. As a result, no monitoring will be done.

A possible solution is to hire a single monitor to check what the borrower is doing. The problem then becomes one of monitoring the monitor, to make sure that she actually monitors the borrowers. Diamond (1984) develops a model of delegated monitoring to solve this problem. Intermediaries have a diversified portfolio of projects for which they provide finance. They pre-commit to monitor borrowers by promising lenders a fixed return. If the intermediary does not monitor, then it will be unable to pay the promised return to lenders. Diamond's model thus illustrates how banks have an incentive to act as a delegated monitor and produce the information necessary for an efficient allocation of resources.

Boot and Thakor (1997) develop a model of financial system architecture that builds on this view of banks as delegated monitors. They assume there are three types of information problem. The first is that there is incomplete information about the future projects a firm has available to it. Outside investors can gather information about these possibilities. The second problem is that lenders cannot observe whether borrowers invest the funds in a risky or safe project. The third problem is the likelihood that borrowers will have the opportunity to invest in a risky project. Boot and Thakor are able to show that the first problem can best be solved by a financial market and the second and third problems can best be solved by intermediaries. They argue that banks will predominate in an emerging financial system, while the informational advantages of markets may allow them to develop in a mature financial system.

The risk sharing role of banks

One of the most important functions of the financial system is to share risk and it is often argued that financial markets are well suited to achieve this aim. As shown in Figure 2.3 and discussed in the Introduction, if both direct holdings of equities and indirect holdings in insurance companies and mutual funds are taken account of, a large amount of household assets is held in equity and only a small amount in banks in the US and UK. In both countries households are exposed to substantial amounts of risk through their holdings of equities. At the other extreme, households in Japan are shielded from risk because they ultimately hold a majority of their assets in banks and very little in equities. Although not as safe as in Japan, households' asset holdings in the euro area are much safer than in the US and UK.

Although the proportions of risky assets held by households in the US and UK are much higher than in Japan and the euro area, this does not necessarily mean that the absolute amount of risk borne by households is greater because the amount invested in financial assets could be higher in the latter countries. However, it can be seen from Figure 2.2 that the euro area has a significantly lower amount of financial assets relative to GDP. Thus, taking into account the amount of wealth held in financial assets increases the differences in the amount of risk borne by households in the different countries, rather than reducing it. Not only do households hold much higher proportions in risky securities in the US and UK, they also hold more financial assets.

How can one explain these differences in the amount of risk households are apparently exposed to in different financial systems? Standard financial theory suggests that the main purpose of financial markets is to improve risk sharing. Financial markets in the US and UK are more developed by most measures than in Japan and the euro area. How can it be that households are exposed to much more risk in the US and UK than in Japan and the euro area?

Allen and Gale (1997; and 2000a: chap. 6) have provided a resolution to this paradox. They point out that traditional financial theory has little to say about hedging non-diversifiable risks. It assumes that the set of assets is given and focuses on the efficient sharing of these risks through exchange. For example, the standard diversification argument requires individuals to exchange assets so that each investor holds a relatively small amount of any one risk. Risks will also be traded so that more risk-averse people bear less risk than people who are more risk-tolerant. This kind of risk sharing is termed cross-sectional risk sharing, because it is achieved through exchanges of risk among individuals at a given point in time. However, importantly, these strategies do not eliminate macroeconomic shocks that affect all assets in a similar way.

Departing from the traditional approach, Allen and Gale focus on the intertemporal smoothing of risks that cannot be diversified at a given point in time.

They argue that such risks can be averaged over time in a way that reduces their impact on individual welfare through intertemporal smoothing by banks. This involves banks building up reserves when the returns on the banks' assets are high and running them down when they are low. The banks can thus pay a relatively constant amount each period and do not impose very much risk on depositors. The authors show that the incentives for engaging in intertemporal smoothing are very different in market-based financial systems. Incomplete financial markets, on the one hand, may not allow effective intertemporal smoothing. The problem is that the long-lived asset 'crowds out' the storage technology because it can be bought and sold for the same price and in addition it pays a dividend. Long-lived banks, on the other hand, can achieve intertemporal smoothing as explained above. However, for this result to hold it is necessary that the banks are not subject to substantial competition from financial markets. In fact, competition from financial markets can lead to disintermediation and the unraveling of intertemporal smoothing provided by long-lived institutions.

Banking crises

Banks perform an important role in terms of maturity transformation. They collect demandable deposits and raise funds in the short-term capital markets and invest them in long-term assets. This maturity mismatch allows them to offer risk sharing to depositors but also exposes them to the possibility that all depositors withdraw their money early. Runs can involve the withdrawal of funds by depositors (retail runs) or the drying up of liquidity in the short-term capital markets (wholesale runs). In the case of the run on Northern Rock in the UK in late 2007, both occurred. These runs can originate in two ways. They can either occur spontaneously as a panic resulting from 'mob psychology' or 'mass hysteria' (e.g., Kindleberger, 1978) or they may arise from fundamental causes that are part of the business cycle (see, e.g., Mitchell, 1941).

The 'panics' view suggests that crises are random events, unrelated to changes in the real economy. The seminal papers developed by Bryant (1980) and Diamond and Dybvig (1983) show bank runs are self-fulfilling prophecies. Given the assumption of first-come, first-served and costly liquidation of some assets, there are multiple equilibria. If everybody believes no panic will occur only those with genuine liquidity needs will withdraw their funds and these demands can be met without costly liquidation of assets. However, if everybody believes a crisis will occur then it becomes a self-fulfilling prophecy as people rush to avoid being last in line. Which of these two equilibria occurs depends on extraneous variables or

'sunspots'. Although sunspots have no effect on the real data of the economy, they affect depositors' beliefs in a way that turns out to be self-fulfilling.

The key issue in theories of panics is which equilibrium is selected and in particular what is the equilibrium selection mechanism. Sunspots are convenient pedagogically but this explanation does not have much content. It does not explain why the sunspot should be used as a coordination device. There is no real account of what triggers a crisis. This is particularly a problem if there is a desire to use the theory for policy analysis.

Carlsson and van Damme (1993) showed how the introduction of a small amount of asymmetric information could eliminate the multiplicity of equilibria in coordination games. They called the games with asymmetric information about fundamentals 'global games'. Their work showed that the existence of multiple equilibria depends on the players having common knowledge about the fundamentals of the game. Introducing noise ensures that the fundamentals are no longer common knowledge and thus prevents the coordination that is essential to multiplicity. Morris and Shin (1998) applied this approach to models of currency crises. Rochet and Vives (2004) and Goldstein and Pauzner (2005) have applied the same technique to banking crises.

Using a global games approach to ensure the uniqueness of equilibrium is theoretically appealing. However, what is really needed in addition to logical consistency is empirical evidence that such an approach is valid. In an important recent contribution, Chen, Goldstein, and Jiang (2007) develop a global games model of mutual fund withdrawals. Using a detailed data set they find evidence consistent with their model. This represents significant evidence supporting the global games approach.

An alternative to the sunspot view is that banking crises are a natural outgrowth of the business cycle. An economic downturn will reduce the value of bank assets, raising the possibility that banks are unable to meet their commitments. If depositors receive information about an impending downturn in the cycle, they will anticipate financial difficulties in the banking sector and try to withdraw their funds, as in Jacklin and Bhattacharya (1988). This attempt will precipitate the crisis. According to this interpretation, crises are not random events but a response of depositors to the arrival of sufficiently negative information on the unfolding economic circumstances. This view is consistent with the evidence in Gorton (1988) that in the US in the late nineteenth and early twentieth centuries, a leading economic indicator based on the liabilities of failed businesses could accurately predict the occurrence of banking crises.

An extensive number of authors have developed models of banking crises caused by aggregate risk. For example, Chari and Jagannathan (1988) focus on a signal extraction problem where part of the population observes a signal about future returns. Others must then try to deduce from observed withdrawals whether an unfavorable signal was received by this group or whether liquidity needs happen to

be high. Chari and Jagannathan are able to show that crises occur not only when the outlook is poor but also when liquidity needs turn out to be high.

Building on the empirical work of Gorton (1988) that nineteenth-century banking crises were predicted by leading economic indicators, Allen and Gale (1998) develop a model that is consistent with the business cycle view of the origins of banking crises. They assume that depositors can observe a leading economic indicator that provides public information about future bank asset returns. If there are high returns then depositors are quite willing to keep their funds in the bank. However, if the returns are sufficiently low they will withdraw their money in anticipation of low returns. There is thus a crisis.

Allen and Gale (2004b) develop a general equilibrium framework for understanding the normative aspects of crises. This framework is used to investigate the welfare properties of financial systems and to discover conditions under which regulation might improve the allocation of resources. An interesting feature of the Allen–Gale framework is that it explicitly models the interaction of banks and markets. Financial institutions are the main players in financial markets, which allow banks and intermediaries to share risks and liquidity. Individuals do not have direct access to markets; instead, they access markets indirectly by investing in intermediaries. Financial intermediaries and markets play important but distinct roles in the model. Intermediaries provide consumers with insurance against idiosyncratic liquidity shocks. Markets allow financial intermediaries and their depositors to share risks from aggregate liquidity and asset return shocks.

Financial markets are said to be complete if it is possible for intermediaries to hedge all aggregate risks in the financial markets. This would be possible if securities contingent on all the possible combinations of aggregate liquidity and asset return shocks, or in other words all the states of nature, were available. Similarly, the risk sharing contracts between intermediaries and consumers are said to be complete if the payoffs can be explicitly conditioned on all the possible combinations of aggregate liquidity and asset return shocks. An example of an incomplete contract would be something like debt, where the payoff on the contract does not depend explicitly on the aggregate state of liquidity demand and asset returns. Allen and Gale (2004b) show that the laissez-faire allocation of resources is efficient provided markets are complete. This is the case even if contracts are incomplete. However, crises are inefficient if markets are incomplete. In this case financial fragility and contagion can occur.

The crisis that started in 2007 provides a dramatic example of how damaging banking crises can be. The causes for its occurrence are not fully understood yet, but many attribute them to the bad incentives in the origination of mortgages and their securitization, the provision of ratings for securitizations, and the risk management systems of investment firms. The large global impact of the crisis suggests, however, that the problems with subprime mortgages are a symptom rather than

the cause. One main problem is that there was a bubble, first in stock prices and then in property prices, and the economic system is now suffering the fallout from the collapse of that bubble. The monetary policies of central banks, particularly the US Federal Reserve, appear to have been too loose and have focused far too much on consumer-price inflation while ignoring asset price inflation. Moreover, the Asian crisis of 1997 and the policies of the IMF during that crisis led to a desire among Asian governments to hoard funds. This created important global imbalances that expanded the credit available and helped to fuel the bubble. Allen and Gale (2000c) show how such an expansion of credit can create a bubble.

Whatever are the reasons behind the crisis, its effects have now certainly spread to the real economy. Most industrialized and non-industrialized countries are experiencing problems with many of their industries entering into recession. The problems are multiple. On the one hand, the difficulties of the financial sectors induce intermediaries to tighten their credit standards thus making it more difficult for firms to obtain credit and at good rates. On the other hand, the sharp fall in consumer demand decreases sales and future orders. As in the financial sectors, the problems are not confined to single firms but affect whole industries. The car industry is one dramatic example, but other manufacturing industries, construction, and many more are very much under pressure.

Banks and contagion

The prevalence of financial crises has led many to conclude that the financial sector is unusually susceptible to shocks. One theory is that small shocks can have a large impact. A shock that initially affects only a particular region or sector or perhaps even a few institutions can spread by contagion through interlinkages between banks and financial institutions to the rest of the financial sector and then infect the larger economy.

The theoretical literature on contagion takes two approaches. On the one hand, there is a number of papers that look for contagious effects via direct linkages. Allen and Gale (2000) study how the banking system responds to contagion when banks are connected under different network structures. In a setting where consumers have the Diamond and Dybvig (1983) type of liquidity preferences, banks perfectly insure against liquidity shocks by exchanging interbank deposits. The connections created by swapping deposits expose the system to contagion. The authors show that incomplete networks are more prone to contagion than complete structures. Better-connected networks are more resilient to contagion since

the proportion of the losses in one bank's portfolio is transferred to more banks through interbank agreements.

Other models capture well the network externalities created from an individual bank risk. Freixas, Parigi, and Rochet (2000) consider the case of banks that face liquidity needs as consumers are uncertain about where they are to consume. In their model, the connections between banks are realized through interbank credit lines that enable these institutions to hedge regional liquidity shocks. In the same way as in Allen and Gale (2000), interbank connections enhance the resilience of the system to the insolvency of a particular bank. The drawback is that this weakens the incentives to close inefficient banks. Moreover, the authors find that the stability of the banking system depends crucially on whether many depositors choose to consume at the location of a bank that functions as a money center or not.

Dasgupta (2004) uses a global games approach to show how a unique equilibrium with contagion can arise when banks hold cross deposits. In the same spirit, Brusco and Castiglionesi (2007) show that there is a positive probability of bankruptcy and propagation of a crisis across regions when banks keep interbank deposits and may engage in excessive risk taking if they are insufficiently capitalized.

Recent contributions have linked the risk of contagion to financial innovation and the accounting system in use. The common feature in this analysis is the presence of incomplete markets where liquidity provision is achieved by selling assets in the market when required. Asset prices are determined by the available liquidity or, said differently, by the 'cash in the market'. It is necessary that people hold liquidity and stand ready to buy assets when they are sold. These suppliers of liquidity are no longer compensated for their opportunity cost of providing liquidity state by state. The cost must be made up on average across all states. This implies volatility in the asset prices that can in turn lead to costly and inefficient crises. In order for people to be willing to supply liquidity they must be able to make a profit in some states. In equilibrium, prices of assets will be such that the profit in the states where banks and intermediaries sell assets is sufficient to compensate the providers of liquidity for all the other states where they are not called upon to provide liquidity and simply bear the opportunity cost of holding it. In other words, asset prices are low in the states where banks and intermediaries need liquidity. But, from an efficiency point of view, this is exactly the wrong time for there to be a transfer from the banks and intermediaries who need liquidity to the providers of liquidity. This is because the banks' depositors who need liquidity will already have low income because they have to withdraw early.

Allen and Carletti (2006) rely on cash in the market pricing to show how financial innovation in the form of credit risk transfer can create contagion across sectors and lower welfare relative to the autarky solution. They focus on the

structure of liquidity shocks hitting the banking sector as the main mechanism determining contagion. When banks face a uniform demand for liquidity, they keep a sufficient amount of the short-term asset and do not need to raise additional liquidity in the market. In this case credit risk transfer is beneficial as it improves risk sharing across sectors. Differently, when banks face idiosyncratic liquidity shocks, they invest also in the long risk-free asset and trade it in the market. The presence of credit risk transfer turns out now to be detrimental as it induces a higher need of liquidity in the market and consequently a greater variability in the asset prices. This in turn affects banks' ability to face their liquidity shocks as it implies a severe reduction in the price of the long asset which banks use to hedge their liquidity risk. The banks that are selling the long asset receive a lower amount and may be unable to pay their depositors.

The effect of introducing credit risk transfer depends crucially also on the accounting system in use, be it historical cost or mark-to-market accounting, as shown by Allen and Carletti (2008). The intuition is similar to the one in the previous chapter. When banks need to liquidate a long-term asset on an illiquid market, it may not be desirable to value such assets according to market values as it reflects the price volatility needed to induce liquidity provision.

The second approach to modeling contagion focuses on indirect balance sheet linkages. Lagunoff and Schreft (2001) construct a model where agents are linked in the sense that the return on an agent's portfolio depends on the portfolio allocations of other agents. In their model, agents who are subject to shocks reallocate their portfolios, thus breaking some linkages. Two related types of financial crisis can occur in response. One occurs gradually as losses spread, breaking more links. The other type occurs instantaneously when forward-looking agents preemptively shift to safer portfolios to avoid future losses from contagion. Similarly, de Vries (2005) shows that there is dependency between banks' portfolios, given the fat tail property of the underlying assets, and this carries the potential for systemic breakdown. Cifuentes, Ferrucci, and Shin (2005) present a model where financial institutions are connected via portfolio holdings. The network is complete as everyone holds the same asset. Although the authors incorporate in their model direct linkages through mutual credit exposures as well, contagion is mainly driven by changes in asset prices.

Complementary to the literature on network effects, Babus (2007) considers a model where banks form links with each other in order to reduce the risk of contagion. The network is formed endogenously and serves as an insurance mechanism. At the base of the link-formation process lies the same intuition developed in Allen and Gale (2000): better connected networks are more resilient to contagion. The model predicts a connectivity threshold above which contagion does not occur, and banks form links to reach this threshold. However, an implicit cost associated to being involved in a link prevents banks from forming connections more than required by the connectivity threshold. Banks manage to form

networks where contagion rarely occurs. Castiglionesi and Navarro (2007) are also interested in whether banks manage to decentralize the network structure a social planner finds optimal. In a setting where banks invest on behalf of depositors and there are positive network externalities on the investment returns, fragility arises when banks that are not sufficiently capitalized gamble with depositors' money. When the probability of bankruptcy is low, the decentralized solution approximates the first best.

Besides the theoretical investigations, there has been a substantial interest in looking for evidence of contagious failures of financial institutions resulting from the mutual claims they have on one another. Most of these papers use balance sheet information to estimate bilateral credit relationships for different banking systems. Subsequently, the stability of the interbank market is tested by simulating the breakdown of a single bank. For example, Upper and Worms (2004) analyze the German banking system. They show that the failure of a single bank could lead to the breakdown of up to 15 percent of the banking sector in terms of assets. Cocco, Gomes, and Martins (2005) consider Portugal; Furfine (2003) the US; Boss, et al. (2004) Austria; and Degryse and Nguyen (2007) Belgium. Iyer and Peydró-Alcalde (2006) conduct a case study of interbank linkages resulting from a large bank failure due to fraud. Upper (2006) contains a survey of this literature. The main conclusion of the literature is that contagion is usually not a serious risk provided there are not significant price movements in response to the turmoil. If there are, as in Cifuentes, Ferrucci, and Shin (2005) then contagion effects can be significant.

The current crisis illustrates the practical importance of contagion. The usual justification for intervention by central banks and governments to prevent the bankruptcy of systemic financial institutions is that this will prevent contagion. This was the argument used by the Federal Reserve for intervening to ensure Bear Sterns did not go bankrupt in March 2008, for example (see Bernanke, 2008). The bankruptcy of Lehman Brothers a few months later in September 2008, illustrated quite how damaging contagion can be. The process did not work in quite the way envisaged in the academic literature and occurred despite the judgment of the Federal Reserve and Treasure that Lehman should not be saved. The first spillover was to the money market mutual fund sector. Reserve Capital 'broke the buck' as it held a significant amount of paper issued by Lehman. This led to many withdrawals from other money market mutual funds and four days after Lehman announced bankruptcy the government was forced to announce guarantees for the entire sector. After seeing Lehman Brothers collapse, confidence in the creditworthiness of banks and other financial institutions and firms fell significantly and this is when the financial crisis started to spill over into the real economy and had such a damaging effect on it. Going forward, much more research is needed to understand the many channels of contagion in a crisis.

Banks and growth

Another important role of banks is in spurring growth. There has been a debate on the relative effectiveness of banks compared with financial markets in doing this. This debate was originally conducted in the context of German and UK growth in the late nineteenth and early twentieth centuries. Gerschenkron (1962) argued that the bank-based system in Germany allowed a closer relationship between bankers providing the finance and industrial firms than was possible in the market-based system in the UK. Goldsmith (1969) pointed out that although manufacturing industry grew much faster in Germany than the UK in the late nineteenth and early twentieth centuries the overall growth rates were fairly similar. More recently, Levine (2002) uses a broad database covering forty-eight countries over the period 1980–95. He finds that the distinction between bank-based and market-based systems is not an interesting one for explaining the finance-growth nexus. Rather, elements of a country's legal environment and the quality of its financial services are most important for fostering general economic growth. In contrast, in a study of thirty-six countries from 1980 to 1995, Tadesse (2002) does find a difference between bank-based and market-based financial systems. For underdeveloped financial sectors, bank-based systems outperform market-based systems, while for developed financial sectors, market-based systems outperform bank-based systems. Levine and Zervos (1998) show that higher stock market liquidity or greater bank development lead to higher growth, irrespective of the development of the other. There is some evidence that financial markets and banks are complements rather than substitutes. Demirgüç-Kunt and Maksimovic (1998) show that more-developed stock markets tend to be associated with increased use of bank finance in developing countries.

There is a large theoretical literature on the relative merits of bank-based and market-based systems for innovation and growth. Bhattacharya and Chiesa (1995) consider a model of R&D incentives and financing. In a market system, lenders learn the value of each firm's R&D at the interim stage after R&D has been undertaken but before production takes place. The lenders can share the information among the firms and will do so if it is in their interest. Bhattacharya and Chiesa show that their incentives to do this correspond to maximizing the aggregate value of the firms' R&D projects. Also, a collusive agreement can be structured so that only one firm actually produces at the production stage. However, this collusion creates a free-rider problem and reduces incentives to undertake the R&D at the first stage. If this incentive problem is severe enough, bilateral financing may be preferable. Under this arrangement, each firm is financed by one bank and there is no scope for information sharing. As a result, each firm's R&D information remains proprietary.

Allen and Gale (1999; and 2000a: chap. 13) ask whether financial markets or banks are better at providing finance for projects where there is diversity of opinion as in the development of new technologies. Diversity of opinion arises from differences in

prior beliefs, rather than differences in information. The advantage of financial markets is that they allow people with similar views to join together to finance projects. This will be optimal provided the costs necessary for each investor to form an opinion before investment decisions are made are sufficiently low. Finance can be provided by the market even when there is great diversity of opinion among investors. Intermediated finance involves delegating the financing decision to a manager who expends the cost necessary to form an opinion. There is an agency problem in that the manager may not have the same prior as the investor. This type of delegation turns out to be optimal when the costs of forming an opinion are high and there is likely to be considerable agreement in any case. The analysis suggests that market-based systems will lead to more innovation than bank-based systems.

The corporate governance role of banks

The importance of equity ownership by financial institutions in Japan and Germany, and the lack of a strong market for corporate control in these countries have led to the suggestion that the agency problem in these countries is solved by banks acting as outside monitors for large corporations. In Japan, this system of monitoring is known as the main bank system. The characteristics of this system are the long-term relationship between a bank and its client firm, the holding of both debt and equity by the bank, and the active intervention of the bank should its client become financially distressed. It has been widely argued that this main bank relationship ensures that the bank acts as delegated monitor and helps to overcome the agency problem between managers and the firm. However, the empirical evidence on the effectiveness of the main bank system is mixed (see, e.g., Hoshi, Kashyap, and Scharfstein, 1990; Hoshi, Kashyap, and Scharfstein, 1993; Aoki and Patrick, 1994; and Hayashi, 2000). Overall, the main bank system appears important in times of financial distress, but less important when a firm is doing well.

In Germany, the counterpart of the main bank system is the hausbank system. Banks tend to have very close ties with industry and form long-run relationships with firms not only because of the loans they make and the shares they directly own but also because of the proxies they are able to exercise. A number of studies have provided evidence on the effectiveness of the outside monitoring of German banks (see, e.g., Gorton and Schmid, 2000).

In an important book, Edwards and Fischer (1994) have argued that in Germany the corporate governance role of banks has been overemphasized in the literature. They provide a variety of evidence that banks do not have the degree of influence as

lenders, shareholders, or voters of proxies that is usually supposed. For example, they find that the number of votes controlled in a company is only weakly related to the number of representatives the bank has on the supervisory board. Hellwig (1991; 1994) also provides a number of theoretical arguments concerning the disadvantages of the banking system in Germany.

Relationship banking

There is a growing literature that analyzes the advantages and disadvantages of relationships in banking (see, for reviews, Boot, 2000; Gorton and Winton, 2003; and Degryse and Ongena, 2008). If, on the one hand, close and durable relationships provide better access to firms and ameliorate some of the information problems characterizing lending relationships, on the other hand, they also involve inefficiencies related to the hold-up and the soft-budget-constraint problems. The hold-up problem refers to the possibility that a relationship bank uses the superior private information it possesses about the firm to extract rents, thus distorting entrepreneurial incentives and causing inefficient investment choices (Sharpe, 1990; Rajan, 1992; and von Thadden, 1995). The soft-budget-constraint problem concerns the inability of a relationship lender to commit itself to a particular course of action in advance. Although it is optimal to threaten to terminate the availability of credit in advance, once the borrower has defaulted the first loan becomes a 'sunk cost'. If the firm has another good project we should expect that the lender will continue to extend credit, even if the borrower defaults. Renegotiation thus creates a time-consistency problem. The threat to terminate credit creates good incentives for the borrower to avoid the risk of default. Termination of credit is not Pareto-efficient *ex post*, but the incentive effect makes both parties better off. However, if the borrower anticipates that the lender will not carry out the threat in practice, the incentive effect disappears. Although the lender's behavior is now *ex post* optimal, both parties may be worse off *ex ante*.

Multiple bank relationships can help mitigating the drawbacks of single-bank relationships in terms of the hold-up and the soft-budget-constraint problems. As for the former, borrowing from multiple banks can restore competition among banks and, consequently, improve entrepreneurial incentives (Padilla and Pagano, 1997). As for the latter, Dewatripont and Maskin (1995) argue that by complicating the refinancing process and making it less profitable multiple bank lending allows banks to commit not to extend further inefficient credit. Similarly, Bolton and Scharfstein (1996) show that multiple bank lending reduces entrepreneurial incentives to default strategically because it complicates debt renegotiation.

The number of bank relationships also has important implications for banks' role as monitors. In a context where both firms and banks are subject to moral hazard problems, Carletti (2004) analyzes how the number of bank relationships influences banks' monitoring incentives, the level of loan rates, and a firm's choice between single and multiple bank relationships. Multiple bank lending suffers from duplication of effort and free-riding but it benefits from diseconomies of scale in monitoring, thus involving a lower level of monitoring but not necessarily higher loan rates than single lending. Since banks choose their monitoring effort to maximize their expected profits, they may choose a level of monitoring which is excessive from the firms' perspective. When this is the case, the firm may choose multiple bank relationships in order to reduce the overall level of monitoring. The attractiveness of such a choice increases with the cost of monitoring, the firm's private benefit, and expected profitability. In a similar framework, Carletti, Cerasi, and Daltung (2007) analyzes the circumstances where banks with limited diversification opportunities find it profitable to enter into multiple bank relationships. They show that sharing lending allows banks to diversify better their portfolios but still entails duplication of effort and free-riding. When the benefit of greater diversification dominates, multiple bank lending leads to higher overall monitoring as a way to mitigate the agency problem between banks and depositors and achieve higher banks' expected profits. The attractiveness of multiple bank lending now decreases with the level of banks' (inside) equity and firms' prior profitability, while it increases with the cost of monitoring.

Other rationales for multiple bank relationships relate to firms' desire to reduce liquidity risk and disclose information through credit relationships. Detragiache, Garella, and Guiso (2000) show that, when relationship banks face internal liquidity problems, borrowing from multiple banks can avoid early liquidation of profitable projects. Yosha (1995) suggests that firms may prefer multiple bank lending as a way to disclose confidential information about the quality of their projects and to avoid aggressive behavior by competitors.

As a final remark, note that there are ways other than multiple bank relationships to solve the problem of lack of commitment affecting exclusive bank relationships. For example, financial institutions may develop a valuable reputation for maintaining commitments. In any one case, it is worth incurring the small cost of a sub-optimal action in order to maintain the value of the reputation. Incomplete information about the borrower's type may lead to a similar outcome. If default causes the institution to believe it is more likely that the defaulter is a bad type, then it may be optimal to refuse to deal with a firm after it has defaulted. Institutional strategies such as delegating decisions to agents who are given no discretion to renegotiate may also be an effective commitment device. Several authors (Huberman and Kahn, 1988; Hart and Moore, 1988; Gale, 1991; and Allen and Gale, 2000a: chap. 10) have argued that, under certain circumstances, renegotiation is welfare-improving. In that case, the argument is reversed. Intermediaries

that establish long-term relationships with clients may have an advantage over financial markets precisely because it is easier for them to renegotiate contracts.

Concluding remarks

We have covered a number of roles of banks in the financial system in this chapter. Banks act as delegated monitors and ensure that firms use the resources allocated to them effectively. They also play an important role in sharing risk in the economy by diversifying and smoothing fluctuations over time. These are positive aspects of the roles banks play. However, the fixed nature of the claims they issue can cause fragility in the financial system. Banks are often at the center of financial crises as in the crisis that started in the summer of 2007. They can help spread crises if there is contagion and small shocks can have a large effect on the financial system and the economy. Banks play an important role in providing funds for firms and helping them and the economy to grow. They are also important for corporate governance, particularly in countries like Germany where bankers sit on boards and control a significant number of proxy votes. Finally, banks can help overcome asymmetric information problems by forming long-lived relationships with firms.

There a number of other roles that we have not covered as they are the subjects of other chapters of the book. These include the role of banks in underwriting securities, covered in Chapter 7 and the role of banks in payments systems, covered in Chapter 28. There remain other roles that are important that are less well understood. Many of these involve the interaction of banks with financial markets of various kinds. The recent crisis has illustrated that securitization can lead to significant problems because bank incentives are fundamentally different when loans are sold rather than retained. The role that banks play in derivative markets is also not fully understood. If there is a chain of counterparties how can that risk be fully assessed if the chain is opaque as it usually is? Finally, how can banks be prevented from taking risks if they retain the profits when there are good outcomes but are bailed out by the government in times of crisis? These are all important issues for future research.

References

Allen, F. and E. Carletti (2006). 'Credit Risk Transfer and Contagion', *Journal of Monetary Economics*, 53: 89–111.

——— ——— (2008). 'Mark-to-Market Accounting and Liquidity Pricing', *Journal of Accounting and Economics*, 45: 358–78.

——M. Chui, and A. Maddaloni (2004). 'Financial Systems in Europe, the USA, and Asia', *Oxford Review of Economic Policy,* 20: 490–508.
——and D. Gale (1997). 'Financial Markets, Intermediaries, and Intertemporal Smoothing', *Journal of Political Economy,* 105: 523–46.
————(1998). 'Optimal Financial Crises', *Journal of Finance,* 53: 1245–84.
————(1999). 'Diversity of Opinion and the Financing of New Technologies', *Journal of Financial Intermediation,* 8: 68–89.
————(2000a). *Comparing Financial Systems*, Cambridge, MA: MIT Press.
————(2000b). 'Financial Contagion', *Journal of Political Economy,* 108: 1–33.
————(2000c). 'Bubbles and Crises', *Economic Journal,* 110: 236–55.
————(2004). 'Financial Intermediaries and Markets', *Econometrica,* 72: 1023–61.
Aoki M. and H. Patrick (eds.) (1994). *The Japanese Main Bank System: Its Relevancy for Developing and Transforming Economies*, New York: Oxford University Press, 592–633.
Babus, A. (2007). 'The Formation of Financial Networks', Discussion Paper 06–093, Tinbergen Institute.
Bernanke, B. (2008). 'Reducing Systemic Risk', Opening Speech at the 2008 Jackson Hole Symposium, Federal Reserve Bank of Kansas City.
Bhattacharya, S. and G. Chiesa (1995). 'Financial Intermediation with Proprietary Information', *Journal of Financial Intermediation,* 4: 328–57.
Bolton, P. and D. Scharfstein (1996). 'Optimal Debt Structure and the Number of Creditors', *Journal of Political Economy,* 104: 1–25.
Boot, A. (2000). 'Relationship Banking: What Do We Know?, *Journal of Financial Intermediation,* 9: 7–25.
——and A. Thakor (1997). 'Financial System Architecture', *Review of Financial Studies,* 10: 693–733.
Boss, M., H. Elsinger, S. Thurner, and M. Summer (2004). 'Network Topology of the Interbank Market', *Quantitative Finance,* 4: 1.8.
Brusco S. and F. Castiglionesi (2007). 'Liquidity Coinsurance, Moral Hazard and Financial Contagion', *Journal of Finance,* 62: 2275–302.
Bryant, J. (1980). 'A Model of Reserves, Bank Runs, and Deposit Insurance', *Journal of Banking & Finance,* 4: 335–44.
Carletti, E. (2004). 'The Structure of Bank Relationships, Endogenous Monitoring and Loan Rates', *Journal of Financial Intermediation,* 13: 58–86.
——V. Cerasi, and S. Daltung (2007). 'Multiple Bank Lending: Diversification and Free-Riding in Monitoring', *Journal of Financial Intermediation,* 16: 425–51.
Carlsson, H. and E. van Damme (1993). 'Global Games and Equilibrium Selection', *Econometrica,* 61: 989–1018.
Castiglionesi, F. and N. Navarro (2007). 'Optimal Fragile Financial Networks', Working Paper, Tilburg University.
Chari, V. and R. Jagannathan (1988). 'Banking Panics, Information, and Rational Expectations Equilibrium', *Journal of Finance,* 43: 749–60.
Chen, Q., I. Goldstein, and W. Jiang (2007). 'Payoff Complementarities and Financial Fragility: Evidence from Mutual Fund Flows', Working Paper, University of Pennsylvania.
Cifuentes, R., G. Ferrucci, and H. Shin (2005). 'Liquidity Risk and Contagion', *Journal of European Economic Association,* 3.
Cocco, J., F. Gomes, and N. Martins (2005). 'Lending Relationships in the Interbank Market', Working Paper, London Business School.

Dasgupta, A. (2004). 'Financial Contagion through Capital Connections: A Model of the Origin and Spread of Bank Panics', *Journal of the European Economic Association,* 6: 1049–84.

Degryse, H. and G. Nguyen (2007), 'Interbank Exposures: An Empirical Examination of Systemic Risk in the Belgian Banking System', *International Journal of Central Banking,* 3: 123–71.

—— and S. Ongena (2008), 'Competition and Regulation in the Banking Sector: A Review of the Empirical Evidence on the Sources of Bank Rents', in A. Thakor and A. Boot (eds.), *Handbook of Financial Intermediation and Banking,* Amsterdam: Elsevier.

Demirgüç-Kunt, A. and V. Maksimovic (1998). 'Law, Finance, and Firm Growth', *Journal of Finance,* 53: 2107–37.

Detragiache, E., P. Garella, and L. Guiso (2000). 'Multiple vs. Single Banking Relationships: Theory and Evidence', *Journal of Finance,* 55: 1133–61.

Dewatripont, M. and E. Maskin (1995). 'Credit and Efficiency in Centralized and Decentralized Economies', *Review of Economic Studies,* 62: 541–55.

de Vries, C. (2005). 'The Simple Economics of Bank Fragility', *Journal of Banking & Finance,* 29: 803–25.

Diamond, D. (1984). 'Financial Intermediation and Delegated Monitoring', *Review of Economic Studies,* 51: 393–414.

—— and P. Dybvig (1983). 'Bank Runs, Deposit Insurance, and Liquidity', *Journal of Political Economy,* 91: 401–19.

Edwards, J. and K. Fischer (1994). *Banks, Finance and Investment in Germany*, Cambridge, U.K.: Cambridge University Press.

Freixas, X., B. Parigi, and J. Rochet (2000). 'Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank', *Journal of Money, Credit and Banking,* 32: 611–38.

Furfine, C. (2003). 'The Interbank Market During a Crisis', *Journal of Money, Credit and Banking,* 35: 111–28.

Gale, D. (1991). 'Optimal Risk Sharing through Renegotiation of Simple Contracts', *Journal of Financial Intermediation,* 1: 283–306.

Gerschenkron, A. (1962). *Economic Backwardness in Historical Perspective*, Cambridge, MA: Harvard University Press.

Goldsmith, R. (1969). *Financial Structure and Development*, New Haven, CT: Yale University Press.

Goldstein, I. and A. Pauzner (2005). 'Demand-Deposit Contracts and the Probability of Bank Runs', *Journal of Finance,* 60: 1293–327.

Gorton, G. (1988). 'Banking Panics and Business Cycles', *Oxford Economic Papers,* 40: 751–81.

—— and F. Schmid (2000). 'Universal Banking and the Performance of German Firms', *Journal of Financial Economics,* 58: 29–80.

—— and A. Winton (2003). 'Financial Intermediation', in G. Constantinides, M. Harris, and R. Stulz (eds.), *Handbook of the Economics of Finance*, Volume 1A, Chapter 8, Amsterdam: North Holland, 431–552.

Hart, O. and J. Moore (1988). 'Incomplete Contracts and Renegotiation', *Econometrica,* 56: 755–85.

Hayashi, F. (2000). 'The Main Bank System and Corporate Investment: An Empirical Reassessment', in M. Aoki and G. Saxonhouse (eds.), *Finance, Governance, and Competitiveness in Japan*, Oxford and New York: Oxford University Press, 81–97.

Hellwig, M. (1991). 'Banking, Financial Intermediation and Corporate Finance', in A. Giovannini and C. Mayer (eds.), *European Financial Integration*, New York: Cambridge University Press, 35–63.

——(1994). 'Liquidity Provision, Banking, and the Allocation of Interest Rate Risk', *European Economic Review,* 38: 1363–89.

Hoshi, T., A. Kashyap, and D. Scharfstein (1990). 'The Role of Banks in Reducing the Costs of Financial Distress in Japan', *Journal of Financial Economics,* 27: 67–68.

Kindleberger, C. (1978). *Manias, Panics, and Crashes: A History of Financial Crises*, New York: Basic Books.

Huberman, G. and C. Kahn (1988). 'Limited Contract Enforcement and Strategic Renegotiation', *American Economic Review,* 78: 471–84.

Iyer, R. and J. L. Peydro-Alcalde (2007). 'The Achilles Heel of Interbank Markets: Financial Contagion Due to Interbank Linkages', Working Paper, University of Amsterdam.

Lagunoff, R. and S. Schreft (2001). 'A Model of Financial Fragility', *Journal of Economic Theory,* 99: 220–64.

Levine, R. (2002). 'Bank-Based or Market-Based Financial Systems: Which is Better?', *Journal of Financial Intermediation,* 11: 398–428.

——and S. Zervos (1998). 'Stock Markets, Banks and Economic Growth', *American Economic Review,* 88: 537–58.

Mitchell, W. (1941). *Business Cycles and Their Causes*, Berkeley: University of California Press.

Morris, S. and H. Shin (1998). 'Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks', *American Economic Review,* 88: 587–97.

Padilla, A. J. and M. Pagano (1997). 'Endogenous Communication Among Lenders and Entrepreneurial Incentives', *Review of Financial Studies,* 10: 205–36.

Rajan, R. (1992). 'Insiders and Outsiders: The Choice Between Informed and Arm's-Length Debt', *Journal of Finance,* 47: 1367–400.

Rochet, J. and X. Vives (2004). 'Coordination Failures and the Lender of Last Resort: Was Bagehot Right after All?', *Journal of the European Economic Association,* 2: 1116–47.

Sharpe, S. (1990). 'Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships', *Journal of Finance,* 45: 1069–87.

Tadassee, S. (2002). 'Financial Architecture and Economic Performance: International Evidence', *Journal of Financial Intermediation*, 11: 429–54.

Upper, C. (2006). 'Contagion Due to Interbank Credit Exposures: What Do We Know, Why Do We Know It, and What Should We Know?', Working Paper, Bank for International Settlements.

——and A. Worms (2004). 'Estimating Bilateral Exposures in the German Interbank Market: Is There a Danger of Contagion?', *European Economic Review,* 48: 827–49.

Von Thadden, E. (1995). 'Long-Term Contracts, Short-Term Investment and Monitoring', *Review of Economic Studies,* 62.

Yosha, O. (1995). 'Information Disclosure Costs and the Choice of Financing Source', *Journal of Financial Intermediation,* 4: 3–20.

CHAPTER 3

THE ACCELERATING INTEGRATION OF BANKS AND MARKETS AND ITS IMPLICATIONS FOR REGULATION

ARNOUD W. A. BOOT
ANJAN V. THAKOR

Introduction

The financial sector has evolved rapidly over the last decade, with the impetus for change provided by deregulation and advances in information technology. Competition has become more intense. Interbank competition within domestic markets as well as across national borders and competition from financial markets have gained importance. Both the institutional structure of financial institutions and the

boundary between financial institutions and financial markets have been transformed. This chapter reviews the literature related to these developments and uses it to examine the importance of this changing landscape for the structure of the financial services industry and the design and organization of regulation.

As we will argue, the increasingly intertwined nature of banks and financial markets is not without costs. In particular, as the financial crisis of 2007–9 has illustrated, systemic risks may have become more prevalent. In this chapter, we seek to provide a fundamental analysis of the underlying forces that could explain the evolution of the banking industry. We begin by discussing the key insights from the financial intermediation literature, including the potential complementarities and conflicts of interest between intermediated relationship banking activities and financial market activities (underwriting, securitization, etc.). While debt contracts dominate the financial intermediation literature, the impressive growth of private equity firms has turned the spotlight on equity. In a sense, one could interpret private equity (PE) as intermediation driven from the equity side. Given their economic functions as debt and equity intermediaries, respectively, how do banks and PE firms interact?

Our discussion reveals that the interaction between banks and PE firms is only one aspect of an increasing integration of banks and markets. Banks have a growing dependence on the financial markets not only as funding sources for hedging purposes but also for engaging in various transactions like securitization for their customers. The multiple dimensions of bank dependence on markets generate both risk reduction and risk elevation possibilities for banks. For example, while hedging may reduce risk, proprietary trading, liquidity guarantees for securitized debt, and positions in credit default swaps can increase risk. This raises potential regulatory concerns. What do these developments imply for prudential regulation and supervision? Will the increasing interactions between banks and markets increase or decrease financial system fragility? The financial crisis of 2007–9 suggests an increase in fragility, but how much can we generalize from this crisis? These questions have become particularly germane not only because of growing banks–markets integration, but also due to the growing cross-border footprint of financial institutions.

These developments have also focused attention on the role of 'gatekeepers' (Coffee, 2002), like credit rating agencies. While the financial intermediation literature has acknowledged the role of credit rating agencies as information processors and sellers for some time now (e.g. Allen, 1990 and Ramakrishnan and Thakor, 1984), the literature has not discussed how rating agencies may affect the fragility of the financial sector through the important role they play as 'spiders in the web of institutions and markets'. We take up this issue in our discussion.

The organization of the chapter is as follows. In the next section, we focus on the economic role of financial intermediaries. The primary focus here is on the

banks' role in lending and how this compares to non-intermediated finance directly from the financial market. We will also analyze the effects of competition on the banks' lending relationships. Does competition harm relationships and reduce their value and hence induce more transaction-oriented banking, or does competition augment the value of relationships? This discussion will summarize the key insights from the modern literature of financial intermediation. In the next section we discuss the increasingly interconnected nature of banks and financial markets, with a focus on securitization. This 'technology' has been at the center of the 2007–9 financial crisis. What are the future prospects for securitization? The proliferation of non-banking financial institutions, and particularly private equity firms, is discussed in the following section. We will argue that much of this activity is complementary to the role of banks, rather than threatening their *raison d'être*. Subsequently, we focus on the role of credit rating agencies. These agencies have been indispensable for the explosive growth (and temporary demise) of securitization. How will their role develop? We then discusses regulatory implications. Here we link the role of banks in lending (as emphasized in our earlier discussions) to their role as providers of liquidity. This brings in the issue of fragility which is at the heart of the current regulatory debate.

Understanding banks as information-processing intermediaries

In this section we discuss two issues: (1) what is the key role of banks vis-à-vis markets? and (2) how does competition impinge on this role?

The economic role of banks

We first discuss the role of banks in qualitative asset transformation—i.e., the process by which banks absorb risk to transform both the liquidity and credit risk characteristics of assets (see Bhattacharya and Thakor, 1993). For example, banks invest in risky loans but finance them with riskless deposits (e.g. Diamond, 1984 and Ramakrishnan and Thakor, 1984). They also invest in illiquid loans and finance them with liquid demandable deposits (e.g. Diamond and Dybvig, 1983). The theory of financial intermediation has placed special emphasis on the role of banks in monitoring and screening borrowers in the process of lending. Bank lending is typically contrasted with direct funding from the financial markets.

What are the comparative advantages of bank loans over public capital market bond financing?

The most striking insight of the contemporary theory of financial intermediation is that banks are better than markets at resolving informational problems. The possession of better information about their borrowers allows banks to get closer to their borrowers. Interestingly, a feedback loop is generated as this proximity between the financier and the borrowing firm in bank lending arrangements may also help mitigate the information asymmetries that typically plague arm's length arrangements in market transactions. This has several aspects. A borrower might be prepared to reveal proprietary information to its bank that it may have been reluctant to reveal to the financial markets (Bhattacharya and Chiesa, 1995). A bank might also have better incentives to invest in information acquisition. While costly, the substantial stake that it has in the funding of the borrower and the enduring nature of its relationship with the borrower—with the possibility of information reusability over time—increase the marginal benefit of information acquisition to the bank.[1]

Such borrower–lender proximity may also have a dark side. An important one is the hold-up problem that stems from the information monopoly the bank may develop due to the spontaneous generation of proprietary information on borrowers. Such an informational monopoly may permit the bank to charge higher loan interest rates *ex post* (see Sharpe, 1990; Rajan, 1992; and Boot, 2000, for a review). The threat of being 'locked in', or informationally captured by the bank, may dampen loan demand *ex ante*, causing a loss of potentially valuable investment opportunities. Alternatively, firms may opt for multiple bank relationships (see Carletti, Cerasi, and Daltung, 2007). This may reduce the informational monopoly of any individual bank, but possibly at a cost. Ongena and Smith (2000) show that multiple bank relationships indeed reduce the hold-up problem, but can worsen the availability of credit (see Thakor, 1996 for a theoretical rationale).

Another aspect is that relationship banking could accommodate an intertemporal smoothing of contract terms (see Allen and Gale, 1995 and Allen and Gale, 1997) that would entail losses for the bank in the short term that are recouped later in the relationship. Petersen and Rajan (1995) show that credit subsidies to young or 'de novo' companies may reduce the moral hazard problem and informational frictions that banks face in lending to such borrowers. Banks may be willing to

[1] Ramakrishnan and Thakor (1984) and Millon and Thakor (1985) focus on pre-contract information asymmetries to rationalize the value financial intermediaries add relative to markets. Diamond (1984) focuses on post-contract information asymmetries to rationalize intermediation. Coval and Thakor (2005) show that financial intermediaries can provide an institutional resolution of the problem of cognitive biases at the individual investor level, acting as a 'belief's bridge' between pessimistic investors and optimistic entrepreneurs. James (1987), Lummer and McConnell (1989), and Gande and Saunders (2005) provide empirical evidence on the informational value of bank financing. See also the 'stories' provided by Berlin (1996) supporting the special role of banks.

provide such subsidized funding if they can expect to offset the initial losses through the long-term rents generated by these borrowers. The point is that, without access to *subsidized* credit early in their lives, 'de novo' borrowers would pose such serious adverse selection and moral hazard problems that *no* bank would lend to them. Relationship lending makes these loans feasible because the *proprietary* information generated during the relationship produces 'competition-immune' rents for the bank later in the relationship and permits the early losses to be offset. The importance of intertemporal transfers in loan pricing is also present in Berlin and Mester (1999). They show that rate-insensitive core deposits allow for intertemporal smoothing in lending rates. This suggests a complementarity between deposit taking and lending. Moreover, the loan commitment literature has emphasized the importance of intertemporal tax subsidy schemes in pricing to resolve moral hazard (see Boot, Thakor, and Udell, 1991 and Shockley and Thakor, 1997) and also the complementarity between deposit taking and *commitment* lending (see Kashyap, Rajan, and Stein, 1999).

The bank–borrower relationship also displays greater contractual flexibility than that normally encountered in the financial market. This flexibility inheres in the generation of hard and soft proprietary information during a banking relationship. This information gives the bank the ability to adjust contractual terms to the arrival of new information and hence encourages it to write 'discretionary contracts' *ex ante* that leave room for such *ex post* adjustments. This is in line with the important ongoing discussion in economic theory on rules versus discretion, where discretion allows for decision-making based on more subtle—potentially non-contractible—information. See, for example, Simon (1936), and Boot, Greenbaum, and Thakor (1993). The papers by Stein (2002), and Berger, Miller, Petersen, Rajan and Stein (2005) highlight the value of 'soft information' in lending. This could be an example of this more subtle and non-contractible information. On this issue, two dimensions can be identified. One dimension is related to the nature of the bank–borrower relationship, which is typically long-term, with accompanying reinforcing incentives for both the bank and the borrower to enhance the durability of the relationship. This allows for *implicit*—non-enforceable—long-term contracting. An optimal information flow is crucial for sustaining these 'contracts'. Information asymmetries in the financial market, and the non-contractibility of various pieces of information, would rule out long-term alternative capital market funding sources as well as *explicit* long-term commitments by banks. Therefore, both the bank and the borrower may realize the added value of their relationship, and have an incentive to foster the relationship.[2]

The other dimension is related to the structure of the explicit contracts that banks can write. Because banks write more discretionary contracts, bank loans are

[2] Mayer (1988) and Hellwig (1991) discuss the commitment nature of bank funding. Boot, Thakor, and Udell (1991) address the *credibility* of commitments.

generally easier to renegotiate than bond issues or other public capital market funding vehicles (see Berlin and Mester, 1992). Such renegotiability may be a mixed blessing because banks may suffer from a 'soft-budget constraint' problem: borrowers may realize that they can renegotiate *ex post*, which could give them perverse *ex ante* incentives (see Bolton and Scharfstein, 1996 and Dewatripont and Maskin, 1995). The soft-budget-constraint problem is related to the potential lack of toughness in enforcing contracts due to the *ex post* distribution of 'bargaining power' linked with relationship banking proximity (see Boot, 2000). In practice, one way that banks can deal with this issue is through the priority structure of their loan contracts. If the bank has priority/seniority over other lenders, it could strengthen the bank's bargaining position and allow it to become tougher. These issues are examined in Diamond (1993), Berglöf and von Thadden (1993), and Gorton and Kahn (1993).

The bank could then credibly intervene in the decision process of the borrower when it believes that its long-term interests are in jeopardy. For example, the bank might believe that the firm's strategy is flawed, or a restructuring is long overdue. Could the bank push for the restructuring? If the bank has no priority, the borrower may choose to ignore the bank's wishes. The bank could threaten to call the loan, but such a threat may lack credibility because the benefits of liquidating the borrower's assets are larger for higher-priority lenders, and the costs from the termination of the borrower's business are higher for lower-priority lenders. When the bank loan has sufficiently high priority, the bank could *credibly* threaten to call back the loan, and this may offset the deleterious effect of the soft-budget constraint. This identifies a potential advantage of bank financing: *timely intervention*. Of course, one could ask whether bondholders could be given priority and allocated the task of timely intervention. Note that bondholders are subject to more severe information asymmetries and are generally more dispersed (i.e., have smaller stakes). Both characteristics make them ill-suited for an 'early intervention' task.

Intermediation and competition

Since relationship banking is an integral part of the economic services provided by banks and generates rents for banks, it also potentially invites multiple bank entry, which then generates interbank competition. An interesting question this raises is how competition might affect the *incentives* for relationship banking. While this may ultimately be an empirical question, two diametrically opposite points of view have emerged theoretically. One is that competition among financiers encourages borrowers to switch to other banks or to the financial market. The consequent shortening of the expected 'life-span' of bank–borrower relationships may induce banks to reduce their relationship-specific investments, thereby inhibiting the reusability of information and diminishing the value of information (Chan, Greenbaum, and Thakor, 1986). Banks may then experience weaker incentives to

acquire (costly) proprietary information, and relationships may suffer. There is empirical evidence that an increase in relationship length benefits the borrower. Brick and Palia (2007) document a twenty-one-basis point impact on the loan interest rate due to a one standard deviation increase in relationship length.

Moreover, increased credit market competition could also impose tighter constraints on the ability of borrowers and lenders intertemporally to share surpluses (see Petersen and Rajan, 1995). In particular, it becomes more difficult for banks to 'subsidize' borrowers in earlier periods in return for a share of the rents in the future. Thus, the funding role for banks that Petersen and Rajan (1995) see in the case of young corporations (as we discussed) may no longer be sustainable in the face of sufficiently high competition. This implies that interbank competition may have an *ex post* effect of diminishing bank lending.[3]

An issue related to competition is the effect of consolidation. An extensive empirical literature focuses on the effect of consolidation in the banking sector on small business lending. This consolidation may in part be a response to competitive pressures. The effects on small business lending, however, are not clear-cut. Sapienza (2002) finds that bank mergers involving at least one large bank result in a lower supply of loans to small borrowers by the merged entity. This could be linked to the difficulty that larger organizations have in using 'soft information' (Stein, 2002 and Berger, Miller, Petersen, Rajan, and Stein, 2005). However, Berger, Saunders, Scalise, and Udell (1998) show that the actual supply of loans to small businesses may not go down after bank mergers, since they invite entry of 'de novo' banks that specialize in small business lending (see also Strahan, 2007).

The opposite point of view is that competition may actually *elevate* the importance of a relationship-orientation as a distinct competitive edge. The idea is that competition pressures profit margins on existing products and increases the importance of financier differentiation, and more intense relationship lending may be one way for the bank to achieve this. Boot and Thakor (2000) formalize this argument to show that a more competitive environment may encourage banks to become more client-driven and customize services, thus generating a *stronger* focus on relationship banking.[4] They distinguish between 'passive' transaction lending and more intensive relationship lending by banks. Transaction lending competes head-on with funding in the financial market. Competition from the financial market (as well as interbank competition) will lead to more resource-intensive relationship lending, and reduce transaction lending, since this mitigates

[3] Berlin and Mester (1999) provide a related, albeit different, argument. Their analysis suggests that competition forces banks to pay market rates on deposits, which may impede their ability to engage in the potentially value-enhancing smoothing of lending rates.

[4] In related work, Hauswald and Marquez (2006) focus on a bank's incentives to acquire borrower-specific information in order to gain market share, and Dinç (2000) examines a bank's reputational incentives to honor commitments to finance higher-quality firms. Song and Thakor (2007) theoretically analyze the effect of competition on the mix between relationship and transaction lending, and focus on fragility issues in particular.

the margin-reducing effects of price competition. The *absolute* level of relationship lending is, however, non-monotonic in the level of competition: initially competition increases relationship lending, but when competition heats up 'too much' investments in bank lending capacity will suffer and that may start to constrain relationship lending. Berger, Klapper, Martinez-Peria, and Zaidi (2008) find empirically that bank ownership type (foreign, state-owned, or private domestic) affects the bank's choice between transaction and relationship lending.

Relationships may foster the exchange of information, but may simultaneously give lenders an information monopoly and undermine competitive pricing. As discussed above, the informational monopoly on the 'inside' lender's side may be smaller if a borrower engages in multiple banking relationships. This would mitigate the possibilities for rent extraction by informed lenders and induce more competitive pricing (see Sharpe, 1990 and also Petersen and Rajan, 1995). Transaction-oriented finance, however, may give banks little incentive to acquire information but is potentially subject to more competition. This suggests that markets for transaction-oriented finance may fail when problems of asymmetric information are insurmountable without explicit information acquisition and information-processing intervention by banks. This argument is used by some to highlight the virtues of (relationship-oriented) bank-dominated systems (e.g., Germany and Japan) vis-à-vis market-oriented systems. This is part of the literature on the design of financial systems (see Allen, 1993; Allen and Gale, 1995; and Boot and Thakor, 1997). One objective of this literature is to evaluate the economic consequences of alternative types of financial system architecture.

What this discussion indicates is that the impact of competition on relationship banking is complex; several effects need to be disentangled. However, recent empirical evidence (see Degryse and Ongena, 2007) seems to support the Boot and Thakor (2000) prediction that the orientation of relationship banking *adapts* to increasing interbank competition, so higher competition does not drive out relationship lending. Despite this adaptation, there is also evidence that in recent years the geographic distance between borrowers and lenders has increased, and that this has been accompanied by higher loan defaults (see DeYoung, Glennon, and Nigro, 2008).

Bank lending, securitization, and capital market funding

Much of our focus in the previous section was on *interbank* competition. Nonetheless, banks also face competition from the capital market. The standard view is

that banks and markets compete, so that growth in one is at the expense of the other (e.g. Allen and Gale, 1995 and Boot and Thakor, 1997). In this context, Deidda and Fattouh (2008) show theoretically that both bank and stock market development have a positive effect on growth, but the growth impact of bank development is lower when there is a higher level of stock market development. They also present supporting empirical evidence. What this shows is that dynamics of the interaction between banks and markets can have *real* effects. How banks and markets interact is therefore of great interest.

In contrast to the standard view that they compete, the observations in the previous section suggest that there are also potential complementarities between bank lending and capital market funding. We argued that prioritized bank debt may facilitate timely intervention. This feature of bank lending is valuable to the firm's bondholders as well. They might find it optimal to have bank debt take priority over their own claims, because this efficiently delegates the timely intervention task to the bank. The bondholders will obviously ask to be compensated for their subordinated status. This—ignoring the timely intervention effect—is a 'wash'. In other words, the priority (seniority) and subordination features can be priced. That is, as much as senior debt may *appear* to be 'cheaper' (it is less risky), junior or subordinated debt will appear to be more expensive, and there should be no preference for bank seniority, other than through the timely bank-intervention channel. Consequently, the borrower may reduce its total funding cost by accessing both the bank-credit market and the financial market.[5]

Another manifestation of potential complementarities between bank lending and capital market activities is the increasing importance of securitization. Securitization is an example of unbundling of financial services. It is a process whereby assets are removed from a bank's balance sheet, so banks no longer permanently fund assets when they are securitized; instead, the investors buying asset-backed securities provide the funding. Asset-backed securities rather than deposits thus end up funding dedicated pools of bank-originated assets. More specifically, the

[5] This is directly related to the work on bargaining power and seniority; see the work of Gorton and Kahn (1993) and Berglöf and von Thadden (1994). The complementarity between bank lending and capital market funding is further highlighted in Diamond (1991), and Hoshi, Kashyap, and Scharfstein (1993). Diamond (1991) shows that a borrower may want to borrow first from banks in order to establish sufficient credibility *before* accessing the capital markets. Again, banks provide certification and monitoring. Once the borrower is 'established', it switches to capital market funding. Hoshi, Kashyap, and Scharfstein (1993) show that bank lending exposes borrowers to monitoring, which may serve as a certification device that facilitates simultaneous capital market funding. In this explanation, there is a *sequential* complementarity between bank and capital market funding. In related theoretical work, Chemmanur and Fulghieri (1994) show that the quality of the bank is of critical importance for its certification role. This suggests a positive correlation between the value of relationship banking and the quality of the lender. See Petersen and Rajan (1994) and Houston and James (1996) for empirical evidence.

lending function can be decomposed into four more primal activities: origination, funding, servicing, and risk processing (Bhattacharya and Thakor, 1993). Origination subsumes screening prospective borrowers and designing and pricing financial contracts. Funding relates to the provision of financial resources. Servicing involves the collection and remission of payments as well as the monitoring of credits. Risk processing alludes to hedging, diversification, and absorption of credit, interest rate, liquidity, and exchange rate risks. Securitization decomposes the lending function such that banks no longer fully fund the assets, but continue to be involved in other primal lending activities. A potential benefit of securitization is better risk sharing. The proliferation of securitization may, however, also be induced by regulatory arbitrage—for example, as vehicle to mitigate capital regulation (see Gorton and Pennacchi, 1995 for an economic rationale for bank loan sales and securitization).

Central to the extensive academic work on securitization is the idea that it is not efficient for originators to completely offload the risks in the originated assets. The originating bank needs to maintain an economic interest in the assets to alleviate moral hazard and induce sufficient effort on the originating bank's part in screening and monitoring. What this implies is that even with securitization banks do not become disengaged from the assets they originate. Banks still continue to provide the services involved in screening and monitoring borrowers, designing and pricing financial claims, and providing risk management and loan servicing support. As such, securitization preserves those functions that are at the core of the *raison d'être* for banks. This militates against the notion that securitization effectively lessens the importance of banks.

Boyd and Gertler (1994) have argued that the substitution from on-balance sheet to off-balance sheet banking induced by securitization may have falsely suggested a shrinking role for banks. Indeed, by keeping banks involved in their primal activity of pre-lending borrower screening, securitization preserves much of the banks' value added on the asset side.

Up to the 2007–9 financial crisis, securitization was rapidly gaining in importance. In fact, prior to the Summer of 2007, securitization became prevalent for ever-wider types of credits including business credits which were previously thought to be difficult to securitize because of their information opaqueness. Also, a rather new market for securitization involving asset-backed commercial paper (ABCP) conduits emerged as a significant force. As the subprime crisis of 2007 has shown, these developments are not without problems. The structure of real-world securitization transactions appears to have taken a rather fragile form. In particular, it is important to note that much of the securitization leading up to the crisis involved the financing of long-term assets with short-term funding, which induced substantial liquidity risk. While this liquidity risk was sometimes mitigated by liquidity guarantees (e.g., stand-by letters of credit and refinancing commitments), the underwriting institutions often underestimated the risks

involved and overstretched themselves.[6] Recent events may cast doubt on the optimality of such strategies. Also, because the originating institutions appeared to have retained minimal residual risk, monitoring incentives may have been compromised (see Mian and Sufi, 2007).[7] The eagerness of banks to securitize claims—and keep the repackaging 'machine' rolling—may have also adversely impacted the quality of loans that were originated through a dilution of banks' screening incentives due to lower retained residual risks (e.g. subprime lending).

The 2007–9 financial crisis has brought securitization almost to a grinding halt. However, the risk diversification that securitization can accomplish appears to be of more than just ephemeral importance. Thus, we expect securitization to re-emerge, albeit possibly in a form that entails lower levels of liquidity risk, as well as lesser moral hazard in screening (loan underwriting standards) and monitoring. A caveat is that some of the activity in securitization may have been induced merely by capital arbitrage, in which case its social value may be rather limited; the new Basel II capital requirements may diminish such regulatory arbitrage.

Another effect of the interaction between banks and markets is that as markets evolve and entice bank borrowers away banks have an incentive to create new products and services that combine services provided by markets with those provided by banks. This allows banks to follow their customers to the market rather than losing them. There are numerous examples. For instance, when a borrower goes to the market to issue commercial paper, its bank can provide a backup line of credit. Securitization of various sorts is another example in that banks not only originate the loans that are pooled and securitized, but they also buy various securitized tranches as investment securities. The impetus for such market-based activities grows stronger as interbank competition puts pressure on profit margins from traditional banking products and the capital market provides access to greater liquidity and lower cost of capital for the bank's traditional borrowers. As a consequence, there is a natural propensity for banks to become increasingly *integrated* with markets, and a sort of unprecedented 'co-dependence' emerges that makes banking and capital market risks become increasingly intertwined.

[6] Most noteworthy are the bankruptcies among German Lander banks that were involved in providing liquidity guarantees.

[7] Securitization is facilitated in part by credit enhancement, including partial guarantees by the arranger of a securitization transaction (and/or he holds on to the most risky layer of the transaction). In the recent credit crisis, this disciplining mechanism broke down; residual risk with the arranger was minimal or framed as liquidity guarantees to off-balance street vehicles without appropriately realizing the inherent risks. That is, banks have also been underwriting the liquidity risk in securitization transactions by, for example, guaranteeing the refinancing of commercial paper in ABCP transactions via standby letters of credit. Such guarantees have generated profits for banks, but also created risks, as illustrated by the losses incurred by banks in the recent subprime crisis. The marketability of securitized claims has also been facilitated by accreditation by credit rating agencies (see Boot, Milbourn, and Schmeits, 2006). However, even the role of rating agencies has been called into question during the sub-prime lending crisis.

A discussion of whether this is desirable and what the regulatory implications might be appears below.

Banks, equity, and private equity firms

The emergence of non-banking financial institutions like PE firms is considered by some a (further) signal for the diminishing role of banks. However, also here we will argue that these developments are rather complementary to the role of banks. Let us first discuss the role that PE firms play.

The arguments above about the need for banks to have seniority suggest a natural economic inhibiting of investments by banks in the equity of corporations. Equity 'softens' a bank's incentive to intervene for much the same reasons as does junior debt. So, while the emphasis of corporate finance theory on agency problems would suggest that it might be efficient for the bank to have both debt and equity claims on a corporation, this seems not to be advisable from a timely-intervention point of view. This might explain why equity intermediation has largely been in the hands of PE firms and/or bulge-bracket global investment banks that typically engage much less in relationship banking and focus more on transactions and the associated capital market activities.

Some more observations can be made about PE firms. Their activities could be viewed as intermediation driven from the equity side. That is, PE firms attract funding from a group of investors ('partners') and invest the funds as equity in businesses. They are extensively involved in monitoring and advising these businesses. How different is this from the role banks play as debt intermediaries? To address this question, note first that banks do occasionally take equity positions in their role as venture capitalists, particularly for later-stage financing where there is a prospect for developing a valuable relationship on the lending side. Thus, banks participate in venture capital financing with higher probability if there is a greater likelihood of subsequent lucrative lending activity (Hellmann, Lindsey, and Puri, 2008). Banks may also have (participations in) PE subsidiaries that operate independently from the other businesses of the bank. However, this somewhat limited role as an equity financier does not mean that it would be efficient for the bank permanently to become an integrated provider of debt and equity finance, a 'one-stop' financier of sorts (see our earlier discussion of the value of having senior claims). In particular, equity as a junior security may undermine a bank's bargaining power and thus compromise its role in timely intervention. Also, soft-budget constraint problems may then (re)emerge.

At a more general level, one could ask whether the monitoring role of PE firms substitutes for the lending-related monitoring of banks. It might. Note, however, that equity and debt are fundamentally different securities. The type of monitoring needed will differ significantly potentially across debt and equity. What will be true,

however, is that the increasing involvement of PE investors induces banks to partner with these investors. In a sense, banks start building relationships with PE firms rather than the firms that the PE investors take equity positions in. This is not without risks since it may affect the added value of banks in timely intervention vis-à-vis the (underlying) borrower and even the banks' incentives to be involved in this.[8] However, to the extent that PE firms are an integral part of the capital market, this development too makes the involvement of banks in the capital market deeper and more intricate. Such complexity is further exacerbated by the emergence of other intermediaries like hedge funds, particularly because of the growing importance of hedge funds as direct lenders. See Brophy, Ouimet, and Sialm (2009) who point out that hedge funds have emerged as 'lenders of last resort', providing finance to firms that banks do not typically lend to.

Role of credit rating agencies

Credit ratings are a fascinating part of today's financial markets. Their importance is evident from the behavior of market participants. However, academic researchers have generally been skeptical about their incremental value, largely because of the absence of a theory of rating agencies. In the literature on financial intermediary existence, bank debt offers monitoring advantages which would not be available in the financial market. The typical argument for the lack of monitoring in the capital market is that free-rider problems among investors prevent effective monitoring. Boot, Milbourn, and Schmeits (2006) have shown that credit rating agencies (CRAs) add a monitoring-type element to the financial market, and thereby play a role as a 'focal point' to resolve coordination failures among multiple dispersed investors (creditors). The CRA's ability to resolve such coordination failure arises from the effect of its actions—the assigned rating and the 'credit watch' process—on

[8] This suggests potential conflicts of interest. Much of the literature has focused on potential concerns related to banks combining lending and capital market activities, i.e., potential conflicts of interest in universal banking. This literature is motivated by the Glass-Steagall regulation in the US (see Kroszner and Rajan, 1994; Puri, 1996; and Ramirez, 2002). In similar spirit, Drucker (2005) shows that junk-rated firms and companies in local lending relationships are more likely to select an integrated (universal) commercial investment bank when they expect to issue public debt in the future. This revealed preference for commercial investment bank relationships by firms that issue informationally sensitive securities suggests that there are benefits for banks to use private information from lending in investment banking. A similar rather positive picture emerges if one looks at US banking following the 1999 Financial Services Modernization Act. It appears that information collected through the banks' commercial lending businesses may have reduced the costs of underwriting debt and equity (see Drucker and Puri, 2004; Schenone, 2004). Gande (2007) concludes that commercial banks have distinct benefits in underwriting leading to lower-issuer costs. He also concludes that 'the value of banking relationships appears to be largest for non-investment grade, small and IPO firms for whom one would ex ante expect the benefit of bank monitoring to be the highest'.

firm behavior via the conditioning of investors' investment decisions on the assigned rating. In earlier work, Da Rin and Hellmann (2002) showed that banks could also resolve a multiple-equilibria problem among borrowers by helping coordinate the investment decisions of these borrowers. The role that Boot, Milbourn, and Schmeits give to CRAs has some similarity to this.

This role of CRAs in resolving coordination failures in the financial market qualifies the distinction between public debt and bank financing. The mechanism is, however, less 'direct' than in the case of bank financing: the credit rating (and particularly the threat of a downgrade) *induces* good firm behavior rather than preventing bad behavior through direct intervention. Apart from bank loans, the non-bank private debt market also offers a potentially more direct alternative than credit rating agencies in the public debt market. In fact, private debtors often impose more discipline than banks and hence serve even riskier borrowers (Carey, Post, and Sharpe, 1998).

Another mechanism that links banks and CRAs is the certification role of bank loans. Datta, Iskandar-Datta, and Patel (1999) show that the monitoring associated with bank loans *facilitates* borrowers' access to the public debt market. This certification role of banks therefore complements what CRAs do. As rating agencies become more sophisticated and reliable, the certification role of banks diminishes in importance, causing bank borrowers to migrate to the capital market. In this sense, CRAs intensify the competition between banks and markets. But CRAs also pull banks into the capital market. For example, banks originate loans that they securitize, and then seek ratings for the securitized pools from CRAs. The ratings, in turn, facilitate the ability of banks to sell (securitized) asset-backed securities in the capital market.

This rather positive interpretation of CRAs is clouded somewhat by recent negative publicity. In the 2001 crisis surrounding Enron, CRAs were accused of being strategically sluggish in downgrading.[9] More recently, CRAs have been blamed (in part) for the subprime crisis in which they were allegedly too lenient in rating the senior tranches in securitization transactions. Allegations have been made about conflicts of interest for CRAs arising from the fact that structured finance is a source of ever-increasing income for CRAs, which then corrupts their incentives for accurately rating the issuers involved in structured finance (Cantor, 2004). In this context, Coffee and Sale (2008) point out that it is naive to think that reputation-building incentives alone would keep credit rating agencies in check.

[9] See, e.g., discussions in the US Senate: 'On March 20, 2002, the Senate Committee held a hearing entitled "Rating the Raters: Enron and the Credit Rating Agencies"... The hearing sought to elicit information on why the credit rating agencies continued to rate Enron a good credit risk until four days before the firm declared bankruptcy...' (US Senate Hearings, 2002). Similarly, US Senate Staff Report (2002): 'in the case of Enron, credit rating agencies displayed a lack of diligence in their coverage and assessment of Enron.' See also Cantor (2004) and Partnoy (1999).

Of particular concern are the so-called 'rating triggers'. For example, some debt contracts may dictate accelerated debt repayments when the rating falls. The consequences of such accelerated debt repayments might, however, be so severe as to cause rating agencies to become reluctant to lower the ratings of those borrowers in a timely manner. Complications also arise from the role played by the so-called 'monoliners'. These are insurers who traditionally guaranteed municipal bonds but now also guarantee the lowest-risk (best) tranches in securitization transactions. These insurers are virtually indispensible in the sense that the viability of many forms of securitization is predicated on this type of 'reinsurance'. However, the ability of the monoliners to issue credible guarantees (and hence their role in securitization) depends on these institutions themselves having AAA ratings. This potentially generates an indirect chain-reaction mechanism for CRAs. In rating (and monitoring) the monoliners, CRAs affect the viability of the securitization market. Thus, the impact of CRAs is both direct (rating securitization tranches) and indirect (rating the monoliners). The potential failure of such monoliners would have a significant effect on the value of various structured finance products and induce an additional chain reaction among players active in the structured finance market, including investors. This further underscores the increasing interlinkages in the financial markets. Other concerns are related to the oligopolistic nature of the industry, and the importance that ratings have due to regulation. The latter includes the exclusivity given to a few rating agencies via the 'Nationally Recognized Statistical Rating Organization' (NRSRO) classification, recently weakened in the 2006 Credit Rating Agency Reform Act, but also the inclusion of external ratings in the new Basel II capital regulation framework.

Regulation and the second *raison d'être* for banks: liquidity creation

In the second section, we discussed the role of banks as information processors and delegated monitors. That information processing and monitoring referred to credit risk primarily. But banks also perform another important function, which is the provision of liquidity. That is, banks invest in illiquid assets (loans) but finance themselves largely with highly liquid demand deposits, and through this intermediation process create liquidity in the economy. However, in the process of creating liquidity, banks expose themselves to withdrawal risk and become fragile. Our discussion of this issue in this section will focus on 'institution-driven fragility', manifested in the classic run on an individual bank, as well as 'market-driven fragility' that refers to risks that come primarily via the financial market and

interbank linkages, and appear to be more systemic. We will discuss how the increasing integration of banks into financial markets allows banks to shift some of their traditional risks to the markets, and what this implies for *financial system stability* and regulation. Issues related to the economics of bank regulation are covered in Bhattacharya, Boot, and Thakor (1998; 2004).

Fragile banks as liquidity providers

In the classical interpretation, a financial crisis is directly linked to the notion of bank runs. In a fractional reserve system with long-term illiquid loans financed by (liquid) demandable deposits, runs may come about due to a coordination failure among depositors (Diamond and Dybvig, 1983). Even an adequately capitalized bank could be subject to a run if the deadweight liquidation costs of assets are substantial. Regulatory intervention via lender of last resort (LOLR) support, deposit insurance, and/or suspension of convertibility could all help, and perhaps even eliminate the inefficiency. In fact, such intervention can be justified because of its potential to expunge the negative social externalities arising from the possible contagion effects associated with an individual bank failure. While these implications arise theoretically in a rather simple and stylized setting, many have generalized this simple setting by allowing for asymmetric information and incomplete contracts; see Rochet (2004) for a review. The general conclusion is that fragility is real, and information-based runs are plausible. In particular, Gorton's (1988) empirical evidence suggests that bank runs are *not* sunspot phenomena (as in Diamond and Dybvig (1983), but are triggered by adverse information about banks. More importantly, the banking crises stemming from such runs have *independent* negative real effects (see Dell'Ariccia, Detragiache, and Rajan, 2008). Also relevant in this context is the large literature that has now developed on banks and liquidity (see, e.g., Acharya, Gromb, and Yorulmazer, 2007a; Acharya, Gromb, and Yorulmazer, 2007b; and Acharya and Schaefer, 2006).

Given that bank runs are triggered by adverse information that depositors have about the financial health of banks, one might think that a simple solution would be to make banks safer by, for example, imposing higher capital requirements. Calomiris and Kahn (1991) first argued that the threat of bank runs may be a valuable disciplining device to keep bank managers honest, since a greater diversion of bank resources for personal consumption can increase the likelihood of a bank run. Building on this argument, Diamond and Rajan (2001) have suggested that financial fragility may play an important role in inducing banks to create liquidity, and thus a reduction in fragility through higher bank capital may lead to lower liquidity creation. Until recently, there has been no empirical work done on this issue, in part because of a paucity of empirical measures of liquidity creation. In recent work, Berger and Bouwman (2008a) develop measures of liquidity

creation and provide empirical evidence on the relationship between bank capital and liquidity creation. They show that higher capital leads to higher liquidity creation in the case of large banks, and lower liquidity creation in the case of small banks. Since capital requirements also affect the asset portfolios of banks through their lending decisions (see Thakor, 1996) and these requirements may be binding for some banks, this raises issues about the interaction of credit and liquidity risks that need to be explored.

Complicating this issue further is that the liquidity provision function of banks is also affected by the financial markets. Two observations are germane in this regard. First, access to financial markets weakens the liquidity insurance feature of demand-deposit contracts. To see this, note that the root cause of the fragility in the Diamond and Dybvig (1983) world is the underlying demand-deposit contract. The rationale for this contract—as modeled by Diamond and Dybvig (1983)—is the desire for liquidity insurance on the part of risk-averse depositors with uncertainty about future liquidity needs. However, as shown by von Thadden (1998), the very presence of financial markets allows depositors to withdraw early and invest in the financial market, which puts a limit on the degree of liquidity insurance. In fact, when the market investment opportunity is completely reversible, deposit contracts cannot provide any liquidity insurance. This is related to the earlier work of Jacklin (1987) who shows that deposit contracts have beneficial liquidity insurance features provided that restricted trading of deposit contracts can be enforced.[10] In any case, these arguments suggest that the proliferation of financial markets weakens the liquidity-provision rationale for demand deposits, which may help explain the market-based proliferation of close substitutes for deposits.

A second observation has to do with whether the development of financial markets leads to a diminished role for the Central Bank in providing liquidity via its LOLR function. In the Bagehot tradition, one could ask whether the LOLR has a role to play in providing liquidity to liquidity-constrained-yet-solvent institutions when capital markets and interbank markets are well developed. Goodfriend and King (1988) argue that solvent institutions then cannot be illiquid since informed parties in the repo and interbank market would step in to provide the needed liquidity. In this spirit, former European Central Bank (ECB) board member Tommaso Padoa-Schioppa suggested that the classical bank run may only happen in textbooks since the 'width and depth of today's interbank market is such that other institutions would probably replace those which withdraw their funds' (as quoted in Rochet and Vives, 2004b).

While these remarks correctly suggest that the development and deepening of financial markets could reduce the need for a LOLR in providing liquidity support,

[10] Actually, Jacklin (1987) shows that with the 'extreme' Diamond-Dybvig preferences, a dividend-paying equity contract can achieve the same allocations without the possibility of bank runs. However, for basically all other preferences, a demand deposit contract does better, provided that trading opportunities are limited.

we believe that it would be hasty to conclude that there is no role for a LOLR, particularly when information asymmetries are considered. For example, Rochet and Vives (2004b) show that a coordination failure in the interbank market may occur, particularly when fundamentals are weak, and that this may lead to a need for liquidity support by the LOLR for a solvent institution.[11] The 2007–9 financial crisis gives ample reason to believe that coordination failures in interbank markets are real and that the role of a LOLR is still important.

This discussion suggests two somewhat tentative conclusions. First, the development of financial markets (including interbank markets) has improved the risk sharing opportunities available to banks and has probably decreased the likelihood of a run on an individual bank. Whether the total insolvency risk of individual institutions has declined depends on the actual risk taking and capitalization. Evidence in De Nicolo and Tieman (2005) suggests that the insolvency risk of European institutions has remained more or less the same over the last fifteen years despite increases in capital over time and a wider geographic range of operations. Second, because these improved risk sharing opportunities have arisen from a greater degree of integration between banks and markets, they may also have contributed to an *increase* in *systemic* risk. In other words, while the likelihood of an individual bank failing due to an idiosyncratic shock may have declined, there may be a concomitant increase in the probability that localized liquidity and solvency problems may propagate quickly through the financial system as a whole, leading to higher systemic risk. This raises thorny regulatory issues, which we turn to next.

Regulatory implications

The preceding discussion has focused the spotlight on one fact: banks and markets are becoming increasingly integrated. This is happening in part because of greater competition is inducing banks to follow their borrowers to the capital market and offer products that *combine* features of bank-based and market-based financing. It is also happening because banks themselves are using the financial market increasingly for their own risk management purposes. There is thus a multitude of factors that have contributed to an astonishingly rapid melding process.

An important implication of this integration is that it is becoming more and more difficult to isolate banking risks from financial market risks. A financial market crisis inevitably cascades through the banking system, and what happens in the banking system does not take long to reverberate through the financial

[11] Another line of research studies the impact of liquidity on asset pricing (e.g. Acharya and Pedersen, 2005) and the possible role of asset price bubbles as a source of fragility and contagion (see Allen, 2005; De Bandt and Hartmann, 2002, for surveys on contagion).

market. So, if the main task of bank regulators is the safety and soundness of the banking system, they must now also worry about the financial market whose participants are outside the bank regulator's domain.

Moreover, even though the explicit insurance guarantee applies only to bank deposits, the temptation for government regulators to bail out various uninsured participants, including investment banks and financial market investors, in the event of a crisis in the capital market seems increasingly difficult to resist on *ex post* efficiency grounds, particularly because of the implications for *bank safety*.[12] It will be interesting to examine the connotations of this for *ex ante* incentives and the magnitude of the *implicit* 'soft' safety net provided by the government. What seems safe to conjecture is that a perception of a greater regulatory concern with *ex post* efficiency and hence a greater desire to intervene has elevated the importance of moral hazard. And this has happened in an environment in which regulatory issues are becoming increasingly international both due to the cross-border proliferation of financial institutions and the increasing integration of banks with financial markets, which are typically international in scope.

Need for cross-border coordination in regulation and supervision: the European Union example

The regulatory task across national boundaries is rather complex. Consider the European Union as an example. The patchwork of national supervision and European-wide coordination in the European Union has so far held itself up reasonably well, arguably even during the 2007–9 financial crisis.[13] Nevertheless, in crisis situations, important concerns can be raised about the adequacy of information sharing and cooperation between the various supervisors, the European Central Bank (ECB), and the national central banks. In particular, in such situations the question about who will be in charge might become paramount. Potential tensions can easily be envisioned between supervisory agencies, national central banks and the ECB. Moreover, one could ask to what extent these arrangements accomplish the efficiency and effectiveness objectives that regulation and supervision should be subjected to.

[12] The guarantee provided to a collapsing Bear Stearns by the government to facilitate its sale to JPMorgan Chase is an example, as are the general measures to let investment banks qualify for a commercial banking license (and in doing so allow them access to deposits and let them qualify for deposit insurance).

[13] Nevertheless, several things did go wrong, most notably the non-coordinated actions surrounding deposit insurance. Some countries chose to offer blanket guarantees overnight (e.g. Ireland) and in doing so imposed severe externalities on other countries and also foreign banks in their own markets that were not covered. These foreign countries and banks faced an immediate erosion of their deposit base.

Policymakers are aware of these issues. For example, the new Directive on Financial Conglomerates gives the home country supervisor the single coordinating responsibility in all member states for group-wide supervision of the financial conglomerate. Issues of financial stability, however, remain the responsibility of the host countries.

The question is how to coordinate these potentially diverse interests, particularly in crisis situations. The core message of the second Brouwer report (Economic and Finance Committee, 2001) was that *no* mechanism was in place to coordinate in case of a crisis.[14] For that reason a Memorandum of Understanding between virtually all European national central banks and supervisors was formulated that specifies principles and procedures for cooperation in crisis management situations (European Central Bank, 2003). However, the fiscal side, in particular the budgetary obligations imposed on member states in the case of bailouts, also requires the approval of national finance ministries that have to incur the potential financial obligations associated with bailouts. In a follow-up Memorandum of Understanding, these finance ministries were also included (European Central Bank, 2005). Several questions can be raised about the efficiency of the arrangements in general. The decentralized structure may give rise to potential conflicts of interest between the national authorities and 'outsiders'. For example, national authorities might be prone to 'too-big-to-fail' (TBTF) rescues, and this worsens moral hazard on the part of large institutions. Yet one could argue that the moral hazard engendered by TBTF policies could be attenuated somewhat by attaching to TBTF rescues specific provisions that would involve replacing management, wiping out the claims of shareholders and uninsured debtholders, etc. This is true in theory but does not appear to happen often in practice. One reason might be the possibility of capture of local regulators and supervisors due to the closeness of their relationships to the 'national flagship' institutions (Boot and Thakor, 1993). There are also issues of 'too many to fail' (see Acharya and Yorulmazer, 2007) or 'too interconnected to fail' (Herring, 2008), which could also induce regulatory leniency toward these institutions. Alternatively, national authorities may not sufficiently internalize the disruptive consequences that a domestic bank failure could have in other countries. Efficiency might be hampered in other ways as well. For example, the national scope of supervision may help encourage the emergence of 'national champions' among regulators, who may then seek to protect institutions in their countries. More fundamentally, the decentralized structure could give rise to an uneven playing field and regulatory arbitrage possibilities.

Casual observation would seem to suggest that integration and further coordination (if not centralization of authority) of both regulation and supervision might yield substantial efficiency gains not only for the supervisory authorities but also, and perhaps more importantly, for the supervised financial institutions

[14] See Economic and Finance Committee, 2002 for further recommendations.

themselves. There are currently more than 35 supervisory authorities responsible for prudential supervision in the European Union, and a typical large financial institution might have to report to more than 20 supervisors (Pearson, 2003).

Yet, practical considerations suggest that a full integration of all regulatory and supervisory functions at the European level may not (yet) be feasible. While it is clear that regulatory and supervisory integration needs to keep pace with the development of the size and the cross-border footprint of the covered banks, the heterogeneity of underlying supervisory systems and the implied costs of integration should not be underestimated. An interesting illustration is the evidence reported by Barth, Caprio, and Levine (2004) on the variation across the European Union countries in supervisory institutions and practices. Their conclusion is that supervisory arrangements within the European Union are as diverse as in the rest of the world. Also, illustrating this point further, the European Union countries are current or former standard bearers of all major legal origins. A vast literature now documents how legal origin matters for the shape and functioning of the financial system (see La Porta, Lopez DeSilanes, Schleifer, and Vishny, 1998). Bank regulation and supervisory practices differ also considerably between civil and common law countries, typically with a more flexible and responsive approach in the latter.

While common sense suggests that ultimately a more integrated regulatory and supervisory structure is desirable[15], the way we should get there is far from clear. Indeed, practical considerations, including political concerns, suggest at least the short-run inevitability of a fragmented structure. A coordination layer will then need to be superimposed on this structure; assigning single coordinating responsibility to the home country supervisor is one example of that.[16]

The 2007–9 credit crisis may well lead to a situation in which central banks get a heavier role in supervision. While central banks always had a role in safeguarding the stability of the financial system, during the 2007–9 crisis we have seen that both the Federal Reserve and the ECB became directly involved in rescuing depository as well as non-depository financial institutions. An important question in the current debate is whether this expanded role should be formalized. For example, the ECB has hinted at obtaining a mandate for the supervision of systemically relevant banks that operate across national borders. This reflects a significant change in

[15] Actually, some theoretical work suggests the potential value of competition between regulators. See, e.g., Kane (1988).

[16] An important distinction needs to be made between business conduct regulation and prudential regulation. We have focused on the latter. The former is closer to the functioning of financial markets and lends itself more readily for centralization at the European level. But even in context of these financial markets, the Lamfalussy report (Committee of Wise Men, 2001) that is the blueprint for financial market supervision in the European Union does not directly propose authority at the EU level, but introduces a collaboration model that induces regulatory and supervisory convergence. It states that if its proposed approach is not successful, the creation of a single EU regulatory authority should be considered.

thinking. Prior to the crisis, the consensus appeared to be that caution was in order when it came to expanding the mandate of central banks because an expanded mandate could compromise the pivotal function of central banks in conducting monetary policy.

Other reform suggestions

The struggle for better cross-border coordination in regulation and supervision should go hand in hand with more fundamental reforms in the regulatory structure. The first is that the scope of regulation and supervision needs to be clearly identified, and, if possible, contained. Effective supervision and regulation—given the mushrooming cross-sector and cross-border footprint of financial institutions—requires a better delineation of safety and systemic risk concerns. The earlier discussion on the precise propagation mechanism as it relates to systemic risk is actually pointing at the same issue. The cross-sector integration of financial institutions and the increasingly more seamless integration of financial markets and institutions have considerably broadened the scope of regulation and the potential sources of systemic risk.

Another relevant question is whether market discipline could help in containing systemic risks, or whether market responses merely amplify such risks (see Flannery, 1998). Here the picture gets a bit murky. Basel II tries to encourage market discipline via its third pillar that is aimed at greater transparency. The idea is that market discipline could help supervisors in safeguarding the well-being of the financial sector. This has merit on the face of it and has support in the literature as well. The literature has viewed market discipline working in three ways: (1) by providing regulators market-based signals of bank risk taking through the yields on subordinated debt issued by banks; (2) by providing banks disincentives to take excessive risk through the upward adjustments in sub-debt yields in response to greater bank risk; and (3) by choking off the supply of sub-debt when sufficiently high risk taking by the bank is detected by the market, thereby providing additional encouragement to the bank to temper its risk taking. Nonetheless, it has been shown both theoretically and empirically that market discipline can be effective only if the claims of uninsured investors (sub-debt and equity) are not protected via de facto *ex post* insurance in a government-sponsored rescue of a failing institution. For a theoretical treatment of these issues, see Decamps, Rochet, and Roger (2004), and for empirical analyses that support the risk-controlling role of market discipline, see Barth, Caprio, and Levine (2004), and Goyal (2005). However, despite all of the research support for the role of market discipline, our knowledge of whether market discipline facilitates or hinders the regulatory task of maintaining banking stability *during* a financial crisis is quite limited. In particular, when the financial sector is severely stressed, as during the 2007–9 credit crisis, market discipline may induce

herding behavior, as everybody 'heads simultaneously for the exit', and this actually could be a source of instability. This suggests that regulation and supervision in 'normal times' should perhaps be distinguished from that during crisis episodes. Market discipline, although valuable in normal times, may be very distortive in times of systemic stress. This may be one reason why during crises regulators have been inclined to provide more or less blanket guarantees to distressed institutions, ostensibly to counter the potentially adverse effects of market discipline. However, all of this notwithstanding, it would be dangerous to conclude that market discipline, say via the use of market value accounting and other mechanisms, is something that should be relied upon in good times and eschewed in bad times. The key is to figure out the appropriate regulatory actions in *good* times, when banks have the flexibility to comply without compromising their viability, that would enable banks to be more capable of withstanding the stresses of market discipline during bad times. And it will also be important to remember that banks cannot be completely insured from the effects of market stress during bad times (e.g., through the use of blanket guarantees for *all* claimants), or else the *ex ante* effectiveness of market discipline is lost entirely (e.g., Decamps, Rochet, and Roger, 2004).

This brings up the issue of introducing firewalls in the financial sector. For example, does a subsidiary structure reduce systemic risk concerns? We do not think that an answer is readily available. More generally, what type of constraints, if any, should be put on the corporate structure of financial institutions? Until the 2007–9 financial crisis, the general belief was that deregulation in the financial sector would continue further, possibly leading to even bigger and broader financial institutions. But, now it is far from clear what the future will bring. Some have suggested reintroducing the Glass-Steagall Act to insulate local banking from the risks and fads that periodically afflict financial markets. To what extent this is effective, and not overly costly, is open to debate. In any case, changes in the industrial structure of the financial sector are of paramount importance for the design and effectiveness of regulation and supervision.[17] If these issues cannot be satisfactorily addressed, we are not very optimistic about the possibilities for effective and efficient pan-European regulation, let alone globally coordinated regulation, even in the long run.

A second issue has to do with the evolution of capital regulation. Many believe that banks should operate with higher capital buffers. This is somewhat at odds with Basel II, which permits banks to fine-tune their required capital ratios based on their (certified) internal models. There are questions about whether these models induce pro-cyclicality, and whether such model-dependency induces

[17] Earlier we referred to the concentration in the credit rating business and the importance of ratings for the markets for structured finance (securitization). It is interesting to ask what impact a meltdown of one of the main credit rating agencies would have on these markets, and what this in turn would imply for participants in these markets.

systemic risk by itself (e.g., institutions using the same models, and thus potentially being subject to the same shortcomings). We also have a concern about the potential adverse consequences of the discretion that Basel II provides to banks.[18] Perhaps similar concerns led the FDIC to impose a minimum leverage ratio on banks in the Basel II environment. The FDIC has argued that requiring a minimum level of capital—regardless of risk—is essential for timely regulatory intervention in the event of problems. Such timely intervention seems particularly important in cross-border situations, given the complexities created by bank failures in such situations. In particular, timely regulatory intervention could help contain conflicts between local authorities in such cases (see Eisenbeis and Kaufman, 2005).

A third issue is deposit insurance. The 2007–9 financial crisis has made it clear that, when a real crisis hits, national authorities effectively feel compelled to fully guarantee the deposit bases of their financial institutions to eliminate the possibility of massive runs. This heavy dependence on insured deposits is an issue that needs a re-examination. Extant research (see Bhattacharya, Boot, and Thakor, 1998) has clearly shown the moral hazards that insured deposits entail. Moreover, Barth, Caprio, and Levine (2004) have shown that high levels of (de facto or *de jure*) deposit insurance impede the effectiveness of market discipline and increase the likelihood of a banking crisis, controlling for many other aspects of the regulatory environment. A question is whether strict regulatory limits should be put on the risks that institutions can expose these deposits to. Earlier research had at some point advocated narrow banking which fully insulates insured deposits. But are there alternatives? And more generally, can insured deposits be made less important as a funding vehicle for financial institutions?

A fourth issue is whether regulation and supervision sufficiently effectively address macro-prudential issues, in particular systemic concerns. It appears that the majority of regulatory initiatives are focused on the well-being of individual financial institutions. That is, a micro-prudential focus dominates (see Brunnemeier, Crockett, Goodhart, and Shin, 2009). This should be addressed to better reconcile regulation and supervision with the systemic concerns that are paramount.

The fifth issue is that very little is known about the efficiency and effectiveness of various regulatory and supervisory structures. As Barth, Nolle, Phumiwasana, and Yago (2003) put it, 'there is very little empirical evidence on how, or indeed whether, the structure, scope or independence of bank supervision affects the banking industry'. Their own research suggests that the effect is at best marginal, but measurement problems are vexing. They conclude from this that we may thus

[18] This concern stems from the observation that individual banks are unlikely to sufficiently internalize the systemic-risk externalities of their actions. Consequently, the latitude that Basel II grants to banks in having them use their own internal risk assessment models to determine appropriate capital levels is misplaced. Banks appear to have powerful incentives to tweak these models in order to generate prescriptions to keep low levels of capital.

choose to focus only on the effect that regulation has on systemic risk. But here, too, little is known about the regulatory structures that are most efficient in dealing with systemic risk. What this means is that we need considerable additional research to sharpen our identification of the costs and benefits of different regulatory and supervisory arrangements. Given the strikingly different national supervisory arrangements that exist today, our lack of knowledge on this issue is a significant barrier to progress toward a harmonized 'superior' model.[19]

Conclusions

We have reviewed some of the literature on why banks exist, the risks these create, and how interbank competition as well as that from markets affects the economic roles served by banks as well as the attendant risks. One important development is that banks have become increasingly integrated with markets. This integration generates two effects that work in opposite directions. On the one hand, individual banks become better equipped to manage their own risks because it becomes easier and less costly to hedge these risks using the market. This could reduce the risk of an individual bank failing due to an idiosyncratic shock. On the other hand, there is an increase in the probability that a shock to a small subset of banks could generate systemic effects that ripple through the financial market, so that this banks-markets integration may be causing an elevation of systemic risk.

It is easy to see that this substantially complicates the task of prudential regulation of banks and raises the specter of a widening of the 'implicit' governmental safety net as *ex post* efficiency concerns tempt the government to bail out even uninsured players. This is no longer a mere theoretical conjecture, as demonstrated by the bailouts of investment banks and insurance companies in 2008–9. We believe that these are important issues that deserve greater theoretical and empirical attention. In particular, we need to have a better understanding of what the regulatory intervention in a crisis should be. Governmental initiatives such as those witnessed in the US during the 2007–9 crisis—massive governmental injections of liquidity and capital into banks and other financial institutions without an adequate corporate control role for the government—are very costly and possibly ineffective due to daunting moral hazard and asymmetric information problems. Some key

[19] We have not focused on changes that might be needed in the internal incentive structure in banks. As has become clear in the current crisis, internal risk management showed substantial lapses (see Group of Thirty, 2009). Other issues abstained from in this chapter relate to pro-cyclicality in Basel II and IFRS (and market value) accounting.

lessons might be learnt from previous financial crises—for example, the Swedish financial crisis of the 1990s (see Ingves and Lind, 1994).[20]

To conclude, we believe the most important, yet only partially answered, research questions raised by our discussion are the following:

- What are the implications of the ever-increasing integration of banks and markets for *systemic* risk and fragility?
- What issues should we consider in the optimal design of regulation to respond to the (up to recently, at least) growing cross-border footprints of major financial institutions and the increasing integration of banks and financial markets?
- What changes, if any, should be imposed on the structure of the financial services industry, and the banking sector in particular, to contain the 'mushrooming' nature of systemic risk concerns (i.e., to contain the scope of regulation and supervision)?
- What role, if any, can market discipline play in helping safeguard the stability of the financial sector?
- How do banks and private equity firms (and other non-banking financial institutions) interact and what implications does this have for the regulation of banks and financial markets?
- What role do credit rating agencies play in financial markets, how does this affect banks, and what implications does this have for systemic risks that *bank* regulators care about?

These questions represent a rich agenda for future research.

References

Acharya, V., Gromb, D., and Yorulmazer, T. (2007a). 'Failure in the Market for Liquidity Transfers and the Origins of Central Banking', Working Paper.

——— ——— ——— (2007b). 'Imperfect Competition in the Inter-Bank Market for Liquidity', Working Paper.

——— and Heje Pedersen, L. (2005). 'Asset Pricing with Liquidity Risk', *Journal of Financial Economics*, 77: 375–410.

[20] See also the theoretical work of Aghion, Bolton, and Fries (1999). In a recent commentary, Stiglitz (2009) advocates nationalization of banks for some time to facilitate their clean up. A better proposal (made by Richard Herring), that closely follows the Swedish experience, might be that of 'bridge banks', where there is temporary involvement of the government (say, for two years). During this time the government has an adequate corporate control role, asset sales are handled in an orderly manner, incentives are realigned, and the bank is put back in a position to be viable again. The regulatory apparatus for this already exists in many countries, including the US. Actually, many variations on this are observed in the 2008–9 handling of the crisis. For example, in some cases (see Citi and the Dutch bank ING), governments had chosen to install fences around troubled assets within the banks' corporate structures, and provide explicit government guarantees on these assets.

Acharya, V., and Schaefer, S. (2006). 'Liquidity Risk and Correlation Risk: Implications for Risk Management', Working Paper.
—— and Yorulmazer, T. (2007). 'Too Many to Fail—An Analysis Of Time-Inconsistency in Bank Closure Policies', *Journal of Financial Intermediation*, 16: 515–54.
Aghion, P., Bolton, P., and Fries, S. (1999). 'Optimal Design of Bank Bailouts: The Case of Transition Economies', *Journal of Institutional and Theoretical Economics*, 155: 51–70.
Allen, F. (1990). 'The Market for Information and the Origin of Financial Intermediation', *Journal of Financial Intermediation*, 1: 3–30.
—— (1993). 'Stock Markets and Resource Allocation', in *Capital Markets and Financial Intermediation*, C. Mayer and X. Vives (eds.), Cambridge: Cambridge University Press.
—— (2005). 'Modeling Financial Instability', *National Institute Economic Review*, 192: 57–67.
—— and Gale, D. (1995). 'A Welfare Comparison of Intermediaries and Financial Markets in Germany and the U.S.', *European Economic Review*, 39: 179–209.
—— —— (1997). 'Financial Markets, Intermediaries and Intertemporal Smoothing', *Journal of Political Economy*, 105: 523–46.
Barth, J. R., Nolle, D. E., Phumiwasana, T., and Yago, G. (2003). 'A Cross-Country Analysis of the Bank Supervisory Framework and Bank Performance', *Financial Markets, Institutions & Instruments*, 12: 67–120.
—— Caprio, G., and Levine, R. (2004). 'Bank Regulation and Supervision: What Works Best?', *Journal of Financial Intermediation*, 13: 205–48.
Berger, A. N. and Bouwman, C. (forthcoming). 'Bank Liquidity Creation', *Review of Financial Studies*.
—— Klappper, L. F., Martinez-Peria, M. S., and Zaidi, R. (2008). 'Bank Ownership Type and Banking Relationships', *Journal of Financial Intermediation*, 17: 37–62.
—— Miller, N., Petersen, M., Rajan, R., and Stein, J. (2005). 'Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks', *Journal of Financial Economics*, 76: 237–69.
—— Saunders, A., Scalise, J., and Udell, G. (1998). 'The Effects of Bank Mergers and Acquisitions on Small Business Lending', *Journal of Financial Economics*, 50: 187–229.
Berglöf, E. and von Thadden, E.-L. (1994). 'Short-Term Versus Long-Term Interests: Capital Structure with Multiple Investors', *Quarterly Journal of Economics*, 109: 1055–84.
Berlin, M. (1996). 'For Better and for Worse: Three Lending Relationships', *Business Review Federal Reserve Bank of Philadelphia*, 3–12.
—— and Mester, L. (1992). 'Debt Covenants and Renegotiation', *Journal of Financial Intermediation*, 2: 95–133.
—— —— (1999). 'Deposits and Relationship Lending', *Review of Financial Studies*, 12: 579–607.
Bhattacharya, S., Boot, A. W. A., and Thakor, A. V. (1998). 'The Economics of Bank Regulation', *Journal of Money, Credit and Banking*, 30: 745–70.
—— —— —— (eds.) (2004). *Credit Intermediation and the Macro Economy*, Oxford: Oxford University Press.
—— and Chiesa, G. (1995). 'Proprietary Information, Financial Intermediation, and Research Incentives', *Journal of Financial Intermediation*, 4: 328–57.
—— and Thakor, A. V. (1993). 'Contemporary Banking Theory', *Journal of Financial Intermediation*, 3: 2–50.
Bolton, P. and Scharfstein, D. (1996). 'Optimal Debt Structure and the Number of Creditors', *Journal of Political Economy*, 104: 1–25.

Boot, A. W. A. (2000). 'Relationship Banking: What Do We Know?', *Journal of Financial Intermediation*, 9: 7–25.
—— Greenbaum S. G. and Thakor, A. V. (1993). 'Reputation and Discretion in Financial Contracting', *American Economic Review*, 83: 1165–83.
—— Milbourn, T. and Schmeits, A. (2006). 'Credit Ratings as Coordination Mechanisms', *Review of Financial Studies*, 19: 81–118.
—— and Thakor, A. V. (1993). 'Self-Interested Bank Regulation', *American Economic Review*, 83: 206–12.
—— —— (1997). 'Financial System Architecture', *Review of Financial Studies*, 10: 693–733.
—— —— (2000). 'Can Relationship Banking Survive Competition?', *Journal of Finance*, 55: 679–713.
—— —— and Udell, G. (1991). 'Credible Commitments, Contract Enforcement Problems and Banks: Intermediation as Credibility Assurance', *Journal of Banking & Finance*, 15: 605–32.
Boyd, J. H. and Gertler, M. (1994). 'Are Banks Dead, or Are the Reports Greatly Exaggerated?', *Federal Reserve Bank of Minneapolis Quarterly Review*, 18.
Brick, I. E. and Palia, D. (2007). 'Evidence of Jointness in the Terms of Relationship Lending', *Journal of Financial Intermediation*, 16: 452–76.
Brophy, D., Ouimet, P. P., and Sialm, C. (2009). 'Hedge Funds as Investors of Last Resort?', *Review of Financial Studies*, 22: 541–74.
Brunnemeier, M., Crockett, A., Goodhart, C., and Shin, H. (2009). 'The Fundamental Principles of Financial Regulation, Preliminary Draft of *Geneva Reports on the World Economy*', 11, International Center for Monetary and Banking Studies, Geneva.
Calomiris, C. and Kahn, C. (1991). 'The Role of Demandable Debt in Structuring Optimal Banking Arrangements', *American Economic Review*, 81: 497–513.
Cantor, R. (2004). 'An Introduction to Recent Research on Credit Ratings', *Journal of Banking & Finance*, 28: 2565–73.
Carletti, E., Cerasi, V., and Daltung, S. (2007). 'Multiple Bank Lending: Diversification and Free-Riding in Monitoring', *Journal of Financial Intermediation*, 16: 425–51.
Carey, M., Post, M., and Sharpe, S. A. (1998). 'Does Corporate Lending by Banks and Finance Companies Differ? Evidence on Specialization in Private Debt Contracting', *Journal of Finance*, 53: 845–78.
Chan, Y.-S., Greenbaum, S. G., and Thakor, A. V. (1986). 'Information Reusability, Competition and Bank Asset Quality', *Journal of Banking & Finance*, 10: 243–53.
Chemmanur, T. J. and Fulghieri, P. (1994). 'Reputation, Renegotiation, and the Choice between Bank Loans and Publicly Traded Debt', *Review of Financial Studies*, 7: 475–506.
Coffee, J. C. (2002). 'Understanding Enron: It's about the Gatekeepers, Stupid', Working Paper, *Columbia Center for Law and Economics Studies*, 207.
—— and H. A. Sale (2008). 'Redesigning the SEC: Does the Treasury Have a Better Idea?', Working Paper, *Columbia Center for Law and Economics Studies*, 342.
Committee of Wise Men (2001). 'Final Report of the Committee of Wise Men on the Regulation of the European Securities Markets', Lamfalussy Report, Brussels.
Coval, J. and Thakor, A. V. (2005). 'Financial Intermediation as a Beliefs-Bridge between Optimists and Pessimists', *Journal of Financial Economics*, 75: 535–70.
Da Rin, M. and Hellmann, T. (2002). 'Banks as Catalysts for Industrialization', *Journal of Financial Intermediation*, 11: 366–97.

Datta, S., Iskandar-Datta, M., and Patel, A. (1999). 'Bank Monitoring and Pricing of Corporate Public Debt', *Journal of Financial Economics*, 51: 435–49.

De Bandt, O. and Hartmann, P. (2002). 'Systemic Risk: A Survey', in *Financial Crises, Contagion and the Lender of Last Resort*, C. Goodhart and G. Illing (eds.), Oxford: Oxford University Press.

De Nicoló, G. and Tieman, A. F. (2005). 'Economic Integration and Financial Stability: A European Perspective', Working Paper, International Monetary Fund, 06/296.

Decamps, J., Rochet, J., and Roger, B. (2004). 'The Three Pillars of Basel II: Optimizing the Mix', *Journal of Financial Intermediation*, 13: 132–55.

Degryse, H. and Ongena, S. (2007). 'The Impact of Competition on Bank Orientation', *Journal of Financial Intermediation*, 16: 399–424.

Deidda, L. and Fattouh, B. (2008). 'Banks, Financial Markets and Growth', *Journal of Financial Intermediation*, 17: 6–36.

Dell'Ariccia, G., Detragiache, E., and Rajan, R. (2008). 'The Real Effect of Banking Crises', *Journal of Financial Intermediation*, 17: 89–112.

Dewatripont, M. and Maskin, E. (1995). 'Credit and Efficiency in Centralized and Decentralized Economies', *Review of Economic Studies*, 62: 541–55.

DeYoung, R., Glennon, D., and Nigro, P. (2008). 'Evidence from Informational-Opaque Small Business Borrowers', *Journal of Financial Intermediation*, 17: 113–43.

Diamond, D. (1984). 'Financial Intermediation and Delegated Monitoring', *Review of Economic Studies*, 51: 393–414.

—— (1991). 'Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt', *Journal of Political Economy*, 99: 689–721.

—— (1993). 'Seniority and Maturity of Debt Contracts', *Journal of Financial Economics*, 33: 341–68.

—— and Dybvig, P. H. (1983). 'Bank Runs, Deposit Insurance and Liquidity', *Journal or Political Economy*, 91: 401–19.

—— and Rajan, R. G. (2001). 'Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking', *Journal of Political Economy*, 109: 287–327.

Dinç, I. S. (2000). 'Bank Reputation, Bank Commitment, and the Effects of Competition in Credit Markets', *Review of Financial Studies*, 13: 781–812.

Drucker, S. (2005). 'Information Asymmetries and the Effects of Banking Mergers in Firm–Bank Relationships', Proceedings, Federal Reserve Bank of Chicago, 140–47.

—— and Puri, M. (2005). On the Benefits of Concurrent Lending and Underwriting', *Journal of Finance*, 60: 2763–99.

Economic and Finance Committee (2001). 'Report on Financial Crisis Management', European Economy, Economic Papers 156, Commission of the EC, Directorate-General for Economic and Financial Affairs.

—— (2002). 'Financial Regulation, Supervision and Stability', Economic and Financial Committee document EF76/ECOFIN 324.

Eisenbeis, R. A. and Kaufman, G. G. (2005). 'Bank Crises Resolution and Foreign-Owned Banks', *Federal Reserve Bank of Atlanta Economic Review*, 90: 1–18.

European Central Bank (2003). 'Memorandum of Understanding on High-Level Principles of Cooperation', Press Release.

—— (2005). 'Memorandum of Understanding on Cooperation between the Banking Supervisors, Central Banks and Finance Ministries of the European Union in Financial Crises Situations', Press Release.

Flannery, M. (1998). 'Using Market Information in Prudential Bank Supervision: A Review of the U.S. Empirical Evidence', *Journal of Money, Credit and Banking*, 30: 273–305.

Gande, A. (2007). 'Commercial Banks in Investment Banking', Working Paper, Vanderbilt University.

—— and Saunders, A. (2005). 'Are Banks Still Special When There Is a Secondary Market for Loans?', Working Paper, New York University.

Goodfriend, M. and King, R. (1988). 'Financial Deregulation, Monetary Policy and Central Banking', in *Restructuring Banking and Financial Services in America*, W. Haraf and R. M. Kushmeider (eds.), American Enterprise Institute Studies 481, Lanham, MD: University Press of America.

Gorton, G. B. (1988). 'Banking Panics and Business Cycles', *Oxford Economic Papers*, 40: 751–81.

—— and Kahn, J. A. (1993). 'The Design of Bank Loan Contracts, Collateral, and Renegotiation', National Bureau of Economic Research Working Paper W4273.

—— and Pennacchi, G. (1995). 'Banks and Loan Sales: Marketing Nonmarketable Assets', *Journal of Monetary Economics*, 35: 389–411.

Goyal, V. (2005). 'Market Discipline of Bank Risk: Evidence from Subordinated Debt Contracts', *Journal of Financial Intermediation*, 14: 318–50.

Group of Thirty (2009). 'Financial Reform: A Framework for Financial Stability', Report by the Working Group on Financial Reform, Washington, DC.

Hauswald, R. and Marquez, R. (2006). 'Competition and Strategic Information Acquisition in Credit Markets', *Review of Financial Studies*, 19: 967–1000.

Hellmann, T., Lindsey, L., and Puri, M. (2008). 'Building Relationships Early: Banks in Venture Capital', *Review of Financial Studies*, 21: 513–41.

Hellwig, M. (1991). 'Banking, Financial Intermediation and Corporate Finance', in *European Financial Integration*, A. Giovanni and C. P. Mayer (eds.), New York: Cambridge University Press, 35–63.

Herring, R. J. (2008). 'The U.S. Subprime Crisis: Lessons for Regulators', Proceedings of the 44th Annual Conference on Bank Structure and Competition, Federal Reserve Bank of Chicago, 48–55.

Hoshi, T., Kashyap, A., and Scharfstein, D. (1993). 'The Choice between Public and Private Debt: An Analysis of Post-Deregulation Corporate Financing in Japan', National Bureau of Economic Research Working Paper 4421.

Houston, J. and James, C. (1996). 'Bank Information Monopolies and the Mix of Private and Public Debt Claims', *Journal of Finance*, 51: 1863–89.

Ingves, S. and Lind, G. (1994). 'The Management of the Bank Crisis: In Retrospect', *Sverigs Riksbank Quarterly Review*, 1: 5–18.

Jacklin, C. J. (1987). 'Demand Deposits, Trading Restrictions and Risk Sharing', in *Financial Intermediation and Intertemporal Trade*, E. Prescott and N. Wallace (eds.), Minneapolis: University of Minnesota Press.

James, C. (1987). 'Some Evidence on the Uniqueness of Bank Loans', *Journal of Financial Economics*, 19: 217–35.

Kane, E. J. (1988). 'How Market Forces Influence the Structure of Financial Regulation', in *Restructuring Banking and Financial Services in America*, W. S. Haraf and R. M. Kushmeider (eds.), Washington, DC: American Enterprise Institute Press, 343–82.

Kashyap, A., Rajan, R., and Stein, J. (2002). 'Banks as Liquidity Providers: An Explanation for the Co-Existence of Lending and Deposit-Taking', *Journal of Finance*, 57: 33–73.

Kroszner, R. S. and Rajan, R. G. (1994). 'Is the Glass-Steagall Act Justified? A Study of the US Experience with Universal Banking before 1933', *American Economic Review*, 84: 810–32.

LaPorta, R., Lopez-de-Silanes, L., Schleifer, A., and Vishny, R. W. (1998). 'Law and Finance', *Journal of Political Economy*, 106: 1113–55.

Lummer, S. L. and McConnell, J. J. (1989). 'Further Evidence on the Bank Lending Process and the Reaction of the Capital Market to Bank Loan Agreements', *Journal of Financial Economics*, 25: 99–122.

Mayer, C. (1966). 'New Issues in Corporate Finance', *European Economic Review*, 32: 1167–183.

Mian, A. R. and Sufi, A. (2008). 'The Consequences of Mortgage Credit Expansion: Evidence from the 2007 Mortgage Default Crisis', Working Paper, University of Chicago.

Millon, M. and Thakor, A. V. (1985). 'Moral Hazard and Information Sharing: A Model of Financial Information Gathering Agencies', *Journal of Finance*, 40: 1403–22.

Ongena, S. and Smith, D. C. (2000). 'What Determines the Number of Bank Relationships? Cross-Country Evidence', *Journal of Financial Intermediation*, 9: 26–56.

Partnoy, F. (1999). 'The Siskel and Ebert of Financial Markets: Two Thumbs Down for the Credit Rating Agencies', *Washington University Law Quarterly*, 77: 619–712.

Pearson, P. J. (2003). 'Comment', in *Financial Supervision in Europe*, J. Kremer, D. Schoenmaker, and P. Wierts (eds.), Cheltenham: Edward Elgar, 51–7.

Petersen, M. and Rajan, R. G. (1994). 'The Benefits of Lending Relationships: Evidence from Small Business Data, *Journal of Finance*, 49: 1367–400.

——(1995). 'The Effect of Credit Market Competition on Lending Relationships', *Quarterly Journal of Economics*, 110: 407–43.

Puri, M. (1996). 'Commercial Banks in Investment Banking: Conflict of Interest or Certification Role?', *Journal of Financial Economics*, 40: 373–401.

Rajan, R. G. (1992). 'Insiders and Outsiders: The Choice between Informed and Arm's Length Debt', *Journal of Finance*, 47: 1367–400.

Ramakrishnan, R. and Thakor, A. V. (1984). 'Information Reliability and a Theory of Financial Intermediation', *Review of Economic Studies*, 51: 415–32.

Ramírez, C. (2002). 'Did Banks' Security Affiliates Add Value? Evidence from the Commercial Banking Industry during the 1920s', *Journal of Money, Credit and Banking*, 34: 391–411.

Rochet, J.-C. (2004). 'Bank Runs and Financial Crises: A Discussion', in *Credit Intermediation and the Macro Economy,* S. Bhattacharya, A. W. A. Boot, and A. V. Thakor (eds.), Oxford: Oxford University Press.

——and Vives, X. (2004). 'Coordination Failures and the Lender of Last Resort: Was Bagehot Right After All?', Working Paper, Institut d'Economie Industrielle 294, Toulouse.

Sapienza, P. (2002). 'The Effects of Banking Mergers on Loan Contracts', *Journal of Finance*, 57: 329–67.

Schenone, C. (2004). 'The Effect of Banking Relationships on the Firm's Ipo Underpricing', *Journal of Finance*, 59: 2903–3058.

Sharpe, S. A. (1990). 'Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships', *Journal of Finance*, 45: 1069–87.

Shockley, R. and Thakor, A. V. (1997). 'Bank Loan Commitment Contracts: Data, Theory and Tests', *Journal of Money, Credit and Banking*, 29: 517–34.

Simons, H.C. (1936). 'Rules versus Authorities in Monetary Policy', *Journal of Political Economy*, 44: 1–30.

Song, F. and Thakor, A. V. (2007). 'Relationship Banking, Fragility and the Asset-Liability Matching Problem', *Review of Financial Studies*, 20: 2129–77.

Stein, J. C. (2002). 'Information Production and Capital Allocation: Decentralized versus Hierarchical Firms', *Journal of Finance*, 57: 1891–921.

Stiglitz, J. E. (2009). 'Commentary: How to Rescue the Bank Bailout', *CNNPolitics.com*.

Strahan, P. E. (2007). 'Bank Structure and Lending: What We Do and Do Not Know', Working Paper, Boston College.

Thakor, A. V. (1996). 'Capital Requirements, Monetary Policy and Aggregate Bank Lending: Theory and Empirical Evidence', *Journal of Finance*, 51: 279–324.

US Senate (2002). 'Rating the Raters: Enron and the Credit Rating Agencies', Hearings before the Senate Committee on Governmental Affairs: Washington, DC.

US Senate Staff Report (2002). 'Financial Oversight of Enron: The SEC and Private-Sector Watchdogs', Report of the Staff to the Senate Committee on Governmental Affairs, Washington, DC.

von Thadden, E.-L. (1998). 'Intermediated versus Direct Investment: Optimal Liquidity Provision and Dynamic Incentive Compatibility', *Journal of Financial Intermediation*, 7: 177–97.

CHAPTER 4

RISK MANAGEMENT IN BANKING

LINDA ALLEN
ANTHONY SAUNDERS

Introduction

If you open the vaults of any bank, you might think you would know what you would find there. You would be wrong. What is really there, hidden behind the stacks of currency, is the bank's inventory of risk. The bank exists to take on the risks of its customer base. By offering its clients risk management products, the bank absorbs an inventory of risk that is contributed with each transaction. The bank prices those products by estimating its costs of managing the risks inherent in each transaction. Financial institutions are specialists in risk management. Indeed, their primary expertise stems from their ability both to measure and manage risk exposure on their own behalf and on behalf of their clients—either through the evolution of financial market products to shift risks or through the absorption of their clients' risk into their risk inventory on their own balance sheets. Because financial institutions are risk intermediaries, they maintain an inventory of risk that must be measured carefully so as to ensure that the risk exposure does not threaten the intermediary's

solvency. Thus, accurate measurement of risk is an essential first step for proper risk management, and financial intermediaries, because of the nature of their business, tend to be leading developers of new risk measurement and risk pricing techniques.

When financial institutions misprice risk, however, as was the case of subprime mortgage securities, the size of the risk inventory may overwhelm the financial system's capacity to absorb risk, thereby resulting in global market failures such as the credit crisis of 2007–2009. For example, banks provided equity and backup lines of credit to the Structured Investment Vehicles (SIVs) that they formed during the build-up of the subprime mortgage securitization bubble. The SIV is a structured operating company that invests in assets that are designed to generate higher returns than the SIV's cost of funds. Rather than selling the asset-backed securities directly to investors in order to raise cash (as special purpose vehicles (SPVs) do in standard securitizations), the SIV sells bonds or commercial paper to investors in order to raise cash. The SIV then holds the loans purchased from originating banks on its own balance sheet until maturity. These loan assets held by the SIV back the debt instruments issued by the SIV to investors. Thus, in essence the SIV itself becomes an asset-backed security, and the SIV's commercial paper liabilities are considered asset-backed commercial paper. Investors buy the SIV's liabilities (most often, asset-backed commercial paper (ABCP)), providing the proceeds for the purchase of loans from originating banks. The SIV's debt is backed by the loan portfolio held by the SIV. However, the SIV does not simply pass through the payments on the loans in its portfolio as in a traditional collateralized mortgage obligation (CMO). Indeed, investors have no direct rights to the cash flows on the underlying loans in the portfolio. They are entitled to the payments specified on the SIV's debt instruments. The SIV's ABCP obligations carry interest obligations that are independent of the cash flows in the underlying loan portfolio. Thus, in the traditional form of securitization, the SPV only pays out what it receives from the underlying loans in the pool of assets backing the asset-backed securities. In the newer form of securitization, the SIV is responsible for payments on its ABCP obligations whether or not the underlying pool of assets generates sufficient cash flow to cover those costs. Of course, if the cash flow from the asset pool exceeds the cost of ABCP liabilities, then the SIV keeps the spread and makes a profit. However, if the assets in the underlying pool do not generate sufficient cash flows, the SIV is still obligated to make interest and principal payments on its debt instruments.

The SIV's operating methodology should seem very familiar to bankers. SIVs are banks minus the regulations! The SIV acts similar to a traditional bank—holding loan assets until maturity and issuing debt instruments (such as ABCP) to fund its asset portfolio. The major difference between a SIV and a traditional bank is that the SIV cannot issue deposits to fund its asset base (i.e., it is not technically a 'bank'). However, to the extent that many of these SIVs use

commercial paper and interbank loans (such as repurchase agreements)[1] to finance their asset portfolios, then they may be subject to even more liquidity risk than are traditional banks. This is because, in the modern world of banking, sophisticated lenders (so-called suppliers of 'purchased funds') are prone to 'run' at the first sign of trouble, whereas depositors are slower to react. That is, interbank lenders and commercial paper buyers will withdraw funds (or refuse to renew financing) quicker than traditional 'core' depositors, who may rely on their bank deposits for day-to-day business dealings or may be protected by government deposit insurance. Thus, the well-publicized problems of UK's Northern Rock bank in August of 2007 were precipitated by the withdrawal of funds by interbank lenders and other purchased fund suppliers. Core depositors represented only approximately 25 percent of Northern Rock's funded assets. The liquidity risk problem is exacerbated when the SIV relies on short-term sources of funding, such as commercial paper, which must be renewed within nine months, and repurchase agreements, which must be fully backed by collateral at all points in time. Thus, if the value of the portfolio declines owing to credit conditions worsening, for example, then the SIV may be forced to sell long-term, illiquid assets at firesale prices in order to meet its short-term debt obligations.

Many SIVs were sponsored and originated by banks anxious to remove the risky subprime mortgages (and other obligations) from their balance sheets. The banks and bank regulators believed that these off-balance sheet SIVs posed little risk to the bank itself. However, most of these SIVs had ABCP programs that were backed with bank lines of credit. When the ABCP market seized up during the summer of 2007, the SIVs took down their lines of credit and, all of a sudden, the risks that were believed to be off the balance sheet came back to haunt the banks. The banks were exposed to the risks associated with the poorly underwritten subprime mortgage securities because they were forced to lend to SIVs that had no assets other than these risky securities. Bank shareholders and stakeholders suffered, top executives lost their jobs, global credit markets dried up—all because risk was improperly measured and priced in the mortgage securities market.

Banks are exposed to several major sources of risk. The first, market risk, includes interest rate risk. Thus, if, for example, interest rates increase unexpectedly, the bank's cost of funds may increase and the value of its longer-term, illiquid assets may fall, to the detriment of both the bank's profitability (net interest margin) and the market value of the bank's equity.

A second source of risk is credit risk. Since the most substantial asset classification on the bank's balance sheet consists of loans (whether to businesses, residential

[1] A repurchase agreement allows a bank to borrow against collateral (securities) transferred to a counterparty. This transaction is typically reversed within a short time period—from a week to three months. Moreover, the collateral is marked-to-market on a daily basis.

households, or even sovereign governments), banks face the risk of default or deterioration in the borrower's credit quality.[2] As many of the subprime mortgages in the pools originated in 2005 and 2006 began to show delinquencies as early as one year or less after origination, there were concerns in the market about the credit risk exposure of the securities, despite their AAA and AA credit ratings.

A third source of risk, described above in the context of the 2007–2009 credit crisis, is liquidity risk. Banks transform short-term, liquid liabilities (such as demand deposits) into longer-term, illiquid assets (such as loans). If there is a sudden demand for liquidity, the bank will be unable to meet all withdrawal demands because of the costs of selling an illiquid portfolio at firesale prices.

Another source of risk is operational risk. Banks undertake clearing and custodial transactions on behalf of their customers. Fraud, mismanagement, computer failure, and human error can result in losses to customers, which the bank may have to reimburse in order to protect its reputation. Strategic business errors cause catastrophic losses that may threaten the bank's viability. Loss of reputation may spell the end of the firm's independence for a financial institution—as in the case of the venerable Barings Bank. In the context of the credit crisis of 2007, HSBC absorbed $45 billion in assets from its SIVs in order to protect its reputation in the market. Moreover, firms such as Citi, Merrill Lynch, UBS, Bear Stearns, etc. have all had their reputations tarnished by their participation in the subprime mortgage debacle.

This brief thumbnail survey of risk exposures highlights the importance of measuring the amount of risk in the bank's risk inventory on a continual basis. Thus, before we can even talk about risk management, we first have to discuss risk measurement. The second section of this chapter will describe a commonly used model of risk measurement for banks—VAR. Only after the bank's risk exposure is measured can we discuss how to manage that risk.

What if upon measuring the amount of risk in the bank's risk inventory we find that the exposure is too high from the perspective of top management's risk tolerance? Can banks simply refuse to take on more risk? The answer is no. The business of banking requires that banks stand ready to absorb the risks of their customers at a price. If customers are willing to pay that price, it is bad business practice to refuse it. Customers will be forced to go elsewhere and it may be impossible to win them back. Instead, the bank should continue to take on their customers' risk exposures—whether by making loans with credit risk, or absorbing currency risk by offering import/export firms cross-currency letters of credit, or by executing trust agreements, thereby exposing the bank to operational risk. However, once that risk is placed into inventory, the bank's risk management team can then decide whether to hold that risk or resell it in the global marketplace. This 'risk reselling' is accomplished using financial derivatives. Banks can manage their

[2] For example, in July 1998, Russia defaulted on its debt, followed by Argentina in 2001.

risk inventory using financial futures, forward contracts, options, and swaps. This is a much more efficient way for the bank to manage risk than disappointing long-standing customers. Thus, risk management takes place almost exclusively using derivatives transactions, rather than balance sheet adjustments. After reviewing risk management opportunities available to banks in the derivatives markets, the chapter concludes with a discussion of the ongoing global financial crisis, which is what happens when financial intermediaries fail to manage risk properly.

Risk measurement

Risk measurement has preoccupied financial market participants since the dawn of financial history. However, many past attempts have proven to be impractically complex. For example, upon its introduction, Harry Markowitz's Nobel Prize-winning theory of portfolio risk measurement was not adopted in practice because of its onerous data requirements.[3] Indeed, it was Bill Sharpe who, along with others,[4] made portfolio theory the standard of financial risk measurement in real-world applications through the adoption of the simplifying assumption that all risk could be decomposed into two parts: systematic, market risk and the residual, company-specific or idiosyncratic risk. The resulting Capital Asset Pricing Model theorized that since only undiversifiable market risk is relevant for securities pricing, only the market risk measurement β is necessary, thereby considerably reducing the required data inputs. This model yielded a readily measurable estimate of risk that could be practically applied in a real-time market environment. The only problem was that β proved to have only a tenuous connection to actual security returns, thereby casting doubts on β's designation as the true risk measure.[5]

With β questioned, and with asset prcing in general being at a bit of a disarray with respect to whether the notion of 'priced risk' is really relevant, market

[3] Modern portfolio theory is based on Markowitz's insight that diversification can reduce, but not generally eliminate risk, thereby necessitating a risk–reward guide to portfolio selection. To estimate the efficient investment frontier in a mean-variance world requires data on expected returns, standard deviations of returns and correlations between returns for every possible pair of financial securities. On the occasion of the fiftieth anniversary of the publication of the seminal Markowitz's (1952) paper, Rubinstein (2002) offers an interesting discussion of the development of modern portfolio theory by Markowitz and others.

[4] For example, Sharpe's (1963) paper was followed by Mossin (1968).

[5] Dissatisfaction with the β measure began as early as Douglas (1969), with mounting doubts leading to Roll's 1977 paper. The practitioner world closely followed the academic debate with articles such as Wallace's 1980 'Is Beta Dead?' article. Beta's death knell was sounded by Fama and French's (1992) paper that found that after controlling for firm size and the market to book ratio, the firm's β had no statistically significant power to explain returns on the firm's equity.

practitioners searched for a replacement risk measure that was both accurate and relatively inexpensive to estimate. Despite the consideration of many other measures and models, VAR has been widely adopted. Part of the reason leading to the widespread adoption of VAR was the decision of J.P. Morgan to create a transparent VAR measurement model, called 'RiskMetrics'. RiskMetrics was supported by a publicly available database containing the critical inputs required to estimate the model.[6] Another reason behind the widespread adoption of VAR was the introduction in 1998[7] by the Bank for International Settlements (BIS) of international bank capital requirements that allowed relatively sophisticated banks to calculate their capital requirements based on their own internal models such as VAR.

In the past, many of the risk measurement models were private, internal models, developed in-house by financial institutions. Internal models were used for risk management in its truest sense. Indeed, the VAR tool is complementary to many other internal risk measures—such as risk-adjusted return on capital (RAROC) developed by Bankers Trust in the 1970s.[8] However, market forces during the late 1990s created conditions that led to the evolution of VAR as a dominant risk measurement tool for financial firms.

The US financial environment during the 1990s was characterized by the *de jure* separation of commercial banking and investment banking dating back to the Glass-Steagall Act of 1933.[9] However, these restrictions were undermined in practice by Section 20 affiliates (that permitted commercial bank holding companies to engage in investment banking activities up to certain limits), mergers between investment and commercial banks, and commercial bank sales of some 'insurance' products, especially annuities. Thus, commercial banks competed with investment banks and insurance companies to offer financial services to clients in an environment characterized by globalization, enhanced risk exposure, and rapidly evolving securities and market procedures. Concerned about the impact of the increasing risk environment on the safety and soundness of the banking system, bank regulators instituted (in 1992) risk-adjusted bank capital requirements that levied a capital charge for both on- and off-balance sheet credit risk exposures.

[6] In their introduction, Mina and Xiao (2001) stress that RiskMetrics is not strictly a VAR model, although it can be used to estimate a VAR model. RiskMetrics' critical role in the dissemination of VAR among financial market practitioners stems in large part from the availability of real time data on financial market fluctuations provided freely in the public domain. Recognizing that value added, RiskMetrics has currently formed a separate data service, DataMetrics which covers almost 100,000 data series.

[7] The market risk amendment to the Basel capital requirements was adopted in November 1996 in Europe and in January 1998 in the US.

[8] RAROC models are risk-sensitive measures of economic performance that can be used to allocate risk capital within the firm. See Saunders and Allen (2002: chap. 13).

[9] The Gramm-Leach-Bliley Act of 1999 permitted the creation of financial service holding companies that could include commercial banking, investment banking and insurance subsidiaries under a single corporate umbrella, thereby effectively repealing the Glass-Steagall Act.

Risk-adjusted capital requirements initially applied only to commercial banks, although insurance companies[10] and securities firms had to comply with their own reserve and 'haircut' regulations as well as with market forces that demanded capital cushions against insolvency based on economic-model-based measures of exposure—so-called economic capital. Among other shortcomings of the BIS capital requirements were their neglect of diversification benefits in measuring a bank's risk exposure. Thus, regulatory capital requirements tended to be higher than economically necessary, thereby undermining commercial banks' competitive position *vis-à- vis* largely unregulated investment banks. To compete with other financial institutions, commercial banks had the incentive to track economic capital requirements more closely notwithstanding their need to meet regulatory capital requirements. The more competitive the commercial bank was in providing investment banking activities, for example, the greater its incentive to increase its potential profitability by increasing leverage and reducing its capital reserves.

J.P. Morgan (now JPMorgan Chase) was one of a handful of globally diversified commercial banks that were in a special position relative to the commercial banking sector on the one hand and the investment banking sector on the other. These banks were caught in between, in a way. On the one hand, from an economic perspective, these banks could be thought of more as investment banks than as commercial banks, with large market risks due to trading activities, as well as advisory and other corporate finance activities. On the other hand, this group of globally diversified commercial banks was holding a commercial banking license, and, hence, was subject to commercial bank capital adequacy requirements. This special position gave these banks, J.P. Morgan being a particular example, a strong incentive to come out with an initiative to remedy the capital adequacy problems that they faced. Specifically, the capital requirements for market risk in place were not representative of true economic risk, due to their limited account of the diversification effect. At the same time, competing financial institutions, in particular investment banks such as Merrill Lynch, Goldman Sachs, and Salomon Brothers, were not subject to bank capital adequacy requirements. As such, the capital they held for market risk was determined more by economic and investor considerations than by regulatory requirements. This allowed these institutions to bolster significantly more impressive ratios, such as return on equity (ROE) and return on assets (ROA), compared with banks with a banking charter.

In response to the above pressures, J.P. Morgan took the initiative to develop an open architecture (rather than in-house) methodology, called 'RiskMetrics'. RiskMetrics quickly became the industry benchmark in risk measurement. The publication of RiskMetrics was a pivotal step moving regulators toward adopting economic capital-based models in measuring a bank's capital adequacy. Indeed,

[10] Insurance regulators in the US adopted their own risk-based capital requirements for life and property casualty insurers in the mid- to late 1990s.

bank regulators worldwide allowed (sophisticated) commercial banks to measure their market risk exposures using internal models that were often VAR-based. The market risk amendments to the Basel accord made in-house risk measurement models a mainstay in the financial sector. Financial institutions worldwide moved forward with this new approach and never looked back.

It was Dennis Weatherstone, at the time the Chairman of J.P. Morgan, who clearly stated the basic question that is the basis for VAR as we know it today—**'how much can we lose on our trading portfolio by tomorrow's close?'** Note that this is a risk measurement, not a risk management question. Also, it is not concerned with obtaining a portfolio position to maximize the profitability of the bank's traded portfolio subject to a risk constraint, or any other optimization question. Instead, this is a pure question of risk measurement. VAR takes a statistical or probabilistic approach to answering Weatherstone's question of how much could be lost on a 'bad day'. That is, we define a 'bad day' in a statistical sense, such that there is only an x percent probability that daily losses will exceed this amount given a distribution of all possible daily returns over some recent past period. That is, we define a 'bad day' so that there is only an x percent probability of an even worse day.

Implementing VAR models requires estimation of a probability distribution of returns (or losses) so that we can measure the cut-off point that designates the loss that will be exceeded with an x percent probability on any given day. The simplest forms of RiskMetrics—for example, the Rule 415 model—assume that financial securities are normally distributed. This makes estimation of VAR quite easy because all we have to do is estimate the mean and standard deviation of securities prices using historical data. Unfortunately, it is often the case that the simplicity of the VAR measures used to analyze the risk of the equity portfolio for example is in large part obtained with assumptions not supported by empirical evidence. The most important (and most problematic) of these assumptions is that daily equity returns are normally distributed. In general, there is a trade-off between the accuracy of assumptions and ease of calculation, such that greater accuracy is often accompanied by greater complexity.[11]

This problem of complexity is exacerbated when there is a paucity of data available to be used to estimate the model's fundamental assumptions. Market risk exposure arises from unexpected security price fluctuations, estimated using long histories of daily price fluctuations. Unfortunately, measuring a loan's credit risk exposure, for example, is far more difficult. Since loans are not always traded, and even when traded they trade infrequently, there is often no history of daily price fluctuations available to build a (loan) loss distribution. Moreover, credit events such as default or rating downgrades are rare, often non-reoccurring events.

[11] For specific methodologies used to estimate VAR models, see Allen, Boudoukh, and Saunders (2004).

Thus, we often have insufficient statistical power to estimate a daily VAR for credit risk exposure—that is, data limitations create special challenges in adapting VAR techniques to estimate credit risk exposure. However, we can use VAR techniques to estimate losses due to credit events if the time interval we consider is longer. Indeed, the convention in the new generation of credit risk models is to assume that the credit risk time horizon is one year, thereby estimating losses during the next year if it is a 'bad year', defined according to a specified VAR level—for example, a 99.5 percentile VAR (i.e., x percent equals 0.5 percent) estimates the minimum losses in the worst five years out of a 1,000. A VAR model, such as CreditMetrics, measures the probability that the credit rating of any given debt security will change over the course of the one-year credit horizon. The tabulation of potential changes in credit ratings—known as the credit migration matrix—considers the entire range of credit events, including upgrades and downgrades as well as actual default. Historical migrations of publicly traded debt instruments, such as corporate bonds, are used to tabulate the annual probability of any given change in credit risk. These loss probabilities are then applied to specific debt instruments, such as untraded loans, to calculate the loan portfolio's VAR.

Because of the problems applying the VAR model to credit risk assessment, banks often use other credit risk measurement models. There has been widespread adoption of credit scoring models in all arenas of bank lending—mortgage lending, commercial lending, credit card, and revolving debt, etc. Following Ed Altman's seminal work, credit scoring models (e.g., 'FICO scores'[12]) apply discriminant analysis to a class of borrowers by identifying certain key factors that determine the probability of default (as opposed to repayment), and combine or weight them into a quantitative score. In some cases, the score can be literally interpreted as a probability of default; in others, the score can be used as a classification system: it places a potential borrower into either a 'good' or a 'bad' group, based on a score and a cut-off point.

VAR models are probably best suited to measuring operational risk exposure. VAR measures losses from unexpected, extreme shocks that are in the tail of the probability distribution (i.e., at the end of outcomes that are extremely unlikely to occur). Thus, the probability of a VAR-size event (x percent) is very small (i.e., 5 percent or 1 percent or 0.5 percent). However, when these improbable events occur, they are catastrophic for the firm and typically result in insolvency. Indeed, Allen and Bali (2007) find that operational risk events are likely to be the cause of large unexpected catastrophic losses. They use a comprehensive approach to measuring operational risk that includes reputational risk and strategic-business risk and shows that approximately 18 percent of financial institutions' returns represent compensation for operational risk. In contrast, Basel II mandates a narrow definition of operational risk for regulatory purposes that focuses on

[12] FICO is a registered trademark of Fair Isaac Corporation.

day-to-day loss events emanating from computer failures and human error—for example, while excluding catastrophic operational-risk events resulting from reputational losses and strategic business errors. Although this definitional decision may be warranted on pragmatic grounds (i.e., the absence of reliable industry databases on extreme tail operational loss events), the eventual goal is to develop a more comprehensive measure of operational risk that is more consistent with the designation of regulatory capital as a cushion against unexpected loss. VAR enables banks to accomplish this (for economic capital, if not for regulatory capital purposes) because of the methodology's focus measuring the impact of extremely unlikely, but catastrophic, risk events. Thus, the VAR methodology can be used to measure market risk, credit risk, and operational risk.

Risk management

Suppose that the VAR model implemented by the bank provides a measure of risk exposure that is enormous—even in excess of the bank's capital position. What can be done? The first thing is not to panic. The second is to fire up the derivatives traders. The bank can manage its risk position by trading in derivatives markets. If the initial risk inventory is too high, the bank can undertake hedging transactions to reduce its risk exposure without turning away profitable and long-standing customers. On the other hand, if the initial risk inventory is too low, and therefore not profitable enough, the bank can undertake speculative transactions to increase its risk exposure. Derivatives markets are the thermostat used by the bank to control its risk temperature.

In either extreme, pure hedging or pure speculating, the derivatives transaction is tied to, indeed motivated by, another transaction, or series of transactions that constitute the underlying cash position. The Commodity Futures Trading Commission (CFTC) estimates that up to 85 percent of all futures trades are explicitly linked to other transactions. If the cash flows on the derivatives transaction are opposite to those of the underlying cash position, we consider the derivatives trade to be a hedge. If, on the other hand, the cash flows move in the same direction, we consider the derivatives trade to be speculative. The cash flows on derivatives are determined by fluctuations in interest rates, exchange rates, equity prices, default probabilities, etc. That is, derivatives can be used to manage all types of risk exposure.

Suppose, for example, the bank has an underlying cash position that is exposed to interest rate increases. This is a very common position for a bank and as a result of the process of 'borrowing short to lend long'. Thus, the bank's assets have a longer maturity (duration) than the bank's liabilities, leading to a positive duration

gap. Under these circumstances, the underlying cash position (the bank's portfolio) will decline in value and profitability will fall if interest rates go up. To hedge that risk, the bank can undertake a derivatives position that generates positive (offsetting) cash flows when interest rates go up—that is, a short position. Short positions can be implemented by selling interest rate futures or forwards, buying put options on interest-rate-sensitive instruments and/or buying fixed-for-floating swaps. We examine each of these markets briefly.

Financial futures and forwards

The concept of a forward contract originated in sixteenth-century Japan when landowners raised money by selling rice in advance of delivery to rice merchants. A more formal, exchange-based contract, the precursor to the modern futures contract, originated in the US Midwest during the early nineteenth century. In 1848, some eighty-two merchants met above a flour store on Chicago's South Water Street and formed the Chicago Board of Trade (CBOT). Today, merged with the Chicago Mercantile Exchange (CME), the CBOT trades millions of futures contracts, as well as options and swaps.

Financial futures or forwards are obligations to make (sell) or take (buy) delivery of some underlying financial asset at a predetermined price (i.e., futures or forward price) on a specified delivery date. The counterparty that buys the contract agrees to buy the underlying financial asset and holds a long position. The counterparty that sells the contract is obligated to sell the underlying financial asset and holds a short position. The long position gains if the price upon delivery date is higher than the predetermined price, whereas the short position gains if the price declines below the predetermined price. Typically, there is no actual delivery of the underlying financial asset in financial futures/forward contracts (in contrast to commodity futures/forwards). Instead of physical delivery, the contracts are usually cash settled, with the losing party paying the winning party for the difference between the spot price upon delivery minus the predetermined futures/forward price.

For example, if the bank has a positive duration gap and wants to hedge its exposure to rising interest rates, it may take a short position in an interest rate-futures contract, such as the US Treasury bills futures contract or the three-month Eurodollar futures contract.[13] If interest rates go up, the price of the contract falls and the short (selling) counterparty gains. For each basis point increase in interest rates, the Treasury bill and Eurodollar futures contracts gain $25 per $1 million face

[13] The Eurodollar CD is not related to the currency named the euro. Eurodollar CDs refer to US dollar-denominated deposits held by banks outside of the US or in international banking facilities within the US. LIBOR is the offer rate on interbank loans of Eurodollar deposits. See Allen (1997: chap. 12).

value. This cash inflow would offset some (or all) of the losses on the underlying cash position emanating from the bank's positive duration gap.[14]

Banks can hedge interest rate risk, currency risk, equity price risk, commodity risk, credit risk, and operational risk using futures and forward contracts. The methodology is the same as illustrated above—that is, short futures/forwards positions hedge underlying cash exposures to price declines and long futures/forwards positions hedge underlying cash exposures to price increases. The only difference is the identity of the reference security. Thus, when hedging currency risk, the reference security's value must fluctuate with shifts in foreign exchange rates. When hedging credit risk, the derivative's underlying security fluctuates with shifts in default risk.

Financial options

Financial futures and forwards are useful tools to protect an underlying cash position from losses due to risk exposure. However, because of their symmetric cash flow payout, they also protect an underlying cash position from gains. That is, when the positive duration gap bank puts on a short-futures position, and interest rates decline rather than increase, the bank's portfolio will make money, but the hedge will lose money. Thus, there was a demand for a hedging instrument that would protect against losses, but not against gains—that is, an insurance policy against losses. This insurance policy is an 'options contract'.

An options contract is a derivative that gives the holder the right, not the obligation, to buy (call option) or sell (put option) an underlying reference financial asset at a predetermined price (the exercise or strike price) for a time period up until the specified expiration date.[15] The buyer (holder) of the option retains the right to exercise the option if it is worthwhile. That is, if the holder has a call option, they will benefit when prices increase above the exercise price. If prices do not exceed the exercise price at expiration date, the option expires worthless. Thus, if the bank wants to use an option to hedge its exposure to rising interest rates, it would purchase a put option on an interest-rate-sensitive instrument (such

[14] If all of the losses are hedged, we consider that a 'perfect' or 'naïve' hedge. In practice, we do not observe such hedges because (1) they are difficult to get exactly right and (2) they are undesirable since while a 'perfect' futures/forward hedge eliminates all possibility of loss, it also eliminates all possibility of gain.

[15] An American option can be exercised at any time up until expiration date, whereas a European option cannot be exercised prior to expiration date. Unless are interim cash flows (such as dividend payments), it would not desirable to exercise an American option prior to expiration since the option is worth more alive than dead because of its time value. Therefore, in practical terms, there is no difference between American and European options on financial securities with no interim cash flows (e.g. zero coupon bonds).

as a Eurodollar futures contract),[16] which would generate positive cash flows if interest rates increase (and prices fall), thereby offsetting the bank's loss due to its underlying cash position with a positive duration gap. If, however, interest rates decline, the positive duration gap bank generates positive cash flows and the option hedge expires worthless, thereby allowing the bank to keep its gains.

The exception to this is that the options buyer must pay an upfront cost—the premium—which is non-refundable to the buyer if the option expires worthless.[17] Options premiums are quite substantial. Therefore, we have seen the development of compound options positions, such as straddles, collars, butterflies, etc. that were originated in order to reduce the upfront premium cost of options trades. As market participants experimented with these 'lower-cost options hedges', however, they found that they were viable products in their own right. Therefore, today, collars are sold as stand-alone risk management products to the customers of financial institutions. Alternatively, they can be packaged with other financial products, as in adjustable rate mortgages that contain collars.

Swaps

It was August 1981. The US dollar was entering a period of strength against European currencies. In 1979, IBM had issued debt denominated in Swiss francs and Deutschmarks, in the course of its regular financing program. With the increase in the dollar, the dollar cost of IBM's liabilities declined significantly. IBM could realize a significant cash inflow if only the liabilities could be repurchased and converted into US dollars. But the retirement of debt at a discount would expose IBM to a considerable tax liability. Moreover, in the European bearer-bond market, it would have been difficult for IBM to find the bonds for repurchase. It seemed that the opportunity would pass IBM by.

Enter the World Bank. The World Bank typically borrows in all major currencies to finance its activities. Because of the upheaval in the European currencies, the World Bank was concerned that future borrowing would soak up the credit available in those markets. How could the World Bank borrow Swiss francs and Deutschmarks without competing with other borrowers?

Enter Salomon Brothers, who saw the opportunity to match the needs of IBM and the World Bank. IBM wanted to replace Deutschmark and Swiss franc

[16] In general, financial options on futures contracts tend to be more liquid than financial options on cash instruments. Thus, for example, we see more activity in the market for US Treasury bill futures options than in the market for US Treasury bill options.

[17] In contrast, futures contracts require an upfront margin (paid to the exchange's clearing corporation, which acts as third-party guarantor) which is a good faith deposit and is refunded to the contract holders (both buyer and seller) upon fulfillment of their obligations under the futures contract. Because forward markets are limited to financial intermediaries with reputations to uphold, there is no margin or third-party guarantor in the forwards market.

borrowings with US dollar borrowings. The World Bank wanted those Deutschmark and Swiss franc borrowings and was willing to borrow US dollars in order to avoid disrupting the European debt markets. The synergies were obvious, once someone pointed them out, and a new financial instrument was born—the cross-currency swap.

It did not take long for financial market professionals to see the extensions—to fixed-for-floating rate swaps (to hedge interest rate risk) and credit default swaps (to hedge credit risk exposure). A swap is essentially a portfolio of forward contracts with predetermined payment dates, called 'reset' dates, and predetermined prices. In a fixed-for-floating rate swap, for example, the buyer of the swap exchanges floating rate payments (say, tied to London Interbank Offered Rate (LIBOR)) for fixed rate payments. If interest rates increase, the swap buyer gains because instead of paying the higher LIBOR payments, the swap buyer pays the lower, predetermined fixed rate. Thus, the positive duration gap bank can purchase fixed-for-floating rate swaps in order to hedge its exposure against increasing interest rates.

Upon reset dates (which can occur monthly, quarterly, semi-annually, annually) for the life of the swap (which can last for up to five or ten years), the swap intermediary calculates the payments required, nets them out, and supervises the transfer of the net cash flow (the difference between the fixed and floating rate as of the reset date) between the counterparties. Thus, if interest rates have increased, the swap seller pays the swap buyer an amount equal to the difference between the fixed rate minus the floating rate times the notional value of the swap, and vice versa if interest rates have declined. The swap intermediary also acts as the guarantor to insure that each swap counterparty meets its obligations. In exchange for setting up the transaction, monitoring its cash flows, and guaranteeing the counterparty credit risk, the swap intermediary receives a fee that is paid on each reset date.

In recent years, there has been an explosive growth in the use of credit derivatives. BIS data for 2001 show that the market for interest rate derivatives totaled $65 trillion (in terms of notional principal), foreign exchange rate derivatives totaled $16 trillion, and equities almost $2 trillion. In contrast, estimates in June 2001 put the market for credit derivatives at approximately $1 trillion in notional value worldwide. However, the BIS reported the notional amount on outstanding 'over-the-counter' (OTC) 'credit default swaps' (CDSs) to be $28.8 trillion in December 2006, up $14.9 trillion (107 percent) from December 2005.[18] It is clear that the market for credit derivatives has grown, and continues to grow, quite rapidly.

The dominant credit derivative to date has been the CDS. Rule (2001) cites a British Bankers' Association survey that found that 50 percent of the notional value of all credit derivatives were CDSs, as compared to 23 percent collateralized

[18] See Bank for International Settlements, 2006d < http://www.bis.org>.

loan obligations, 13 percent credit linked notes, 8 percent baskets,[19] and only 6 percent credit spread options. There are two main types of credit default swap: (1) total return swap and (2) pure credit or default swap.

A total return swap involves swapping an obligation to pay interest at a specified fixed or floating rate for payments representing the total return on a loan or a bond. For example, suppose that a bank lends $100 million to a manufacturing firm at a fixed rate of 10 percent. If the firm's credit risk increases unexpectedly over the life of the loan, the market value of the loan will fall. The bank can seek to hedge an unexpected increase in the borrower's credit risk by entering into a total return swap in which it agrees to pay a counterparty (say, an insurance company) the total return based on an annual rate equal to the promised interest (and fees) on the loan, plus the change in the market value of the loan. Using the total return swap, the bank agrees to pay a fixed rate of interest annually, plus the capital gains or losses on the market value of the loan over the period of the swap. In return, the bank receives a variable market rate payment of interest annually—for example, the one-year LIBOR—from the insuring counterparty.

If the loan decreases in value over the payment period, the bank pays the insurance company a relatively small (possibly negative) amount equal to the fixed payment on the swap minus the capital loss on the loan. For example, suppose the loan was priced at par (100) at the beginning of the swap period. At the end of the swap period (or on the first payment date), the loan has an estimated market value of 90 (90 cents on the dollar) because of an increase in the borrower's credit risk. Suppose that the fixed rate payment as part of the total return swap is 12 percent. The bank would send to the insurance company (swap counterparty) the fixed rate of 12 percent minus 10 percent (the capital loss on the loan), or a total of 2 percent, and would receive in return a floating payment (e.g., LIBOR, say equal to 11 percent in this hypothetical example) from the CDS seller. Thus, the net profit on the swap to the bank/lender is 9 percent (11 percent minus 2 percent) times the notional amount of the swap contract. This gain can be used to offset the loss of market value of the loan over that period. Thus, the seller of credit protection (the insurance company, in this example) would pay the buyer of credit protection (the bank) when there is a credit event (in this example, the 10 percent decline in loan value due to the increase in the borrower's credit risk exposure). If there is no credit event, then the insurance buyer (the bank) simply pays the CDS seller a premium equal to 1 percent in this example (the 12 percent fixed rate minus the 11 percent LIBOR).

[19] Source: Rule, 2001: 117–40. Baskets are credit derivatives based on a small portfolio of loans or bonds, such that all assets included in the underlying pool are individually listed. In contrast, the contents of larger portfolios are described by their characteristics. A basket credit default swap, also known as a first-to-default swap, is structured like a regular CDS, but the reference security consists of several securities. The first reference entity to default triggers a default payment of the par value minus the recovery value and then all payments end.

Total return swaps can be used to hedge credit risk exposure, but they contain an element of interest rate (or market) risk as well as credit risk. For example, in the previous example, if the LIBOR changes, then the *net* cash flows on the total return swap will also change, even though the credit risk of the underlying loans has not necessarily changed. Moreover, if the price of the loan changes due to interest rate or liquidity risk considerations, then the payout on the total return swap will also be affected even if there is no change in the borrower's credit risk exposure.

To strip out the interest rate-sensitive element of total return swaps, an alternate swap, called a 'pure' credit or default swap, has been developed and has dominated the market for credit protection. The CDS is characterized by the following terms:

- The identity of the reference loan (i.e., the notional value, maturity, and the credit spread (over LIBOR) on a risky loan issued by the reference obligor).[20]
- The definition of a credit event (usually any one of the following: bankruptcy, prepayment, default, failure to pay, repudiation/moratorium, and restructuring).
- The compensation that the protection seller (e.g., the insurance company) will pay the protection buyer (e.g., the bank) if a credit event occurs.
- Specification of either physical settlement (delivery of agreed debt instruments) or cash settlement. Early credit swaps were cash settled, but now physical delivery is the most common settlement method. Physical delivery is preferred because it gives the CDS seller more time to recoup the settlement payment through recovery of the value of the reference loan. However, the borrower's consent may be needed to transfer the loan if the credit derivative specifies physical delivery upon occurrence of a credit event.

The protection buyer on a CDS (say, the bank lender) will send (in each swap period) a fixed fee or payment (similar to a premium on an insurance policy or an option) to the protection seller swap counterparty. If the CDS reference loan (or loans) does not default, the protection buyer will receive nothing back from the swap counterparty. However, if the loan (or loans) defaults, the CDS seller will cover the default loss by making a default payment equal to the par value of the original loan minus the secondary market value of the defaulted loan. For example, if the loan's price falls to \$40 upon the borrower's default, then the insurance company selling the CDS will pay the bank CDS buyer \$60.[21] Thus, the CDS pays out par minus the recovery value of the loan in the event of default.

[20] Both the obligor and the specific reference debt instrument must be specified. The reference instrument is usually a senior unsecured debt obligation, although CDS can be written on subordinated debt as well.

[21] Default payments are usually computed in one of three ways: (1) par minus a final loan price as determined by a poll of dealers (such as Creditex and CreditTrade); (2) payment of par by the counterparty in exchange for physical delivery of the defaulted loan; and (3) a fixed-dollar amount contractually agreed to at the swap origination. Increasingly, method (2) is the favored method of settlement because of the difficulty in getting accurate secondary market prices on loans around credit event dates.

A pure credit default swap is similar to buying credit insurance and/or a multi-period credit option. The growth in trading of these credit derivatives has facilitated a net overall transfer of credit risk from banks to non-banks, principally insurance companies. Banks, securities firms, and corporations are net buyers of credit protection, whereas insurance companies, hedge funds, mutual funds, and pension funds are net sellers. Insurance companies view credit derivatives as an insurance product, in which their relatively high credit ratings can be used to insure the buyers of credit protection (e.g., banks) against risk exposure to their loan customers. Credit derivatives such as CDSs allow a bank to alter the risk–return trade-off of a loan portfolio without having to sell or remove loans from the balance sheet. Apart from avoiding an adverse customer relationship effect, the use of credit derivatives (rather than loan sales or other portfolio methods for reducing the bank's credit risk exposure) may allow a bank to avoid adverse timing of tax payments, as well as liquidity problems related to buying back a similar loan at a later date if risk–return considerations so dictate. Thus, for customer relationship, tax, transaction cost, and liquidity reasons, a bank may prefer the credit derivative solution to loan portfolio optimization rather than the more direct (loan-trading) portfolio management solution.

The 2007–9 global financial crisis

The economy relies on financial institutions to act as specialists in risk measurement and risk management. The importance of this is demonstrated by the aftermath of the banks' failure to perform this critical function: the 2007–8 global financial crisis resulting in the worldwide breakdown in credit markets, as well as an astonishing level of equity market volatility. When banks fail to perform their critical risk measurement and risk management functions, the result is a crisis of confidence that paralyzes the entire economy. Even overnight, credit markets seize up. Banks are unwilling to lend to other banks because of uncertainty about their own and their competitors' precarious financial condition. This hoarding of liquidity prevents banks from providing the fundamental credit required to keep businesses moving. Commercial-paper and other debt markets cease to function, thereby leaving even creditworthy firms without a source of either working capital or investment capital. These firms are then unable to do business and the financial contagion is transmitted to the real economy in terms of loss of jobs and declines in economic activity.

As we write this, we are in the middle of this dismal chain of events. While we cannot pretend to know how it will end, it is instructive to understand the phases

of the process. Although it may appear that we have been embroiled in one, seemingly never-ending crisis, in reality there have been different stages that differentiate the 2007–8 crisis (to date) into three separate phases: (1) the initial credit risk crisis; (2) the subsequent liquidity risk crisis; and (3) the realization of an operational risk crisis.

The first phase began in the beginning of the year 2007 with the realization of rising delinquencies on subprime residential-mortgage-backed securities (RMBS). As of February 2007, the percentage of subprime mortgage-backed securities delinquent by 90 days or more was 10.09 percent, up from 9.08 percent in October 2006, which was substantially higher than the 5.37 percent rate in May 2005.[22] The second largest subprime lender, New Century Financial, was hit by large number of mortgage defaults, and filed for bankruptcy on 2 April 2007, after it was unable to meet its lenders' calls for more collateral on its credit lines. Bear Stearns High-Grade Structured Credit Master Fund (the investment vehicle for four Bear Stearns hedge funds heavily invested in subprime CMOs, CLOs, and CDOs) and Dillon Read Capital Management, a subsidiary of UBS (Union Bank of Switzerland), both experienced substantial losses during the spring of 2007 and were ultimately closed down several months later.

The roots of these credit problems can be found in the overheated market conditions that characterized the housing bubble. While it is difficult to date the genesis of the 2007 credit crisis, the preconditions for such a crisis were building from 2001 and, in particular, the events of 9/11. In fact, the immediate response by regulators to 9/11 was to create stability in the financial markets. For example, the Federal Reserve lowered the short-term money market rate that banks and other financial institutions pay in the Federal funds market, the market for overnight borrowings among major banks, and even made LOLR funds available to non-bank financial institutions such as investment banks. This had the immediate effect of lowering short-term borrowing rates for other market instruments, such as short-term borrowings of dollars abroad (LIBOR). In fact, very soon, nominal short-term rates fell to close to 1 percent—historically low levels. Perhaps not surprisingly, given low interest rates and the increased liquidity provided by Central banks, such as the Federal Reserve, there ensued a rapid expansion in borrowing or debt levels in the economy, both among consumer borrowers and commercial borrowers. Thus, consumer demand for mortgages and credit card debt ballooned. Moreover, commercial demand for loans increased, and it became increasingly less expensive for private equity firms to undertake takeovers financed via commercial loans (often in the form of syndicated bank loans).

However, what is important is that it was not just the *quantity* of consumer and commercial debt that increased but also the *quality* of debt simultaneously declined. Specifically, as the demand for mortgage debt grew, especially among

[22] See Shenn (2007).

those who had previously been excluded from participating in the market because of their poor credit quality, banks and other financial institutions began lowering their credit quality cut-off points. Moreover, to boost their earnings in the relatively new area of the market now popularly known as the 'subprime market', banks and other mortgage-supplying institutions often offered relatively low 'teaser' rates or adjustable rate mortgages (ARMs) at exceptionally low initial interest rates, but with substantial step-up rates after the initial rate expired and if market rates rose in the future, in addition to low-documentation or no-documentation loans, now known as 'liar loans' because they did not verify borrowers' claims. Under the traditional banking structure, banks might have been reluctant so aggressively to court low-credit quality borrowers for fear that the loans would default prior to maturity. However, asset securitization and loan syndication allowed banks to retain little or no part of the loans they originate. Thus, banks were able to pass along the risk without performing their fundamental risk measurement and management functions.

The bursting of the bubble occurred with the downturn in US housing prices during the second half of the year 2006. Most importantly, the geographic impact of the subprime mortgage crisis spread across the US, thereby undermining the geographic diversification assumptions used in constructing the subprime asset-backed mortgage pools. At around the same time as housing prices began to fall, the Federal Reserve started to raise interest rates in the money market as it began to fear inflation. Since many subprime mortgages originated in the 2001–5 period had floating rates (i.e., were ARMs) with high step-up rates, the cost of meeting mortgage commitments rose to unsustainable levels for many low-income households. The result was a dramatic increase in delinquencies and defaults. This was the credit risk phase of the crisis.

The second phase began in the late summer of 2007 as the crisis spread to the banking community and to the money markets. The British bank Northern Rock and the German bank IKB experienced runs and were bailed out by their respective regulatory agencies. The major US subprime mortgage lender, Countrywide, announced in August that it was drawing down on backup lines of credit because of its growing losses. Ultimately, a liquidity run on Countrywide was stemmed only after a $2 billion equity investment by Bank of America on 23 August 2007. However, a number of ABCP issuers also began having difficulty refinancing their short-term commercial paper issues because of investors' concerns about the quality of the underlying collateral of subprime mortgages and other assets, despite the AA or AAA ratings these issues may have received from the rating agencies.

This phase can be viewed as the liquidity risk phase. Liquidity hoarding forced overnight interbank rates to astronomical levels. Even at those levels, however, banks were reluctant to part with liquidity and lend to one another. Banks refused to support auctions and there were failures in the Auction Rate Securities markets,

spreading the crisis to municipal and corporate borrowers. Investors began to lose confidence in the quality of credit ratings and the rating agencies—an AAA security did not mean what it appeared to suggest.[23] Thus, all other debt issues—from the interbank market to the corporate bond market, including the so-called investment grade market—were negatively affected by a 'flight to quality'. A flight to quality implies a move away from privately issued debt to default-risk-free securities issued by the government, such as US Treasury securities. This resulted in falling prices (rising interest rates or 'credit spreads') on privately issued debt securities and rising prices and lower rates on government-issued securities. Thus, the crisis spread to the corporate debt market, and credit markets throughout the world suffered from high spreads and drastically curtailed liquidity.

It is at this stage in the crisis that the vulture funds and workout specialists typically hunt for bargains to purchase securities selling at depressed prices. During November to December 2007, there was evidence that this process had begun, thereby firming up some debt prices and lowering credit spreads. Sovereign wealth funds made investments in prominent investment and commercial banks. It appeared that the market was finding a bottom. That was when the third, and perhaps most devastating, phase hit. The year 2008 opened to news that a low-level employee of Société Générale, Jerome Kerviel, managed to run up almost €5 billion of losses without being detected by the bank's state-of-the-art internal risk measurement systems. One sophisticated bank after another announced enormous write-downs, only to find that additional huge write-downs were required weeks after the 'final' damage was announced. For example, after reassuring markets in February 2008 that it had only minimal losses, Credit Suisse was forced to announce a $2.85 billion write-down that had somehow been overlooked because the traders didn't properly value the securities in the portfolio.

The credibility of the banks and the financial community was undermined by a string of 'rogue' traders with eye-popping losses, the poor state of the due diligence conducted by underwriters of asset-backed securities, credit default swap holders who do not know the identity of their counterparties to extract payment, liens against property in mortgage pools that are not perfected, and the Keystone Cops quality to the write-downs (ironically called 'death by a thousand cuts' by a Merrill Lynch analyst). It was becoming apparent to the market that the banks themselves did not know the value of the dodgy securities on their own books. The securitizations were done so hastily, without the proper due diligence or legal protections, that working out the loans would be difficult and time consuming. Operational corners were cut when the deals were originally done, thereby imposing operational

[23] In 2006, Moody's earned 44% of its revenues from rating structured finance deals—see Tomlinson and Evans, 2007. Thus, the rating agencies may have had been disinclined to scrutinize the quality of the loans in the ABS, thereby contributing to the large number of defaults on highly rated securities.

risk on the entire market. Thus, the vulture funds and workout specialists pulled back. When the originators themselves do not know the contents (never mind the valuation) of a security, no one can step in with a bid to put a floor under the market—so it just keeps falling. This is what happens in an operational risk crisis. The result was that the credit markets tightened up even further than they had during the first two phases of the crisis.

Indeed, this third operational risk phase has proven to be the most devastating phase of the crisis. It has been during this third phase that the crisis has spread to the general economy, causing an overall decline in economic activity and increasing the likelihood and severity of a global recession. The economy cannot operate without functioning financial markets, and financial markets cannot operate without confidence in the banks' ability to measure and manage risk. Going back and resolving operational omissions will have to be done on an individual security basis. This will take time—something that is in short supply when financial markets are stressed and fear is rampant. Thus, on 3 October 2008, the US Congress passed a package permitting the Treasury to buy up to $700 billion of distressed securities from the banks in order to remove this overhang from the market. The US government presumably will spend the time needed to resolve the operational details that were left unfinished in the deals hastily arranged during the bubble period.

The crisis of 2007–9 demonstrates that we still have a lot to learn about risk measurement and risk management. However, no system will be effective if financial institutions ignore the warning signals flashed by their risk measurement models in their rush to join in the latest market frenzy—whether it is subprime mortgage-backed securities, high-tech, international government securities, or whatever will be the next mania. Risk measurement and management requires a steady eye and a firm hand as well as effective quantitative and analytical tools.

References

Allen, L. (1997). *Capital Markets and Institutions: A Global View,* New York: John Wiley and Sons.

—— and Bali, T. (2007). 'Cyclicality in Catastrophic and Operational Risk Measures', *Journal of Banking & Finance,* 31: 1191–235.

—— Boudoukh, J., and Saunders, A. (2004). *Understanding Market, Credit and Operational Risk: The Value at Risk Approach,* Oxford: Blackwell Publishing.

Bank for International Settlements (2006). 'OTC Derivatives Market Activity in the Second Half of 2006' <http://www.docstoc.com/docs/1029087/OTC-derivatives-market-activity-in-the-second-half-of-2006>.

Douglas, G. W. (1969). 'Risk in Equity Markets: An Empirical Appraisal of Market Efficiency', *Yale Economic Essays*, 9: 3–4.

Fama, E. F. and French, K. R. (1992). 'The Cross Section of Expected Stock Returns', *Journal of Finance*, 47: 427–65.

Markowitz, H. (1952). 'Portfolio Selection', *Journal of Finance*, 7: 77–91.

Mina, J. and Xiao, J. Y. (2001). *Return to RiskMetrics: The Evolution of a Standard*, New York: RiskMetrics.

Mossin, J. (1968). 'Optimal Multiperiod Portfolio Policies', *Journal of Business,* 41: 215–29.

Roll, R. (1977). 'A Critique of the Capital Asset Theory Tests: Part I: On Past and Potential Testability of the Theory', *Journal of Financial Economics*, 4: 129–76.

Rubinstein, M. (2002). 'Markowitz's Portfolio Selection: A Fifty-Year Retrospective', *Journal of Finance*, 57: 1041–5.

Rule, D. (2001). 'The Credit Derivative Market: Its Development and Possible Implications for Financial Stability', *Bank of England Financial Stability Review*, June, 117–40.

Saunders, A. and Allen, L. (2002). *Credit Risk Measurement: New Approaches to Value at Risk and Other Paradigms,* 2nd edn. New York: John Wiley and Sons.

Sharpe, W. F. (1963). 'A Simplified Model for Portfolio Analysis', *Management Science,* 9: 277–93.

Wallace, A. (1980). 'Is Beta Dead?', *Institutional Investor*, 14: 22–30.

CHAPTER 5

LIQUIDITY PRODUCTION IN TWENTY-FIRST-CENTURY BANKING

PHILIP E. STRAHAN

Introduction

Commercial banks produce credit and they provide liquidity. Credit involves channeling resources from entities with excess funds (savers) to entities with a scarcity of funds (investors). Many institutions produce credit. Banks collect savings from depositors and lend the funds to firms and households; finance companies collect funds in the commercial paper market and lend (or lease) the funds to various investors; insurance companies collect premiums and purchase stocks, bonds, commercial paper, and other securities.

Defining liquidity is more difficult. Brunnemeier and Pederson (2007) contrast 'funding liquidity' with 'market liquidity'. Funding liquidity involves raising cash on short notice. Brunnemeier and Pederson model trading in securities markets, so for them funding liquidity describes traders' ability to raise cash with securities as collateral (i.e. margin requirements). But the concept is quite general. When a

depositor withdraws cash or a firm borrows from a credit line, for example, the issuing bank has supplied funding liquidity to the customer. Market liquidity, in contrast, describes the cost of selling assets. Market liquidity is high when prices net of all transactions costs approximate fundamentals (present values). Real estate brokers produce market liquidity for houses. Broker-dealers (i.e. investment banks) produce market liquidity for stocks, bonds, and derivatives, both as underwriters in primary markets and as market makers and traders in secondary markets.

Banks provide both funding liquidity and market liquidity. Traditional intermediation—making illiquid loans funded with liquid deposits—involves production of funding liquidity. Like investment banks, commercial banks also provide market liquidity in their role as market makers in derivatives markets. Banks also create market liquidity in loans. For example, in securitization banks transform pools of illiquid loans, such as mortgages or credit card receivables, into liquid securities. Securitization only flourishes in the presence of a deep securities market. Banks' function in securitization is similar to the function played by investment bankers when they underwrite debt and equity for non-financial companies.

In this chapter I consider how banks provide funding liquidity and market liquidity, and describe how these roles have evolved. With provision of both sorts of liquidity, banks face unique risks and risk management challenges. I discuss how banks meet those challenges. I argue that banks have a special advantage in managing *funding* liquidity risk but not *market* liquidity risk. Hence, many institutions provide market liquidity, while banks dominate in producing funding liquidity. Their comparative advantage stems from the structure of bank balance sheets as well as their access to government guarantees and central bank liquidity. This advantage became especially clear during the 2007–8 financial crisis, discussed below, when the large stand-alone investment banks in the US all either failed, were purchased, or converted to bank holding companies.

In the end, I argue that liquidity production has always been, and continues to be, the core function of banking, but its form has changed in response to the development of financial technology and deepening of securities markets. Traditional banks issued liquid deposits to finance illiquid loans. In this 'old-school' model, most bank liquidity production came in the form of issuing transactions deposits and supporting the payments system. Modern banks continue to run the payments' system, but the rise of electronic mechanisms has made this function more efficient and reduced the float available to finance lending. At the same time, securities markets have expanded and deepened. Thus, much of the liquidity production supported by the banking system now occurs (in the US) or probably will occur (in developing economies) in support of the capital markets. For example, in the form of off-balance sheet commitments to lend (e.g. lines of credit), credit guarantees (e.g. letters of credit), securitization (and the associated implicit support from originating banks), and syndicated lending. These modern

roles have changed because financial innovations now allow funding to be separated from the information production and financial contracting necessary for credit production.[1]

Funding liquidity

Liquidity production from deposits

Banks have traditionally provided funding liquidity to customers by issuing transactions deposits that act as a close substitute for currency. Transactions deposits allow account holders to take cash on demand from the bank. The float from this business—the average balance depositors hold in their accounts—can be invested by the bank in loans to businesses and households. Because banks tend to invest in illiquid loans, this business model has been called 'asset transformation'—banks *transform* illiquid and hence high-yield assets (loans) into liquid and thus low-yield assets (deposits). The yield spread creates positive carry for the bank. Loans are illiquid because banks lend to small and medium-sized businesses without access to broad securities market. To do such lending, banks collect private information on credit risk and future-growth opportunities and monitor borrowers over the life of the loan.

The early theory emphasized how information and monitoring solve financial contracting problems in bank lending (e.g., Leland and Pyle, 1977). As the intermediary, the bank pools funds from many small and uninformed depositors and lends on their behalf. Given the relatively limited information of these investors, banks tend to be financed with debt (Townsend, 1979). Because banks monitor loans on behalf of depositors, Diamond (1984) argues that they will be large and diversified. This structure minimizes the cost of delegating the monitoring role from the principal depositor to the agent banker. Safety and soundness improves the banker's incentives. So, theory suggests that loans are illiquid because the originating bank has superior information to any potential buyer, and that banks are large, well-diversified, and financed mainly with debt.

In contrast to loans, deposits are low in risk and high in liquidity. Bank deposits possess the three attributes theorists ascribe to money—namely, they act as a store

[1] This chapter will not survey the literature on liquidity and banking. The literature is a vast, with many hundreds of papers focusing on bank runs, contagion, and financial crises. These papers span finance, industrial organization, and macroeconomics. It is simply too much to review here. I will briefly discuss some classic treatments of these topics, but will focus most of the discussion on the current state of liquidity production and liquidity risk that we see emerging as banks change to keep pace with the development of capital markets.

of value, they are denominated in the economy's unit of account (e.g. dollars for the US), and they can be used as a medium of exchange. In order for deposits to act as an effective store of value, banks must minimize the risk to depositors that their claims will not be honored at face value; hence low risk and high liquidity go hand in hand.

Banks also spend substantial resources to enhance customers' ability to use deposits as a medium of exchange. For example, demand depositors use checks as a medium of exchange. In recent years, electronic-payments technologies offer a cheaper alternative to cash or checks in both large and small transactions. Customers can use debit cards to transfer funds electronically from the buyer's deposit account to the seller's without losing interest before making the payment. Credit card transactions work similarly, although the buyer maintains more flexibility by having the option to borrow funds from the issuing bank.

Liquidity production from loans

Banks also produce funding liquidity by issuing lines of credit (sometimes known as 'revolvers' or 'loan commitments'), which allow customers to receive cash by drawing down the line, much as a demand deposit allows customers to take cash at any time. The difference between a demand deposit versus a line of credit—such as a credit card account—is that the line is not pre-funded. Instead, the customer borrows from the bank when they take cash, typically at a pre-arranged rate of interest. Households use unsecured lines in the form of credit card accounts and secured lines in the form of home equity lines. Businesses typically also receive liquidity from banks in the form of credit lines.

The appeal of lines of credit has been studied theoretically from the standpoint of businesses. For example, credit lines mitigate the risk of credit rationing for businesses during downturns (Berger and Udell, 1992; and Morgan, 1998), and they insure firms against shocks to demand for capital (Holmstrom and Tirole, 1997). Firms can also use cash, but Sufi (2007) shows that firms with high cash flow use bank lines as a cheaper source of liquidity. Firms with low cash flow are unable to access bank liquidity on good terms and as a result hold cash instead. Sufi also finds that firms without bank lines adjust their buffer of cash to changes in cash flow, while firms with access to bank lines do not. Together these results suggest that bank lines mitigate liquidity constraints on firms by providing them access to capital when investment opportunities improve (Almeida, Campello, and Weisbach, 2004).

What explains the traditional model of banking?

Why combine liquid deposits with illiquid loans and credit lines? There have been many attempts to understand this traditional structure of banks. Some explanations

suggest a causal chain running from liquid deposits to illiquid lending; some emphasize a chain of logic running from illiquid loans to liquid deposits; a third set emphasize a true synergy in which the causality goes in both directions.

Liquid Deposits → Illiquid Loans

One simple explanation for bank balance sheets stems from government deposit insurance. With insurance, deposits are safe, regardless of the bank's investment and financing decisions. Claims against banks are thus a close substitute for claims against the government and are an equally good store of value. Banks can invest in safe and liquid assets—they can be structured as narrow banks—or they can invest in higher-yielding risky assets like loans. That decision does not matter to a fully insured depositor. With limited liability, bank shareholders receive the upside but have limited downside and benefit from risky, high-yield assets. Thus, the 'moral hazard' from deposit insurance encourages banks to invest in risky assets like loans and may play some role in explaining bank structure. Explicit deposit insurance schemes have become common worldwide. As of 1999, for example, ninety-nine countries had created such programs (Kane and Demirgüç-Kunt, 2002).

Kane and Demirgiüç-Kunt also show that deposit insurance was absent in most countries before the 1960s. Implicit insurance can also induce risk shifting behavior, but history suggests that the structure of banking pre-dates active involvement by governments. Much theory attempts to explain the economics of combining illiquid loans with liquid deposits absent government guarantees. Fama (1985) argues that banks have a comparative advantage in information production flowing from their role managing the payments system. Imagine a bank providing payments services to a small business. Each day the small business makes and receives payments, both in the form of checks as well as currency. These payment flows reflect the current state of business, and if the small firm uses one bank, this bank has the opportunity to know before others if the firm is having problems. Such private information could explain why the bank can lend on better terms than other competing intermediaries.

Some evidence suggests that banks do have an information advantage stemming from deposits. Small firms concentrate their borrowing with a single financial service firm, and these firms usually borrow from commercial banks. More than 80 percent of borrowing from financial institutions comes from commercial banks, and most borrowing comes from lenders where firms have a deposit account (Petersen and Rajan, 1994). In two bank case studies, Mester, Nakamura, and Renault (2006) and Norden and Weber (2007) find that changes in checking account balances help banks monitor small businesses. Similarly, Udell (2004) finds that finance companies that lend with accounts receivable as collateral sometimes require borrowers to set up a special checking account to take payment on the receivables, thus potentially providing the finance company with the same information flows available to banks.

Loan pricing and credit availability, however, do not seem lower for those borrowing from the bank that holds their checking account. Cole (1998), for example, finds no link between the presence of a checking account and the probability that a firm will be granted credit from a bank. Cole, Goldberg, and White (2004) find that for small firms applying to borrow from small banks, the likelihood of approval increases with the presence of a deposit relationship, but no such result is evident when small firms borrow from large banks. Petersen and Rajan (1994) find no relationship between borrower deposits and the interest rate charged (holding constant the length of the bank–borrower relationship). Berger, et al. (2005) find no link between the presence of a checking account and the fraction of trade credit paid late, a measure of credit availability to the firm.

Beyond information, transaction deposits may be supplied inelastically with respect to market interest rates, thereby allowing banks to insure borrowers against credit shocks. Berlin and Mester (1999) present a model in which bank core deposits allow them to insure borrowers against credit shocks, whereas an intermediary funded with, say, commercial paper, would not be able to offer this insurance. Borrowers value this contract either because they are risk averse or because they face costly financial distress. Thus, the deposit franchise may give banks an advantage in lending that is unrelated to information. As evidence, Berlin and Mester show that interest rates on bank loans are less sensitive to economy-wide credit shocks (e.g. corporate bond spreads, changes in unemployment, etc.) when the originating bank holds more core deposits (defined as deposits under $100,000).

Myers and Rajan (1998) argue that because banks are funded with very liquid debt and have such high leverage, they need to some hold illiquid assets to mitigate the risk of expropriation or fraud. It is simply too easy for the banker to 'steal' when assets are highly liquid. In a sense, their model argues against the 'narrow bank' in which deposits are backed 100 percent by low-risk and highly liquid government securities. Too much liquidity on the asset side is dangerous because it becomes too easy for funds to be expropriated quickly. Although not an empirical article, Myers and Rajan argue that the historical development of commercial banking supports their model. Banks historically emerged as payments providers only; the bank began as a 'money changer'. These money changers held high levels of reserves, and the main risk perceived at the time had to do with fraud rather than bank runs. Myers and Rajan argue that the money changers enhanced their reputation for honesty by engaging in lending in the local community, hence the origin of asset transformation.

Illiquid Loans → Liquid Deposits

Several recent theories argue that the structure of bank lending shapes the nature of their liability structure. These arguments reverse the causal chain but lead to the same main implication, which is that illiquid loans go together with liquid

deposits. For example, Calomiris and Kahn (1991) and Diamond and Rajan (2001) argue that demandable deposits, by making the bank vulnerable to a destructive run, improve incentives for monitoring loans. Similarly, Flannery (1994) argues that very short term maturity of deposits improves bank incentives—for example, asset substitution problems are contained by short debt maturity. Moreover, Calomiris and Kahn emphasize that the 'sequential service constraint', whereby deposits are paid on a first-come, first-served basis, strengthens monitoring incentives for informed depositors. Thus, the nature of the bank loan portfolio shapes the structure of its deposits.

Illiquid Loans ← → Liquid Deposits

Diamond and Dybvig (1983) argue that the liquid deposit account offered through a financial intermediary fosters households insurance against liquidity risk and fosters consumption smoothing. In their model, a bank is a mechanism to allow investors to finance illiquid but high-return projects while insuring against unpredictable early-period consumption demands through pooling. The cost of this arrangement is the possibility of a bank run. While this model does not suggest a true synergy between lending and deposits, it does begin to consider links between the two sides of the banking business.

Recent studies suggest that by combining exposure to liquidity risk in both deposit taking and lending yields a risk-reducing synergy. Kashyap, Rajan, and Stein (2002) argue that as long as liquidity demands from depositors and borrowers off lines of credit are not too correlated, an intermediary reduces its cash buffer by serving both customers. Holding cash raises costs for both agency and tax reasons. Thus, their model yields a diversification synergy between transactions deposits and unused loan commitments. Gatev, Schuermann, and Strahan (2007) show that bank stock-return volatility increases with both transactions deposits and unused loan commitments, but that volatility declines when banks are exposed to liquidity on both sides of the balance sheet. Thus, there seems to be a hedge associated with combining these two activities.

Gatev and Strahan (2006) suggest a stronger hypothesis than Kashyap, et al., arguing that liquidity demands may be negatively correlated during episodes of 'flight to quality'. Funds tend to flow into bank transactions deposits during such episodes. At the same time, demands from borrowers for liquidity from credit lines also increase then because of lack of liquidity in commercial paper and bond markets. Why do banks enjoy funding inflows when liquidity dries up? First, the banking system has explicit guarantees of its liabilities. Second, banks have access to emergency liquidity from the central bank. Third, large banks such as Continental Illinois have been supported in the face of financial distress (O'Hara and Shaw, 1990). Thus, funding inflows occur because banks are rationally viewed as a safe haven for funds. Consistent with this notion, Pennacchi (2006) finds that

during the years before federal deposit insurance, bank funding supply did *not* increase when spreads tightened.

Sources of funding liquidity risk

By its very nature, providing funding liquidity makes (non-narrow) banks unstable because they are in the position of promising to disgorge cash on demand. Diamond and Dybvig (1983) present the classic treatment of liquidity risk from traditional asset transformation. Under normal circumstances the bank can meet random liquidity demands from depositors, but the structure is vulnerable to a run based on depositors' expectation that *other* depositors will run. In other words, runs can occur with no basis in fundamentals. Forced early liquidation lowers value and so runs are costly. This bank-run scenario presents the classic rationale for government provision of deposit insurance and backup liquidity from central banks because these safety nets eliminate the incentive to run based on depositor expectations that others may run. The safety net, however, creates a host of other incentive and implementation problems (see Chapters 26 and 27 in this volume).

The empirical evidence suggests that expectations-based runs are the exception rather than the rule. Gorton (1988) studies seven banking panics in the US prior to the creation of the Federal Reserve and finds that in all seven cases rational concerns about solvency motivated depositors. Under such circumstances, runs may cripple even healthy banks if depositors cannot distinguish solvent from insolvent banks. Gorton also finds, however, that the economic shocks preceding the banking panics during the Depression were not sufficiently large to justify the scale of the subsequent runs. Thus, changes in depositor expectations may have exacerbated the severity of bank runs above what one might normally expect. Calomiris and Mason (1997) study a specific run on Chicago banks in 1932 and also find no evidence that solvent banks failed because of runs.

More recently, McCandless, Gabrielli, and Rouillet (2003) studied runs faced by banks in Argentina in 2001 and, consistent with the studies of the US, find that solvency concerns were at the root of the runs. Puri and Iyer (2007) used detailed depositor-level data from a single Indian bank that experienced a run triggered by the insolvency of a large cooperative bank invested heavily in securities-based lending. Sharp declines in stock prices in 2001 led to the panic, but the individual bank studied appears to have been solvent. The study finds that depositors with a longer history with the distressed bank, depositors that also borrow from the bank, and depositors from the majority ethnic groups were less likely to run than other account holders. Relationships between banks and depositors seem to mitigate liquidity risk.

Loss of funding from runs can spill over and constrain bank production of credit. Bernanke (1983) argued, for example, that bank failures reduced credit supply and worsened the US depression during the 1930s. Many subsequent studies have tested how bank insolvency or monetary policy changes affect credit, but several recent studies have found that pure liquidity shocks can also alter credit. Khwaja and Mian (2005) study bank lending following runs on dollar-denominated deposits that occurred in the wake of Pakistan's unexpected nuclear test in 1998. They exploit variation in the magnitude of the liquidity shock across banks (from different levels of dollar deposits), and show that borrowers substituted away from banks experiencing greater runs and toward banks experiencing smaller runs (or no runs). Paravisini (2007) exploits the opposite kind of shock—an injection of liquidity by the Argentine government—and finds that profitable lending expanded following the liquidity infusion.

Banking panics and failures dropped to near zero in US banks from the creation of the FDIC in 1934 until the late 1970s. This stability ended in the 1980s, but most of the failures have been from investments in high-risk loans (e.g., Continental Illinois in business lending; Bank of New England in high-risk commercial real estate), or in the case of savings institutions in taking on interest rate risk. In some cases banks faced funding outflows, but for the most part instability had little to do with liquidity risk from depositors. For example, large, uninsured depositors began removing funds from Continental Illinois in response to large credit losses on business lending. In 2007, depositors ran from the UK bank Northern Rock and the US bank Countrywide, as did depositors and other short-term creditors at Bear Stearns and other financial institutions in 2008. These runs were again based fundamentally on rational concerns about solvency stemming from losses on securities backed by troubled subprime mortgages (see below).

Bank exposure to funding liquidity risk also arises from issuance of lines of credit, which commit the bank to provide cash on demand. Much of the day-to-day variation in liquidity risk is managed by diversification across a large base of customers. Kashyap, Rajan, and Stein (2002) argue that diversification benefits across broad classes of customers—specifically borrowers off lines of credit and demand depositors—can reduce risk and the need to hold cash. As evidence, they report that banks dominate the market for lines of credit and, among banks, those with high-transactions deposits issue more credit lines. Gatev, Schuermann, and Strahan (2007) show that stock-return volatility is lower at banks exposed to liquidity risk in both lines of credit and deposits, suggesting a powerful hedge associated with combining these two products.

Several studies explore why firms and households draw funds from credit lines. Aggarwal, Ambrose, and Lin (2005) find that individuals draw more funds from home equity lines when their credit quality declines. Using data on Spanish firms, Jimenéz, Lopez, and Saurina (2007) find that usage rates on credit lines increase as firms near bankruptcy. These studies suggest that *ex post* demand for liquidity

increases as a borrower's credit quality deteriorates, but other studies suggest that bank *ex ante* supply of lines offsets this risk. Banks restrict access to lines for poor-credit-quality firms and households. For example, Gropp, Schulz, and White (1997) show that total supply of credit to households is lower in states with large housing exemption, and that this reduction restricts access to low-income, high-default risk customers. Sufi (2007) finds that large public firms with low cash flow have less access to liquidity from bank lines and thus hold more balance sheet cash. Similarly, Agarwal (2004) finds that riskier private firms have both less access to credit lines and lower utilization rates than safer firms, and Agarwal, et al. (2006) compare home equity lines with second mortgages (home equity loans) and find greater default risk in the second category of 'spot loans'.

Managing funding liquidity risk

Banks use several tools to manage funding liquidity exposures. First, they hold liquid assets—cash, securities, and loans that may be sold easily. Second, they have access to liquidity from other banks in the interbank market. Third, they have liquidity support from the Central Bank. And, fourth, they are funded with deposits that expand during periods of market uncertainty.[2]

Holding cash is a simple yet costly way to bear liquidity risk. Banks have no particular advantage in bearing risk this way. Cash is costly because it earns a low return, is tax inefficient, and may be easily diverted or misallocated (Jensen, 1986). Nevertheless, during the early part of the twentieth century banks used cash to persuade depositors of their soundness. A. P. Giannini, the founder of what is now the Bank of America, famously withstood runs during the panic of 1906 by displaying gold reserves on the street front, and offering to convert deposits into gold to all comers. This show of strength calmed his depositors while many competing banks failed. Modern evidence also suggests that cash and liquid assets, as well as loans that can potentially be sold or securitized, act as a buffer for banks against funding shocks (e.g. Kashyap and Stein, 2000; and Loutskina, 2005).

Banks' second layer of support comes from borrowing and lending in the interbank market. While not a source of aggregate liquidity, interbank markets can recycle liquidity through the system. Typically, large banks have greater access to interbank credit than small ones. Ashcraft, McAndrews, and Skeie (2007) find that small US banks hold larger cash buffers and excess reserves with the Federal Reserve than large banks, and that they supply funds to large banks in the Federal Funds market.

[2] Banks also face similar liquidity risk management problem in balancing high frequency payments during daylight hours. The Federal Reserve provides intraday credit to help grease the wheels of the payments' system. The funding liquidity risk discussed here involves somewhat longer-lived risks of loss of funding sources beyond a single day.

Even with such access, a key issue for large banks, and one that came into sharp focus during the 2007–8 crisis, is how well this market functions during periods of market stress. Allen and Gale (2000) show theoretically how interbank connections can lead to contagion of liquidity shortages from one part of the banking system to others. Furfine (2002) studies the Fed Funds market during the Long-Term Capital Management (LTCM) crisis and finds that LTCM-exposed banks continued to have access to borrowing during this period. During the summer of 2007, however, spreads in the LIBOR market increased to about 50 basis points above rates in the Fed Funds market as investors appeared to lose confidence in their ability to evaluate the risks of large European banks with potential exposure to losses in the US subprime mortgages market (Kane, 2007). As we will see below, the interbank market dried up spectacularly following the failures of AIG and Lehman Brothers in the Fall of 2008.

The third source of liquidity for banks is from the Central Bank, the 'lender of last resort'. According to Bagehot (1873), central banks should lend to illiquid but solvent banks at a penalty rate. Some have argued that such targeted liquidity support may worsen moral hazard problems associated with bailouts. Instead, open-market operations that expand the total supply of liquidity are preferable. Such broad expansions of liquidity can be recycled in the interbank lending market without (or with less) moral hazard (e.g. Goodfriend and King, 1988). Coordination failures may occur in the interbank market, however, whereby banks hoard liquidity because of concern about counterparty solvency (e.g. Rochet and Vives, 2004). Such coordination failures worry policymakers, who often intervene to overcome them. Historical examples include the 1987 stock market crash, when commercial banks were encouraged to lend by the Federal Reserve to distressed investment banks; the reorganization of the LTCM in 1998, where counterparties were discouraged from forcibly liquidating the hedge fund; the temporary freezing up in the wholesale payments' system following 9/11, when Federal Reserve officials not only injected liquidity through open-market operations and direct lending to banks, but also implored banks to resume making payments to restore the normal patterns of payment coordination (McAndrews and Potter, 2002); and, of course, the summer of the 2007–8 crisis, which we discuss below, when the Federal Reserve explicitly encouraged banks to borrow from the discount window to assure markets that liquidity support would be offered if necessary.

Combining exposure to funding liquidity on both the asset and liability sides of the balance sheet provides the fourth measure of liquidity stability for banks. First, as noted above, Kashyap, Rajan, and Stein (2002) argue that demands for funding liquidity by borrowers and depositors tend to be less than perfectly correlated, so combining the two products offers some diversification benefits. Moreover, Saidenberg and Strahan (2000) study the LTCM crisis during 1998 and find that bank lending increased to satisfy a *systematic* increase in loan demand from firms that normally receive liquidity in the commercial paper market, but that this increase in

liquidity demand from borrowers was offset by funding inflows by depositors. Gatev and Strahan (2006) study these flows across many market conditions and find that both bank loans and their holding of cash and securities increase when market liquidity dries up generally (as proxied by the commercial paper T-Bill (Treasury Bill) spread).

So, recent evidence suggests that by offering liquidity from lines of credit, banks expose themselves to the systematic risk that they may face loan takedowns across many borrowers at the same time. Bearing this risk requires access to funds at exactly the time that most firms find borrowing expensive. Banks enjoy an increase in funding supply at exactly such times because they are viewed as a safe haven for funds. For example, during the 1998 liquidity crisis, banks experienced funding inflows into transactions deposit accounts, and banks with larger transactions deposit bases prior to the shock received the greatest inflows. Since banks tend to combine these two products, flows into the bank deposit accounts tended to balance outflows of funds from unused lines of credit (Gatev, Schuermann, and Strahan, 2006).

Empirical trends in funding liquidity

The importance of funding liquidity produced through deposits seems to be falling consistently over time, as shown graphically in Figure 5.1. The decline in the ratio of

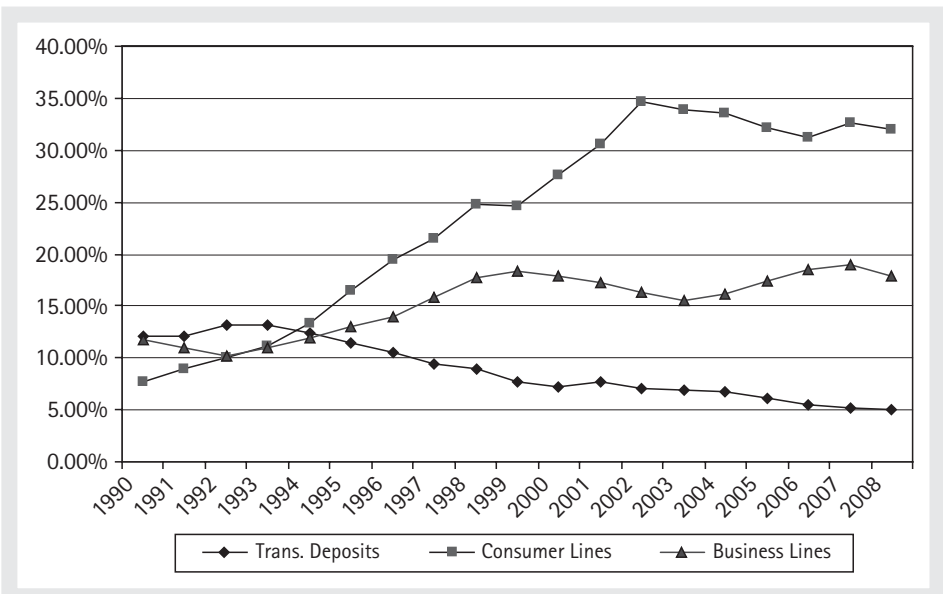

Fig. 5.1. Bank liquidity production to GDP

transactions deposits to GDP has occurred as banks have increasingly offered customers money-like services via electronic technologies. Similar trends are evident in plotting the ratio of M1—the amount of notes and coins in circulation plus demand deposits—to GDP. The secular decline in 'money demand' reflects better payments technologies that allow households to make payments without losing interest on their savings. For example, credit card payments' share rose consistently since 1980. Debit card usage remained very small (around 2 percent or less) until the middle of the 1990s, and then rose sharply to about 20 percent of all payments by 2006.

Figure 5.1 also shows that the drop in transaction deposits has been more than offset by the growth in unused credit lines. All kinds of lines—lines to consumers (mainly home equity lines and credit cards) and business lines—have grown relative to GDP over the past twenty years, in contrast to the declining ratio of transaction deposits to GDP. Lines to households have grown most dramatically, from about 10 percent of GDP in the early 1990s to more the 30 percent by 2006, more than offsetting the decline in transactions deposits to GDP. Moreover, the total amount of liquidity provision by the banking system as a whole has grown. Thus, while the composition of banks provision of funding liquidity has changed, it remains as important as ever (Berger and Bouwman, 2007).

Market liquidity

Many of the changes in banking over the past twenty years reflect moving from a model of 'originate and hold' to one of 'originate-to-distribute, and sell'. The first model involved creation of funding liquidity through asset transformation from loans to deposits. This traditional model has been reshaped by the growth of loan sales and securitization. In the modern approach, the bank creates *market* liquidity rather than *funding* liquidity; that is, the bank (or other intermediary) transforms a hard-to-sell asset like a loan into one that is easier to sell, like a bond or other security. This allows the originating bank to sell the asset to passive investors and recycle their capital to originate new loans, which can in turn be transformed and sold.

Securitization

Banks have increasingly used securitization to finance their lending by creating structures such as collateralized loan, mortgage, and debt obligations (CDOs,

CLOs, CMOs, and, generically, SIVs). These financing arrangements allow the originating bank to remove business loans, credit card loans, and mortgages from the balance sheet. Securitization involves pooling the cash flows from a number of similar assets and selling the pool to a separate legal entity known as a special purpose vehicle (SPV). The SPV purchases those cash flows from the proceeds of the sale of securities, such as bonds or commercial paper. The securities are sold to arm's-length investors like insurance companies and money market mutual funds. Rather than holding the asset on a balance sheet financed with liquid deposits (the traditional model of asset transformation), securitization transforms the asset itself from an illiquid one (pools of loans) into a liquid securities issued by the SPV (bonds and commercial paper).

The pooling process results in a diversified portfolio of cash flows, which are used to support payments on debt securities issued by the SPV. Creating the separate SPV isolates the cash-flow-generating assets and/or collateral so that securities issued by the SPV are not a general claim against the issuer, just against those assets. Cash-flows from the original pool of loans can be further stripped and repackaged based on various characteristics (e.g., the prepayment behavior or payment priority) to enhance their liquidity. Often, the cash flows come with some additional implicit or explicit guarantees from the originating financial institution. For example, the originator may retain the residual or equity tranche in the SPV, thus retaining most of the credit risk. Originating banks also will often issue backup liquidity when SPV funding comes from short-term sources such as commercial paper. The backup liquidity may be necessary if the SPV has trouble rolling over the commercial paper, as occurred during the Summer of 2007 (see below). In cases like this, there is no clean separation between the bank's production of funding and market liquidity because without the backup liquidity (i.e., the funding liquidity), the securitization would probably not be possible.

Securitization is attractive to banks (as well as to non-financial firms) because it lowers the total cost of financing loans. One benefit of securitization is that it avoids bankruptcy costs. In contrast to normal debt finance, owners of the SPV-issued debt have *no claim* against the originator's other assets if the originator files for bankruptcy (Ayotte and Gaon, 2006). Moreover, the SPV itself cannot go bankrupt, although defaults on the underlying loans can create losses for bondholders. This is accomplished contractually by forcing early amortization of the bonds issued by the SPV if cash flows from the underlying assets are lower than expected. With no possibility of default on the bonds, no claim against the originator when cash flows are low, and no decisions to be made by the SPV itself, the bonds sold by the SPV have sidestepped both the agency costs of financial distress as well as direct and indirect costs of bankruptcy (Gorton and Souleles, 2006).

While early amortization avoids financial distress cost, it does impose losses on bondholders. Thus, buyers of bonds created through securitization face a potential 'lemons problem' (Akerlof, 1970) because originators have better information and

may be tempted to securitize their low-quality loans. If the lemons problem were not solved in some manner, securitization would fail to lower the costs of finance. Demarzo (2005) shows how pooling and tranching can reduce the lemons problem by allowing the SPV to fund most of the purchase of the original assets with very safe bonds issued by the SPV. In a typical structure, the SPV will issue senior notes with a high rating (say AA), a mezzanine tranche with a lower rating (say BB), and an equity tranche that is unrated. Both the senior and mezzanine tranches are liquid and held by various sorts of institutional investors; only a small piece of the financing—the equity tranche—remains illiquid. The equity tranche bears all of the losses (unless losses fully deplete this tranche), and is typically held by the originator (Franke and Krahnen, 2004) or the master or special servicer (Sanders, 2004). Thus, most of the credit risk is concentrated in the equity tranche; because the originator holds this tranche, their incentive to place lemons in the pool is reduced. Moreover, the originator typically continues to collect payments and pass these payments to the SPV. Again, holding the first loss also improve incentives to monitor the assets to minimize losses on this riskiest tranche.

To reduce the risk of early amortization further, there have been a number of documented instances in which originating financial institutions voluntarily enhanced the cash flows to the SPV (and thus reduced losses to security holders) to preserve their reputation in the market (e.g., Calomiris and Mason, 2002 and Higgins and Mason, 2004). In 2007, several large banks repurchased billions in assets that had been securitized in SIVs. Gorton and Souleles (2006) show that the pricing of bonds issued by the SPV reflect not only the quality of assets in the pool but also the rating of the issuer, suggesting that investors value implicit support for unexpectedly low cash flows.

Securitization also enhances liquidity by creating classes of assets with risk characteristics suitable to different clienteles. For example, insurance companies may be the natural clientele for the most senior tranches, while hedge funds with either a strong appetite for risk or superior credit risk management models may be the natural clientele for the subordinated tranches. Mortgage securitizations are often tranched according to prepayment risk. Again, this kind of structure allows specialists in prepayment risk to earn returns on their expertise by concentrating that risk in one class of securities.

Securitization of mortgages has grown most dramatically in the US, in large part because of subsidies from government-sponsored enterprises (GSEs)—The Federal National Mortgage Association (Fannie Mae) and the Federal Home Loan Mortgage Corporation (Freddie Mac). Fannie Mae was created by the US Congress in 1934 to promote access to mortgage credit for low- and moderate-income households. During its first three decades, Fannie Mae was operated as a government agency that purchased mainly mortgages insured by the Federal Housing Authority (FHA). In 1968, Fannie Mae became a public corporation; its role in purchasing FHA mortgages (as well as mortgages insured by the Veteran's Administration) was

taken over by a new government agency, the Government National Mortgage Association (GNMA). Freddie Mac was chartered by Congress in 1970 to provide stability and liquidity to the market for residential mortgages, focusing mainly on mortgages originated by savings institutions. Freddie Mac was privatized in 1986.

By the 1990s, both Fannie Mae and Freddie Mac were heavy buyers of mortgages from all types of lenders, with the aim of holding some of those loans and securitizing the rest. Together they have played the dominant role in fostering the development of the secondary market. As shown by Frame and White (2005), the GSEs combined market share has grown rapidly since the early 1980s. In 1990 about 25 percent of the $2.9 trillion in outstanding mortgages were either purchased and held or purchased and securitized by the two major GSEs. By 2003, this market share had increased to 47 percent.[3] GSE access to implicit government support allows them to borrow at rates below those available to private banks. Passmore, Sherlund, and Burgess (2005) argue that most (but not all) of the benefits of GSE-subsidized borrowing benefits their shareholders rather than mortgage borrowers. To take advantage of this subsidy, during the 1990s, the GSEs increasingly opted to hold, rather than securitize, many of the mortgages that they buy. Policymakers have become concerned about the resulting expansion of interest rate risk at the GSEs (Greenspan, 2004). During the expansion of credit to subprime borrowers, the GSEs also encouraged securitization by buying mortgage-backed securities in the secondary market.

Despite the policy concerns, the GSEs do enhance mortgage liquidity either by buying and holding mortgages or by securitizing them. The GSEs operate under a special charter, however, limiting the size of mortgages that they may purchase or securitize. Today, the GSEs may only purchase 'non-jumbo' mortgages, defined in 2006 as those below $417,000 for loans secured by single-family homes. The loan limit, first set at $93,750 in 1980, increases each year by the percentage change in the national average of one-family housing prices, based on a survey of major lenders by the Federal Housing Finance Board. Loutskina and Strahan (2007) show that bank supply of mortgages to the jumbo market are constrained by their liquidity and cost of funds, whereas there are no such supply constraints to the non-jumbo market because banks have the low-cost option of selling those mortgages to the GSEs.

Loan sales and syndication

Banks also create market liquidity in loan syndication and in secondary market trading of loans. Gorton and Pennacchi (1995) describe how adverse selection and

[3] GNMA provides a very important source of mortgage finance to low-income borrowers, holding or securitizing about 10% of all mortgages outstanding.

moral hazard dampen the liquidity of loans. They argue that loan sales took off in 1980s because banks learned to sell only a portion of loans and began to offer implicit guarantees to buyers (e.g. promises to buy back troubled loans).[4] As a result, loan sales grew from $27 billion in 1983 to $291 billion by 1989. Loan syndication, which grew dramatically during the 1990s, works much the same as loan sales. In these structures, financial institutions lend as a group at the outset, with one bank taking the lead in contracting with the borrower.

Loan sales and syndication differ from securitization mainly in the scale of loans considered. Loan sales involve large loans while securitization achieves sufficient scale by pooling of many small loans. But the fundamental contracting problems—asymmetric information and moral hazard—are basically the same. In loan syndication, a lead bank has the primary responsibility for negotiation with the borrower, writing contracts, pricing the loan, and setting the non-price terms (e.g., covenants). The lead bank also manages the relationship over time. The lead bank will often guarantee a commitment amount to the borrower, and then sell pieces of the loan to participant banks. The participant banks thus help fund the loan but are less involved in the relationship on a day-to-day basis. Lead banks thus typically have better information than participants, so there is a potential lemons problem similar to the one described earlier about securitization.

As with securitization, the contracting problem in loan sales and syndication is solved in part through incentives and in part though the lead-bank's reputation. Lead banks generally retain the largest share of syndicated loans, which helps mitigate the information problem faced by less-informed participants. Both the lead-bank's share and the concentration of the syndicate increase with borrower opacity (Dennis and Mullineaux, 2000; Lee and Mullineaux, 2004; Jones, Lang, and Nigro, 2005; and Sufi, 2007). Moreover, Ivashina (2006) shows that the incentive problems built into a syndicate are priced into the yield.

Syndication itself is a kind of liquidity production on the part of the lead bank, similar conceptually to bond underwriting. Since 1995, many cases of loan syndications have also received ratings from Moody's and Standard & Poor's, just as bonds do. In contrast to bonds underwritten by investment banks, however, lead banks are more actively involved in maintaining a relationship with the borrower after syndication. And, as we have seen, the lead bank invariably retains a stake in the loan, again in contrast to the case of bond underwriting.

In recent years, participation in syndicated loans has become increasingly liquid as secondary market trading has flourished. Güner (2006) finds that yields on loans issued by banks that actively sell loans tend to be lower than other loans, consistent with the idea that liquidity reduces the yield required to compensate lenders. Wittenberg-Moerman (2006) finds that bid-ask spreads are higher when borrowers

[4] Such guarantees undermined the spirit of bank capital requirements and have been a concern to bank supervisors.

are more opaque (e.g., unrated, private firms vs. rated, public firms), consistent with private information reducing liquidity. She also finds, however, that loans originated by lead arrangers with greater market share trade at lower spreads. This result echoes studies of securitization, providing further evidence that banks enhance liquidity not only with explicit contracting and credit guarantees but also with their reputation (e.g., through implicit guarantees or recourse).

Much of the secondary market trading occurs as a means for non-bank investors to enter the syndicated lending market. Term loans trade much more than credit lines because non-bank institutional investors do not want the funding risk. Drucker and Puri (2005) study the emergence of secondary-market trading volume in syndicated loans, which has grown from nearly nothing in 1990 to about $180 billion in 2006. They find that loans with more restrictive covenants are more likely to trade because covenants provide assurances to buyers that they will have sufficiently strong control rights to protect their investment. In this case, control rights seem to act as an alternative to information in enhancing liquidity. Their results are surprising because most public debt, which continues to trade more than bank loans, comes with much looser covenants than what is seen in the typical syndicated loan. Overall, however, this paper suggests that information asymmetry between bank lenders and potential buyers continues to dampen liquidity. For example, in their sample, sold loans are rated 88 percent of the time, compared to just 39 percent for loans that have no secondary market liquidity.

So, banks provide market liquidity when they repackage loans via securitization, when they trade loans, and when they act as lead arrangers in loan syndication. In all three cases the asymmetric information creates a barrier to liquidity. To overcome this barrier, banks expose themselves to risk—in the case of loan securitization by taking the first losses and in the syndicated lending context by holding the largest share of the loan. In many cases originating banks also provide implicit guarantees. Reputation in the market mitigates shirking incentives, and helps explain why large, well-capitalized banks tend to dominate as lead arrangers.

Sources of market liquidity risk

Market liquidity risk occurs if banks lose the ability to sell or securitize loans at fair prices. Under such circumstances, market liquidity risk feeds back to funding liquidity if the bank must raise funds on short notice in order to hold those assets on their balance sheet, or if a bank must repurchase assets sold in a securitization due to funding disruptions in the capital markets (see below).

Producing market liquidity requires banks to bear enough risk to maintain incentives to deal responsibly with the borrower in setting prices and enforcing

covenants. Loan syndication also exposes lead banks to 'underwriter risk', similar to the risk born by securities' underwriters offering firm commitments to debt and equity issuers. In the case of syndicated lending, if demand from participant banks is lower than expected, the lead arranger must either fund more of the loan than anticipated or the deal may fail to close.[5] Both outcomes are costly for the lead bank. The first option would require the bank to have additional debt and equity capital; the second would be harmful to the lead bank's reputation.

Securitization also often comes with not only credit guarantees of various kinds but also liquidity support. For example, asset-backed commercial paper structures usually come with a liquidity backstop facility issued by the bank that set up the securitization. These facilities expose banks to funding liquidity risk as described above, but the facilities are necessary to create the securitization in the first place. Many of these kinds of structures could not refinance their commercial paper market during the 'credit crunch' of 2007 and required banks to replace that financing to avoid default. In such cases, there is no bright line between market liquidity and fuinding liquidity.

Empirical trends in market liquidity

Figure 5.2 illustrates the growing quantitative importance of loan securitization for different types of loans over time. In 1976, there was no securitization of commercial mortgages, business loans (commercial and industrial (C and I) loans), or consumer loans. By the end of 2007, 28 percent or $294 billion of commercial mortgages were securitized and $104 billion worth of C and I loans were securitized, along with 27 percent or $658 billion worth of consumer loans had been securitized. But the really explosive growth has occurred in the market for home mortgages: in 1976, the amount of securitized home mortgages was $28 billion; by the end of 2007, the total amount of securitized home mortgages had grown to almost 60 percent of the market (reaching $4.2 trillion). Over the same period, the amount of home mortgages outstanding grew from $489 billion to $7.3 trillion.

Unlike the US, securitization has not been spurred elsewhere by government enterprises—there are no institutions analogous to the GSEs in Europe. Nevertheless, securitization has taken root there as private banks have begun to tap into markets to fund various kinds of loans. Table 5.1 reports the rate of securitization issuance for new loans between 2000 and 2006, and by collateral type in 2005. The

[5] While most loan syndicates are arranged under a firm commitment between the bank and the borrower (opposed to 'best efforts' where borrowers bear all of the risk of the issue failing), some syndicated loans have been arranged under market-flex contracts whereby the pricing of a loan may not be guaranteed by the lead bank (Standard & Poor's, 2006). These kinds of arrangements shift some of the underwriting risk from the lead arranger(s) to the issuer.

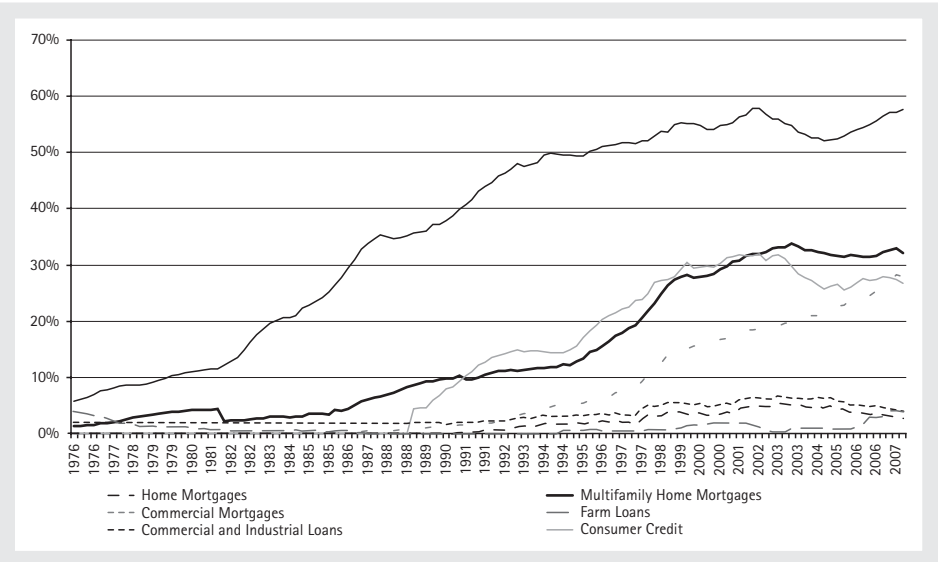

Fig. 5.2. Share of securitized loans in the US economy

figures show rapid growth overall. For example, between 2000 and 2006 securitization origination rates soared by more than 35 percent per year in Europe; in contrast, based on growth in outstandings, securitization in the US rose by only about 9 percent per year over the same period. Securitization of residential mortgage-backed securities grew fastest in Europe, by almost 70 percent between

Table 5.1. Securitization and residential mortgage markets in Europe

Year	Total Securitization (€ million)	Annual Growth Rate	2005 Securitization by collateral		
			Collateral	Total (€ million)	Share
2000	78.2	–	Auto Loans	4.1	1.3%
2001	152.6	95.1%	Credit Card	11.7	3.6%
2002	157.7	3.3%	CDOs	48.9	15.0%
2003	217.3	37.8%	Commercial Mortgages	38.6	11.8%
2004	243.5	12.1%	Loans & Leases	55.1	16.9%
2005	327.0	34.3%	Residential Mortgages	144.9	44.3%
2006	458.9	40.3%	Other	23.7	7.2%
			Total	327.0	100.0%

Sources: Thomson Financial, Dealogic, JP Morgan, Merrill Lynch, Structured Finance International, Bloomberg.

2005 and 2006. Moreover, as in the US, securitization of loans backed by real estate—both residential and commercial—seem to be taking the lead, comprising about 56 percent of total securitization in 2005 (compared with about 80 percent share for real estate loans in the US). Presumably real-estate-backed loans are relatively transparent and thus amenable to purchase by a diffuse class of investors, in contrast to more opaque assets such as loans to businesses.

The financial crisis of 2007–8

The financial crisis of 2007–8 is the biggest shock to the banking and financial system since the 1930s and offers the greatest challenge to our understanding of liquidity production and liquidity risk management, both for private institutions and for regulators and central banks. In broad terms, the financial system experienced runs from various sources, including depositors, customers, counterparties, and short-term creditors. This liquidity crisis mirrors what has happened in the past in that the runs were based on concerns over solvency rather than representing runs based solely on expectations that others might run. The crisis, ongoing at the time of writing, has led to a severe contraction in bank loan originations. Figure 5.3 illustrates the time series of new lending to large businesses from Loan Pricing

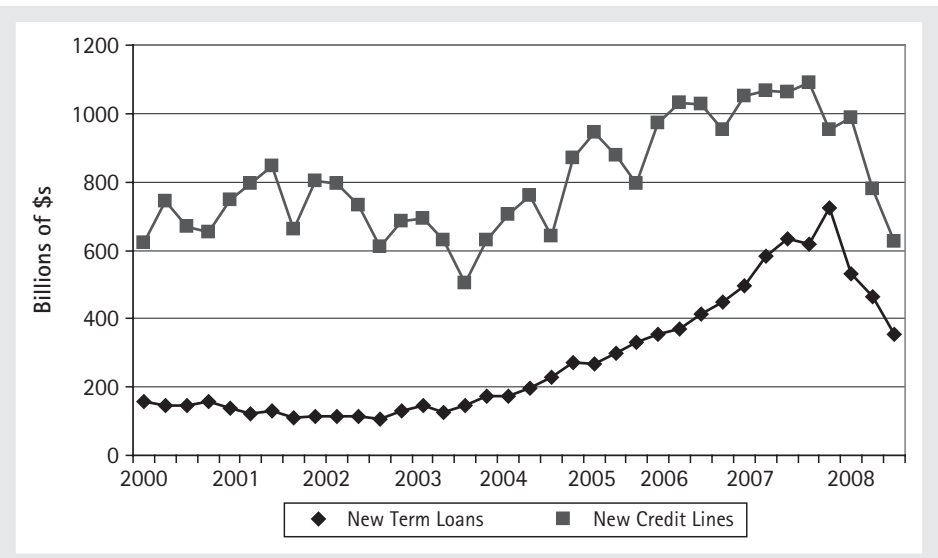

Fig. 5.3. Business loan originations collapse

Corporation's *Dealscan* database from 2000 to the third quarter of 2008. During 2001 and 2002, both lines of credit and term loans declined, as would be expected during a mild recession; but this earlier decline pales relative to the steep drop in new lending beginning in the middle of 2007. One cannot predict how deeply the crisis will reduce consumption and investment, or how profoundly the downturn will change the management of risk and the regulation of financial institutions going forward. Nevertheless, after describing very briefly what happened, I will discuss the early lessons for liquidity production and risk management, focusing mainly on the US financial system.

What caused the crisis?

The roots of the crisis lie in the overvaluation in housing prices and the subsequent crash in those prices beginning around 2007. The popping of this real estate bubble created large losses to lenders. Kindelberger (2000) describes past episodes of asset-pricing bubbles, going back several hundred years. He finds that such bubbles tend to be preceded by loose monetary policy and an over-expansion of credit. The current episode supports his understanding of the historical record. Sufi and Mian (2008) show that markets where credit expanded most experienced both the greatest house price appreciation and the worst subsequent crashes. Demyanyk and Van Hemert (2008) provide evidence that underwriting standards eased with each lending cohort from 2000 to 2006, coinciding with the run-up in prices.

Most analysts have blamed the move from the traditional 'buy and hold' to the new 'originate-to-distribute' model of lending for the credit expansion, and a few recent studies have offered rigorous evidence consistent with this notion. For example, Keys, et al. (2008) show that mortgages expected to be securitized had greater *ex post* default rates than otherwise similar mortgages retained by lenders; Purnanadam (2008) shows that banks with large pipelines of mortgages that were intended to be sold faced losses when liquidity dried up in the mortgage-backed securities market in 2007. Loutskina and Strahan (2008) argue that because banks moved en masse toward a diversified lending model—a model facilitated by securitization—investments in private information about local credit markets declined, thus setting the stage for over-expansion of credit.

An abbreviated chronology of key events

While concern about subprime mortgages began somewhat earlier, the crisis really took hold in the summer of 2007. In June and July, two Bear Stearns hedge funds required assistance, and Countrywide, one of the largest subprime originators,

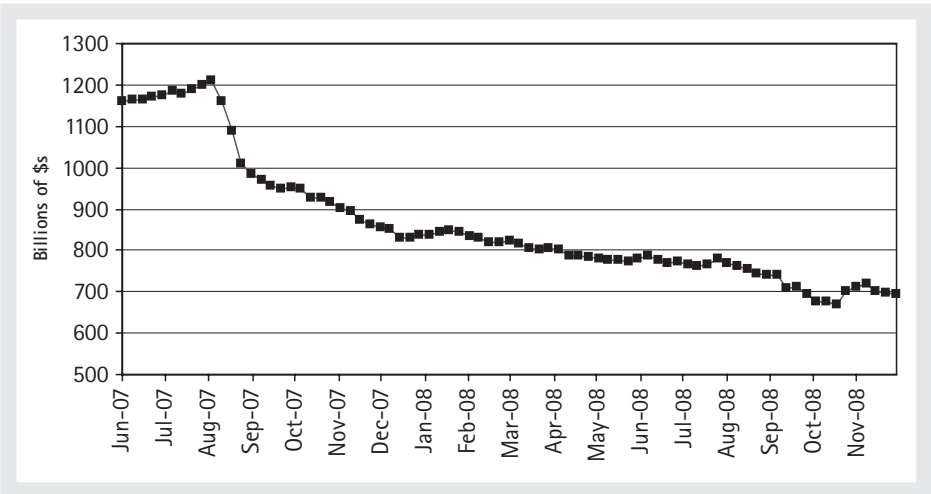

Fig. 5.4. Asset-backed commercial paper dries up

announced unexpectedly large losses. In August 2007, the asset-backed securities market dried up when several issuers failed to provide liquidity to support funding of securitized assets financed with short-term commercial paper (Brunnermeier, 2008). Banks had been moving pools of loans off balance sheet and into so-called SIVs financed with short-term commercial paper. This market peaked in 2007 with about $1.2 trillion outstanding, and then declined by about 50 percent in just six months (Figure 5.4). The funding liquidity risk of these structures, which replaced the old on-balance sheet model of asset transformation, *did not* leave the banking system because issuers provided liquidity backstops to insure against refinancing risk in the asset-backed commercial paper. The market's faith in these backstop facilities wavered when an SIV issued by IKB, a small German bank, was unable to refinance its short-term commercial paper, and IKB could not meet its obligation to refinance the SIV through its line of credit. The asset-backed securitization market collapsed, leading to balance sheet stress for large issuers such as HSBC and Citigroup, who had to take large pools of these assets back onto their balance sheets.

In response to the decline in asset values and an increase in concerns about bank solvency, the interbank market began to lose liquidity. The cost of borrowing at maturities beyond overnight rose especially sharply. In August of 2007, the spread between the three-month bank CDs and the US Treasuries tripled, from about 50 to 150 basis points (Figure 5.5).[6] Spreads fell back to a lower yet still elevated level in the Fall, and then spiked to more than 200 basis points in December 2007. In reaction to this illiquidity, the Federal Reserve created the TAF to sell a fixed

[6] Similar patterns are evident using other bank borrowing rates such as the LIBOR at various maturities.

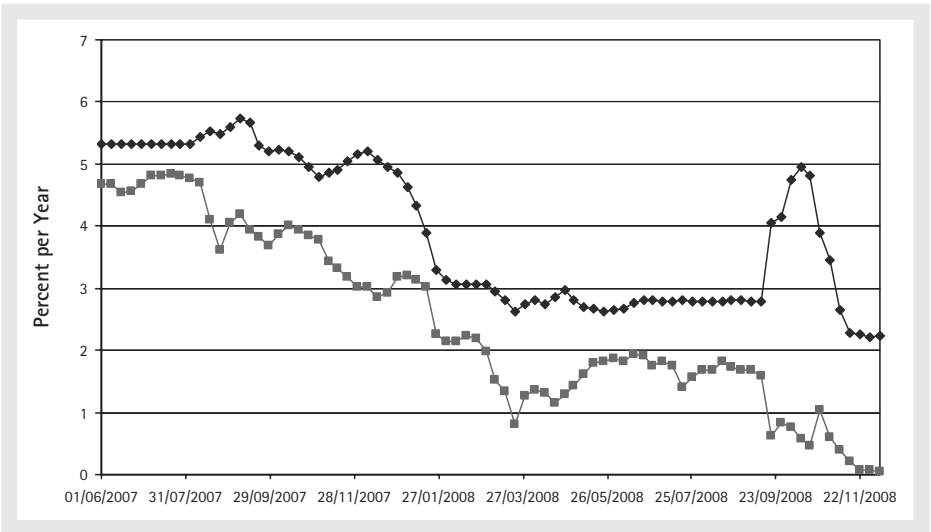

Fig. 5.5. Yield on three-month CDs vs. treasuries

quantity of three-month credit (and later longer-term) in a competitive auction. These auctions reduced borrowing costs temporarily, with spreads falling below 100 basis points by February 2008 (McAndrews, Sarkar, and Wang, 2008).

Then, in March 2008, concern about the value of Bear Stearns' large portfolio of subprime mortgage-backed securities led to a run by many of their counterparties, short-term creditors, and large customers (e.g., hedge funds), again stressing the interbank market. Spreads jumped above 150 basis points (Figure 5.5). The Federal Reserve stepped in, brokering a rescue of Bear Stearns by J.P. Morgan, and guaranteeing most of the losses on Bear's troubled portfolio of subprime assets. The Federal Reserve then launched the Term Securities Lending Facility and the Primary Dealer Credit Facility, essentially opening up its discount window to the remaining large Wall Street investment banks. With these three creative new lending facilities, the Federal Reserve began playing its role as lender of last resort on a massive scale, stepping in to supply the liquidity that had ceased to flow in the interbank credit markets.

Conditions improved following the bailout of Bear Stearns. The cost of funds to banks fell, as did spreads over Treasuries (Figure 5.5). In the summer of 2008, however, mortgage foreclosures continued to rise, leading to further downgrades of mortgage-backed securities by the credit rating agencies and accelerating losses to holders of those securities. In July, Congress passed stopgap legislation, formalizing its previously implicit guarantee of debt issued by Fannie Mae and Freddie Mac. Despite the debt guarantee, the razor-thin capital ratios of these two GSEs were overwhelmed by credit losses, forcing the Treasury to take both into conservatorship

by early September. Similar losses accrued to others with exposure to real estate, leading to the failures of AIG and Lehman Brothers.

The depth of the crisis dramatically expanded when markets were shocked by the AIG/Lehman collapses in mid-September. Lehman was allowed to go bankrupt and, as a result, a large and reputedly conservative money market fund—the Primary Reserve Fund—'broke the buck', meaning that its investors lost principal. This fund had built a reputation for safe investment; hence its exposure to Lehman scared investors, leading to a run on mutual funds broadly.[7] Within a few days more than $200 billion had flowed out of these funds (Krishamurthy, 2008). The US Treasury stopped the run by extending government insurance over money market mutual fund accounts on a temporary basis. Nevertheless, the panic soon spread globally, leading to expansion of insurance on deposits and interbank funds, first in Europe and then very quickly in the US. Public capital was also injected into all of the large banks in an attempt to allay fears about solvency.

Demand and supply of funding liquidity

The demise of AIG and Lehman massively increased the demand for funding liquidity across the whole financial system. US banks' cash holding rose dramatically (Figure 5.6). Until the Fall of 2008, banks held a very stable level of about $300 billion in cash going back at least five years. Starting in September, however, banks begin to hoard cash, with the level rising sharply, peaking at about $850 billion in the middle of November 2008. Why the increase? As noted earlier, concerns about the solvency of financial institutions had damped the supply of credit in the interbank market. The absence of interbank credit thus increased banks' precautionary motive to hold cash as a liquidity buffer.

Non-financial firms also lost access to short-term funds as the commercial paper market dried up (Figure 5.7). Commercial paper had typically been held by money market mutual funds, but their appetite for such assets collapsed in the wake of the Lehman failure; instead, these funds began to fill up their balance sheets with Treasuries. Demand for liquidity from banks by non-financials also increased as issuers drew funds from prearranged backup lines to refinance their commercial paper as it came due, thereby feeding back into banks' demand for cash. This spike in liquidity demands on the banking system can be seen clearly in Figure 5.7, where the drop in outstanding commercial paper coincides exactly with an increase in business loans on bank balance sheets. Thus, loans increased in response to draw

[7] The *Wall Street Journal* reported that the head of the Primary Reserve Fund had criticized competing money market fund managers for holding assets with credit risk. 'Commercial Paper is Anathema to the Concept of the Money Fund', Mr Bent told Reuters in 2001 (*Wall Street Journal*, 8 December 2008).

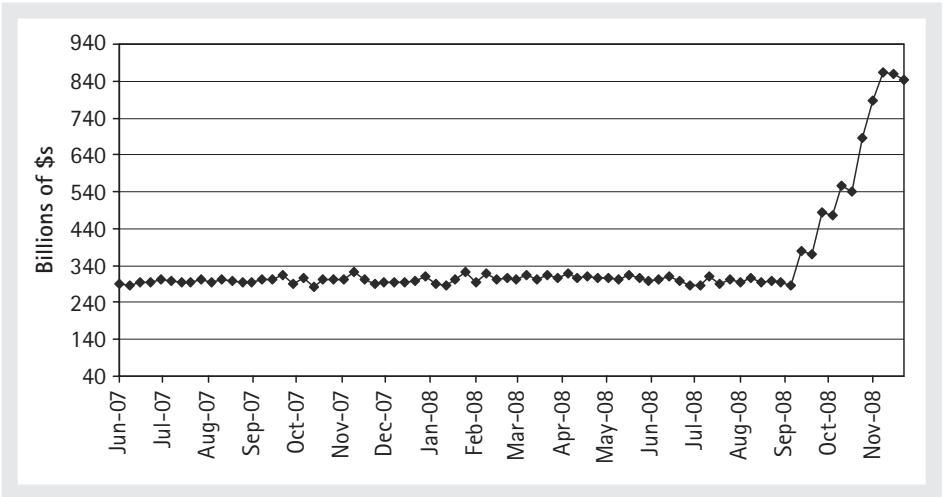

Fig. 5.6. Liquidity demand by US banks: hoarding cash

downs on credit lines; banks responded rationally by hoarding cash (Figure 5.6) and by ceasing to make new loans (Figure 5.3). The increase in loans on bank balance sheets, however, turned around, as did commercial paper outstanding, in the last week of October, because the Federal Reserve began to purchase commercial paper, both directly from issuers and indirectly from mutual funds and other

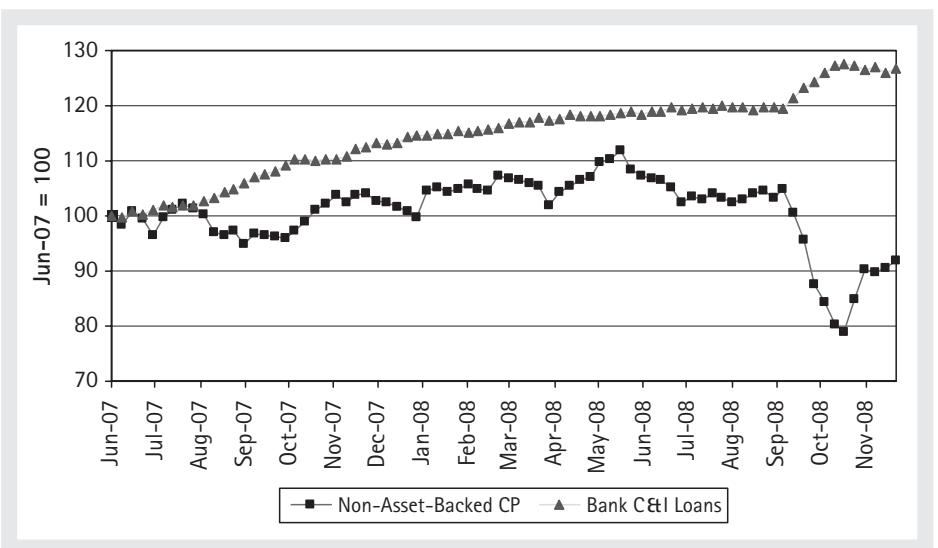

Fig. 5.7. Liquidity demand by US banks: commercial paper moves back on balance sheet

investors. Notice that the turning points in Figure 5.7 for commercial paper outstanding and bank loans correspond *exactly* (week ending on October 29).

Funding liquidity demanded by non-financial firms increased not only to substitute for the absence of market liquidity but also to meet increased *precautionary* demands for cash. Many non-financial firms drew funds from existing lines of credit simply due to fears about disturbances in the credit markets. Ivashina and Scharfstein (2008) present a table summarizing eighteen instances in which large firms drew funds not to meet direct needs for cash but in reaction to concern about debt-market access. To take one example, American Electric Power (AEP) drew down $3 billion from an existing credit line issued by J.P. Morgan and Barclays. According to their SEC filing, 'AEP took this proactive step to increase its cash position while there are disruptions in the debt markets. The borrowing provides AEP flexibility and will act as a bridge until the capital markets improve'. Given cash demands on banks from existing customers, and given the increased cost of borrowing to banks, it is no surprise that new lending has fallen off a cliff (recall Figure 5.3).

How did banks meet the drastic increase in demands for funding liquidity? Figure 5.8 reports the three main sources of funds flowing into US banks. First, deposits flowed in, with the total stock of transactions deposits rising by about $200 billion between the middle of September and the middle of November. The deposit inflows, it should be noted, were *not* evenly distributed across banks. For example, Bank of America and J.P. Morgan experienced large inflows, while others, such as Washington Mutual and Wachovia, faced sufficient outflows to force their closure by the Federal Deposit Insurance Corporation. Second, lending from the

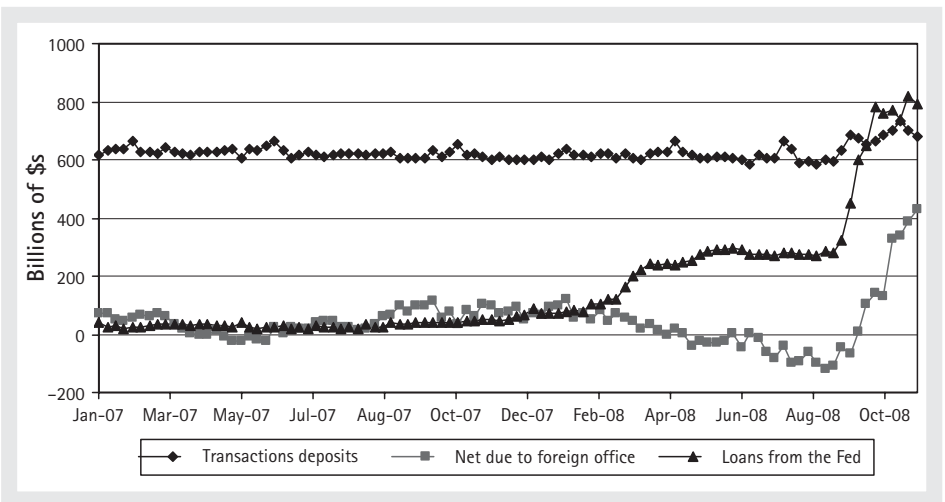

Fig. 5.8. Liquidity supply to US banks comes from Fed loans, transactions deposits, and foreign deposits

Federal Reserve to banks and other financial institutions increased from nearly zero to almost $900 billion. This flood of liquidity, as noted above, did not spur additional lending but merely acted to swell banks' buffer stocks of cash and allow them to meet demands from existing credit lines. Third, starting in the middle of October, US banks raised almost $500 billion from their foreign affiliates and branches. These foreign flows reversed the liquidity outflows from US banks' foreign operations that began in the middle of 2007, when the magnitude of the crisis was first recognized. What changed were perceptions of the impact of the crisis globally. In the middle of October, the contagion into Europe and Asia began, leading to government expansion of guarantees of both deposits and interbank loans, first in Ireland, Germany, and the UK and then very quickly matched by the US. The flow of funds into foreign arms of US banks suggests that investors have greater faith in US guarantees than those of other governments. The appreciation of the dollar and the declining yields on Treasuries, despite a massive increase in supply, are both consistent with this interpretation.

Lessons from the 2007–8 financial crisis

What are the lessons of the crisis of 2008 for liquidity risk management?[8] First, the crisis supports the broad thesis of this chapter, which is that banks have a special advantage in managing funding liquidity risk. As we have seen, access to deposits and central bank liquidity were both critically important factors allowing banks to support liquidity guarantees to the non-financial sector. Depositories that did fail—Countrywide, IndyMac, Washington Mutual, and Wachovia—faced runs having to do with rational concerns about their *solvency*; these institutions were all heavily exposed to subprime mortgages. Moreover, the five large stand-alone investment banks—Bear Stearns, Merrill Lynch, Lehman Brothers, Goldman Sachs, and Morgan Stanley—have either failed, been purchased by a large bank holding company, or registered themselves as bank holding companies. These investment banks had relied on wholesale short-term credit markets, which proved to be the most sensitive to the crisis. Anecdotal evidence suggests that these institutions intend to raise deposits as a more stable source of funds. For example, according to press accounts, Goldman Sachs will develop an Internet bank to raise deposits, expand its existing Utah-based bank, and rebrand its trust business into a full-service bank with its name changed from Goldman Sachs Trust Co. to Goldman Sachs Bank USA (*American Banker*, 4 December 2008).

Second, we have learned that the interbank market, which efficiently circulates liquidity during normal times, can fail spectacularly. Liquidity outside of the

[8] There are certainly lessons beyond liquidity, such as flaws in the originate-to-distribute model of lending.

overnight market dried up during the crisis, raising the price of credit not just to banks but also to non-financial borrowers whose loan rates are typically tied to interbank rates (e.g., three-month LIBOR). As we have seen, bank willingness to extend new credit also dried up. The failure of the interbank market stemmed from uncertainty about the solvency of counter-parties in an environment of asymmetric information. Growth of credit derivatives almost surely worsened the asymmetric information problem by lowering the cost of trading risks. Credible estimates of the aggregate hole in bank balance sheets are on the order of $500 billion to $1 trillion; losses of this magnitude could clearly bankrupt many institutions. Moreover, because it was very hard for outsiders to evaluate *how* these losses were distributed across firms, the rational response was to ration credit to everyone. This asymmetry of information can explain why the interbank market failed to recirculate the liquidity supplied by the Federal Reserve. In fact, banks' cost of borrowing did not begin to decline until the Treasury attempted to reduce concerns about solvency by injecting public funds into banks, raising limits of deposit insurance, and guaranteeing interbank credit (recall Figure 5.5).

Third, much of the funding liquidity risk stemmed from short-term wholesale credit markets and from off-balance sheet commitments, rather than from deposits. To the extent that depositors ran, they ran away from insolvent banks and toward solvent ones. In aggregate, depositors added liquidity to the banking system. The liquidity stresses on the banking system came from increased takedowns on existing lines of credit; these takedowns increased systematically in response to the lack of liquidity in the debt markets. Some of the demand occurred mechanically because commercial paper issuers could not refinance, and other demands stemmed from non-financial firms' precautionary motive to build up a warchest of cash. Ironically, runs from depositors had been the concern that motivated building the safety net for the banking system back in the 1930s. But it is runs *to* the banking system and *away* from the markets that generate systematic funding liquidity risk for banks today.

Conclusions

Banks provide both funding liquidity and market liquidity in various ways. Some of these liquidity-producing activities are unique to banking (or are dominated by banks), such as holding deposits and issuing lines of credit. Others are similar to liquidity provision by non-bank intermediaries like investment banks. For example, both securitization and loan syndication share features in common with bond underwriting. The differences are related to the greater information asymmetry and incentive problems—in the case of bond and equity underwriting,

investment banks sell all of an issuance to arm's-length investors; in securitization and syndication, only a portion of the funding comes at arm's length. Moreover, both commercial as well as investment banks and unregulated finance companies are active securitizers. Both investment banks and large commercial banks also make markets in over-the-counter derivatives such as interest rate swaps and foreign exchange, which enhance market liquidity of those assets. In general, *market* liquidity production is something that is not unique to banking.

What *is* different about commercial banks, what distinguishes them from other intermediaries, are products like checkable deposits and loan commitments. These products supply *funding* liquidity to customers; they offer cash on demand. Banks' 'special role' lies mainly in providing this funding liquidity, but their day-to-day business has increasingly involved provision of market liquidity as a consequence of the growth and deepening of securities markets. This changing role can be seen in the evolution of syndicated lending, where banks typically continue to dominate in the market for credit lines in both the primary and secondary markets (Gatev and Strahan, 2008). In contrast, non-bank institutional investors play an important role in term lending in the syndicated market.

Looking ahead, banks will probably continue to provide liquidity in both dimensions. The Financial Crisis of 2007–8 seems to have strengthened bank dominance at the expense of the large stand-alone investment banking model. Until the crisis, the traditional asset transformation role of banks—holding loans financed with liquid deposits—was on the wane. The growth of securitization seemed to offer cheaper ways to finance loans, although banks continued to provide the funding liquidity support through backup lines of credit. But, this originate-to-distribute model went too far, facilitating the lax underwriting standards that fueled the credit bubble at the heart of the ongoing crisis. Looking ahead, it seems plausible that the traditional model of bank lending and asset transformation may become increasingly attractive.

References

Agarwal, S., Ambrose, B. W., Chomsisengphet, S., and Liu, C. (2006). 'An Empirical Analysis of Home Equity Loan and Line Performance', *Journal of Financial Intemerdiation*, 15: 444–69.

—— Chomsisengphet, S., and Driscoll, J. C. (2004). 'Loan Commitments in Private Firms', mimeo.

Akrelof, G. (1970). 'The Market for Lemons', *Quarterly Journal of Economics*, 84: 488–500.

Almeida, H., Campello, M., and Weisbach, M. (2004). 'The Cash Flow Sensitivity of Cash', *Journal of Finance*, 59: 1777–804.

Allen, F. and Gale, D. (2000). 'Financial Contagion', *Journal of Political Economy*, 108.

Ambrose, B. W., LaCour-Little, M., and Sanders, A. (2004). 'The Effects of Conforming Loan Status on Mortgage Yield Spreads: A Loan Level Analysis', *Real Estate Economics*, 32: 541–69.

Ashcraft, A., McAndrews, J. J., and Skeie, D. (2007). 'Precautionary Reserves and the Interbank Market', mimeo, August 2007.

Ayotte, K. and Gaon, S. (2006). 'Asset-Backed Securities: Costs and Benefits of Bankruptcy Remoteness', mimeo.

Bagehot, W. [1873]. (2006) *Lombard Street: A Description of the Money Market*, London: John Murray (reprinted Gloucester: Dodo Press).

Berger, A. N. and Bouwman, C. H. S. (forthcoming). 'Bank Liquidity Creation', *Review of Financial Studies*.

—— Miller, N., Petersen, M., Rajan, R. G., and Stein, J. (2005). 'Does Function Follow Form? Evidence from the Lending Practices of Large and Small Banks', *Journal of Financial Economics*, 76: 237–69.

—— and Udell, G. (1992). 'Some Evidence on the Empirical Significance of Credit Rationing', *Journal of Political Economy*, 100: 1047–77.

Berlin, M. and Mester, L. J. (1999). 'Deposits and Relationship Lending', *Review of Financial Studies*, 12: 579–607.

Bernanke, B. S. (1983). 'Nonmonetary Effects of the Financial Crisis in Propogation of the Great Depression', *American Economic Review*, 73: 257–76.

—— and Lown, C. (1991). 'The Credit Crunch', *Brookings Papers on Economic Activity*, 2: 205–47.

Brunnemeir, M. and Pedersen, L. (forthcoming). 'Market Liquidity and Funding Liquidity', *Review of Financial Studies*.

Calomiris, C. (1989). 'The Motivations for Loan Commitments Backing Commercial Paper', *Journal of Banking & Finance*, 13: 271–77.

—— (1994). 'Is the Discount Window Necessary: A Penn Central Perspective', *Federal Reserve Bank of Saint Louis Review*, 31–55.

—— and Kahn, C. (1991). 'The Role of Demandable Debt in Structuring Optimal Banking Arrangements', *American Economic Review*, 81: 497–513.

—— and Mason, J. P. (1997). 'Contagion and Bank Failures during the Great Depression: The Chicago Banking Panic of June 1932', *American Economic Review*, 87: 863–84.

—— —— (2004). 'Credit Card Securitization and Regulatory Arbitrage', *Journal of Financial Services Research*, 26: 5–27.

Chava, S. and Purnanandam, A. K. (2006). 'The Effect of Banking Crisis on Bank-Dependent Borrowers', EFA 2006 Zurich Meetings, available at SSRN: http://ssrn.com/abstract=821804.

Cole, R. A. (1998). 'The Importance of Relationship to the Availability of Credit', *Journal of Banking & Finance*, 22: 959–77.

—— Goldberg, L. G., and White, L. J. (2004). 'Cookie Cutter vs. Character: The Micro Structure of Small Business Lending by Large and Small Banks', *Journal of Financial and Quantitative Analysis*, 39: 227–51.

Demarzo, P. (2005). 'The Pooling and Tranching of Securities: A Model of Informed Intermediation', *Review of Financial Studies*, 18: 1–35.

Demyanyk, Y. and Van Hemert, O. (2008), 'Understanding the Subprime Mortgage Crisis', mimeo, http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=1020396.

Dennis, S. A. and Mullineaux, D. J. (2000). 'Syndicated Loans', *Journal of Financial Intermediation*, 9: 404–26.
Diamond, D. (1984). 'Financial Intermediation and Delegated Monitoring', *Review of Economic Studies*, 51: 393–414.
—— and Dybvig, P. (1983). 'Bank Runs, Deposit Insurance, and Liquidity', *Journal of Political Economy*, 91: 401–19.
—— and Rajan, R. G. (2001). 'Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking', *Journal of Political Economy*, 91: 401–19.
Drucker, S. and Puri, M. (2005). 'On the Benefits of Concurrent Lending and Underwriting', *Journal of Finance*, 60: 2763–99.
—— —— (2007). 'On Loan Sales, Contracting and Lending Relationships', available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=920877.
Edwards, F. (1999). 'Hedge Funds and the Collapse of Long-Term Capital Management', *Journal of Economic Perspectives*, 13: 189–209.
Estrella, A. (2002). 'Securitization and the Efficacy of Monetary Policy', *Federal Reserve Bank of New York Economic Policy Review*, 1–13.
Fama, E. F. (1985). 'What's Different about Banks?', *Journal of Monetary Economics*, 15: 29–39.
Flannery, M. (1994). 'Debt Maturity and the Deadweight Costs of Leverage: Optimally Financing Banking Firms', *American Economic Review*, 84: 320–31.
Frame, W. S. and White, L. J. (2005). 'Fussing and Fuming about Fannie and Freddie: How Much Smoke, How Much Fire?', *Journal of Economic Perspectives*, 19: 159–84.
Franke, G. and Krahnen, J. P. (2004). 'Default Risk-Sharing between Banks and Markets: The Contribution of Collateralized Loan Obligations', in Mark Carey and René Stulz (eds.), *Risks of Financial Institutions*, Chicago, IL: University of Chicago Press, 603–31.
Furfine, C. (2002). 'The Costs and Benefits of Moral Suasion: Evidence from the Rescue of Long-Term Capital Management', *Journal of Business*, 79.
Garber, P. and Weisbrod, D. (1990). 'Banks in the Market for Liquidity', working paper 3381, National Bureau of Economic Research.
Gatev, E. (2006). 'Liquidity Risk and Limited Arbitrage: Are Taxpayers Helping Hedge Funds Get Rich?' working paper, Carroll School of Management, Boston College.
—— and Strahan, P. E. (2006). 'Banks Advantage in Hedging Liquidity Risk: Theory and Evidence from the Commercial Paper Market', *Journal of Finance*, 61: 867–92.
—— —— (2007). 'Liquidity Risk and Loan Syndicates', mimeo.
—— Schuermann, T., and Strahan, P. E. (2006). 'How Do Banks Manage Liquidity Risk? Evidence from the Equity and Deposit Markets in the Fall of 1998', in Mark Carey and René Stulz (eds.), *Risks of Financial Institutions*, Chicago, IL: University of Chicago Press, 105–27.
—— —— —— (2007). 'Managing Liquidity Risk: How Deposit-Loan Synergies Vary through Market Conditions', *Review of Financial Studies*, 22: 995–1020.
Goodfriend, M. and King, R. G. (1988). 'Financial Deregulation, Monetary Policy and Central Banking', *Federal Reserve Bank of Richmond Economic Review* (May/June), 3–22.
Gorton, G. (1988). 'Banking Panics and Business Cycles', *Oxford Economic Papers*, 40: 751–81.
—— and Pennacchi, G. (1995). 'Bank Loan Sales: Marketing Non-Marketable Assets', *Journal of Monetary Economics*, 35: 389–411.
—— and Souleles, N. (2005). 'Special Purpose Vehicles and Securitization', in M. Carey and R. Stulz (eds.), *Risks of Financial Institutions*, Chicago, IL: University of Chicago Press, 549–97.

Greenspan, A. (2004). 'Testimony before the Committee on Banking, Housing, and Urban Affairs', US Senate, February 24, 2004.

Gropp, R., Schulz, J. K., and White, M. J. (1997). 'Personal Bankruptcy and Supply and Demand for Credit', *Quarterly Journal of Economics*, 112: 217–51.

Güner, Barak (2006). 'Loan Sales and the Cost of Corporate Borrowing', *Review of Financial Studies*, 19: 687–716.

Higgins, E. J. and Mason, J. R. (2004). 'What is the Value of Recourse to Asset-Backed Securities? A Clinical Study of Credit Card Banks', *Journal of Banking & Finance*, 28: 857–74.

Holmstrom, B. and Tirole, J. (1997). 'Financial Intermediation, Loanable Funds, and the Real Sector', *Quarterly Journal of Economics*, 112: 663–91.

Ivashina, V. (2006). 'Agency Effects on Syndicated Loan Rates', unpublished working paper, Harvard Business School.

—— and Scharfstein, D. (2008). 'Bank Lending during the Financial Crisis of 2008', unpublished working paper, Harvard Business School, November 5, 2008.

—— and Sun, Z. (2007). 'Institutional Stock Trading on Loan Market Information', mimeo.

Iyer, R. and Puri, M. (2007). 'Who runs? The Importance of Relationships in Bank Panics', mimeo.

Jensen, M. (1986). 'Agency costs of free cash flow, corporate finance and takeovers', *American Economic Review*, 76: 323–29.

Jimenez, G., Lopez, J.A., and Saurina, J. (2007). 'Empirical Analysis of Corporate Lines of Credit', mimeo.

Jones, J., Lang W. and Nigro, P. (2005). 'Agent Bank Behavior in Bank Loan Syndications', *Journal of Financial Research*, 28: 385–402.

Kane, E. and Demirgüç-Kunt, A. (2002). Deposit Insurance around the World: Where Does It Work?', *Journal of Economic Perspectives* (Spring), 175–95.

—— (2007). 'Incentive Conflict in Central Bank Responses to Sectoral Turmoil in Financial Hub Countries', mimeo.

Kashyap, A., Rajan, R. G., and Stein, J. (2002). 'Banks as Liquidity Providers: An Explanation for the Co-existence of Lending and Deposit-Taking', *Journal of Finance*, 57: 33–73.

Keys, B. J., Mukerjee, T. K., Seru, A., and Vig, V. (2008). 'Did Securitization Lead to Lax Screening? Evidence from Sub-prime Loans', mimeo, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1093137.

Khwaja, A. I. and Mian, A. (2005). 'Tracing the Impact of Bank Liquidity Shocks', mimeo, June 2005.

Kindleberger, C. P. (2000). *Manias, Panics and Crashes: A History of Financial Crises*, London: MacMillan, 5th edition.

Krishnamurthy, A. (2008). 'The Financial Meltdown: Data and Diagnoses', working paper, http://www.nber.org/~confer/2008/cff08/krishnamurthy.pdf.

Lee, S. and Mullineaux, D. J. (2004). 'Monitoring, Financial Distress, and the Structure of Commercial Lending Syndicates', *Financial Management*, 33: 107–30.

Leland, H. and Pyle, D. E. (1977). 'Information Asymmetries, Financial Structure and Financial Intermediation', *Journal of Finance*, 32: 371–87.

Loutskina, E. (2005). 'Does Securitization Affect Bank lending? Evidence from Bank Responses to Funding Shocks', mimeo, March 2005.

—— and Strahan, P. E. (2007). 'Securitization and the Declining Impact of Bank Financial Condition on Loan Supply: Evidence from Mortgage Acceptance Rates', forthcoming, *Journal of Finance*, 64: 861–89.

——— ——— (2008). 'Informed and Uninformed Investment in Housing: The Downside of Diversification', working paper, September 2008.

McAndless, G., Gabrielli, M. F., and Rouillet, M. J. (2003). 'Determining the Causes of Bank Runs in Argentina during the Crisis of 2001', *Revista de Analisis Económico*, 18: 87–102.

McAndrews, J. J. and Potter, S. M. (2002). 'Liquidity Effects of September 11, 2001', *Federal Reserve Bank of New York Economic Policy Review*, 8: 59–79.

——— Sarkar A., and Wang, Z. (2002). 'The Effect of the Term Auction Facility on the London Inter-Bank Offered Rate?', working paper, Federal Reserve Bank of New York.

Mester, L., Nakamura, L., and Renault, M. (2006). 'Transactions Accounts and Loan Monitoring, forthcoming', *Review of Financial Studies*, 20: 529–56.

Mian, A. and Sufi, A. (2008), 'The Consequences of Mortgage Credit Expansion: Evidence from the 2007 Mortgage Crisis', mimeo, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1072304.

Morgan, D. P. (1998). 'The Credit Effects of Monetary Policy: Evidence from Loan Commitments', *Journal of Money, Credit and Banking*, 20: 102–18.

Myers, S. C. and Rajan, R. G. (1998). 'The Paradox of Liquidity', *Quarterly Journal of Economics*, 113: 733–71.

Nandy, D. and Shao, P. (2007). 'Institutional Investment in Syndicated Loans', mimeo.

Norden, L. and Weber, M. (2007). 'Checking Account Information and The Credit Risk of Bank Customers', mimeo.

O'Hara, M., and Shaw, W. (1990). 'Deposit Insurance and Wealth Effects: The Benefit of Being Too Big to Fail', *Journal of Finance*, 45: 1587–1600.

Paravisini, D. (2007). 'Local Financial Constraints and Firm Access to External Finance', *Journal of Finance*, 63: 2161–93.

Passmore, W., Sherlund, S. M., and Burgess, G. (2005). 'The Effect of Housing Government-Sponsored Enterprises on Mortgage Rates', *Real Estate Economics*, 33: 427–63.

Pennacchi, G. (2006). 'Deposit Insurance, Bank Regulation and Financial System Risk', *Journal of Monetary Economics*, 53: 1–30.

Purnanandam, A., (2008). 'Originate-to-Distribute Model and the Sub-Prime Mortgage Crisis', http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1167786.

Rochet, J. C. and Vives, X. (2004). 'Coordination Failures and Lender of Last Resort: Was Bagehot Right After All?', *Journal of the European Economic Association*, 2: 1116–47.

Saidenberg, M. R. and Strahan, P. E. (1999). 'Are Banks Still Important for Financing Large Businesses?' *Federal Reserve Bank of New York Current Issues in Economics and Finance*, 5: 1–6.

Standard & Poors (2006). *A Guide to the Loan Market*, September 2006.

Sufi, A. (2007a). 'Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans', *Journal of Finance*, 62: 629–68.

——— (2007b). 'Bank Lines of Credit in Corporate Finance: An Empirical Investigation', *Review of Financial Studies*.

Townsend, R. M. (1979). 'Optimal Contracts and Competitive Markets with Costly State Verification', *Journal of Economic Theory*, 21: 1–29.

Udell, G. (2004). *Asset-Based Finance: Proven Disciplines for Prudent Lending*, New York: Commercial Finance Association.

Wittenberg-Moerman, R. (2006). 'The Role of Information Asymmetry and Financial Reporting Quality in Debt Contracting: Evidence from Secondary Loan Market', mimeo.

CHAPTER 6

DIVERSIFICATION IN BANKING

KEVIN J. STIROH[1]

Introduction

The turbulence in financial markets since mid-2007 continues to reshape the industry and raises fundamental questions about how large, complex financial firms operate. Greater scale and scope and wide diversification, both geographically and across products, were expected by some to reduce risk and insulate larger firms from macroeconomic or financial market shocks. While the crisis has affected financial firms of all sizes, it is notable that many large, diversified firms have been among the most impaired as real-estate-related problems spread across a wide range of products and geographies.

While it is too early fully to understand the implications of the financial crisis, it is useful to review what is known about the impact of diversification on the risk and return of US financial institutions. This can help policymakers to better understand the potential and limitations of diversification across alternative business models and to implement a more effective policy response when thinking about longer-term regulatory reform issues.

The second section examines potential explanations for why banks diversify at all. This is a natural first step because classical finance theory suggests that internal diversification is not efficient, as investors can easily shed any firm-specific risk by

[1] I thank Matt Botsch and the editors for helpful comments on an earlier draft.

holding a well-diversified portfolio. There are good reasons reflecting market frictions, however, that explain why managers may choose to diversify and why this may be valuable to a wide range of stakeholders including equity-holders, borrowers, regulators, and the managers themselves.

The third section then reviews the empirical literature that investigates the impact of diversification on the risk and return of financial firms. Both the earlier literature that examined US banks in a relatively regulated environment and the more recent literature that examined their performance in the last few years provide no consensus view—some studies report evidence of significant diversification gains, while other do not. To summarize broadly, studies that looked at counterfactual mergers between banks and non-bank firms tended to find evidence of potential diversification benefits. These studies, however, cannot account for the endogeniety of risk taking and thus must be interpreted cautiously. In contrast, studies that focused on accounting measures tended to showed evidence of greater risk after product diversification, particularly when measured by the growth of non-interest income. Studies that examine equity market returns, however, showed mixed evidence about the impact of diversification on the total risk of financial firms. This divergence undoubtedly reflects differences in methodology, data, sample, and time period and is consistent with basic finance theory, but raises interesting issues about the impact of adopting a more diversified set of financial activities.

The fourth section concludes with potential interpretations of the results and a discussion of implications for financial market participants. I raise some questions for bank supervisors interested in maintaining financial stability and a healthy banking sector in the future, and for researchers interested in better understanding the impact of diversification on financial institutions.

Understanding diversification

The fundamental motivation for this chapter is the observation that large banks, particularly in the US, have become substantially more diversified in terms of product mix and geography over the last two decades. To provide some perspective, Table 6.1 reports summary statistics for the five largest bank holding companies (BHCs) in 1986, 1996, and 2006 in the US (identified by total BHC assets in Y-9C reports[2] in December of each year). In 1986, the five largest BHCs held about 21 percent of aggregate bank assets with about two-thirds of those assets in the form of traditional loans. Approximately 40 percent of their net operating revenue (defined as net interest income plus

[2] Consolidated Financial Statements for Bank Holding Companies.

Table 6.1. Evolution of large bank holding companies

	Total assets		Loans/assets	Non-interest income/ net operating revenue	States	Branches	Concentration	
	($B)	Share					Product	Geography
1986								
Citicorp	196	7.6	67.0	39.6	13	537	4,623	6,176
BankAmerica Corp	104	4.0	70.9	39.9	2	1,145	4,489	8,140
Chase Manhattan Corp	95	3.7	70.0	34.2	6	394	4,864	8,555
J.P. Morgan and Co.	76	3.0	45.6	42.2	3	7	3,930	8,881
Manufacturers Hanover Corp	74	2.9	75.3	39.3	3	231	4,616	9,189
Sum	546	21.2						
1996								
Chase Manhattan Corp	336	7.9	49.5	46.9	7	813	3,768	6,364
Citicorp	281	6.6	63.8	45.4	11	396	4,102	5,073
BankAmerica Corp	251	5.9	67.6	37.5	12	2,003	4,492	4,495
J.P. Morgan and Co.	222	5.2	12.7	75.1	3	4	2,762	9,282
Nationsbank Corp	186	4.3	66.7	35.9	11	1,986	4,567	1,596
Sum	1,276	29.8						
2006								
Citigroup	1,884	15.4	38.1	54.9	14	896	4,100	4,309
Bank of America Corp	1,464	11.9	49.4	53.6	31	5,788	3,591	1,045
J.P. Morgan Chase and Co.	1,352	11.0	35.7	65.8	19	2,629	2,870	2,764
Wachovia Corp	707	5.8	62.0	48.6	17	3,063	3,809	1,456
Wells Fargo & Co.	482	3.9	73.6	43.2	23	3,215	4,272	1,719
Sum	5,889	48.0						
1986 Mean	109	4.2	65.8	39.0	5	463	4,505	8,188
1996 Mean	255	6.0	52.0	48.2	9	1,040	3,938	5,362
2006 Mean	1,178	9.6	51.8	53.2	21	3,118	3,728	2,259

Notes: Data for individual bank holding companies are from Y-9C reports in December of each year. Share of total assets is total assets for the individual bank holding company as a percentage of aggregate banking assets as reported by the FDIC in the Historical Statistics on Banking. Net operating revenue N is defined as net interest income plus non-interest income. Each concentration measure is the sum of the squared shares (multiplied by 100) of all items in a category. Product concentration is calculated as the HHI using revenue shares from net interest income, fiduciary income, services, trading, and fee and other income. Geographic concentratrion is calculated as the HHI using the distribution of a BHC's deposits across states, i.e., the share of a particular BHC's deposits that come from each state in which it operates.

non-interest income) in 1986 came from non-interest sources such as fees and commissions, trading, and fiduciary income. Even the largest firms in the banking industry were relatively concentrated geographically, operating in only five states on average.

Over the following two decades, the industry changed dramatically as regulatory constraints loosened and market pressures evolved. The largest BHC grew relative to the industry due to a steady stream of consolidation and a wave of mega-mergers in the last decade. By 2006, the five largest BHCs accounted for half of aggregate banking assets. Along with greater size came changes in strategy and focus—loans became a smaller share of the balance sheet and non-interest income grew to dominate the income statement. DeYoung and Rice (2004b) show a similar trend for the industry as whole with the relative importance of non-interest income more than doubling between 1970 and 2003. These large banks now operate in a much wider geographic footprint with branches in more than twenty states on average. Indeed, an explicit motivation for many of the large bank mergers in recent years has been the desire to create a nationwide franchise (Clark, et al., 2007 and Hirtle and Stiroh, 2007).

To quantify how diversification for the largest BHCs changed across both product mix and geography, I calculated bank-level measures of concentration via Herfindahl-Hirschman Indices (HHIs) that are based on variation in revenue sources and in interstate activities over time.[3] Both measures show steady increases as these BHCs became increasingly diversified by offering a wider range of financial products that generated more-varied income streams and by operating in a wider geographic area.

What might explain the trend toward provision of broader financial services and geographic expansion? Most obviously, earlier regulations may have prevented banks from entering profitable business lines or forced them to enter in inefficient ways, so consolidation and increased diversification could be the normal response of profit-maximizing firms to the relaxation of external constraints. Berger, Demsetz, and Strahan (1999) review the consolidation wave in the 1990s in the US. More fundamentally, these gains could reflect production synergies between lending and other financial activities that create a comparative advantage for the integrated financial services firm. Alternatively, diversification across products may improve the risk–return frontier by expanding the investment opportunity set.

Why do firms diversify?

Portfolio theory shows that diversification—the expansion of investments into activities that are not perfectly correlated—can reduce the risk of the portfolio. In

[3] An HHI index is calculated as the sum of the squared shares (measured in percentages). I used revenue shares from net interest income, fiduciary income, services, trading, and fee and other income for the measure of product concentration. I used the distribution of a BHC's deposits across states for the measure of geographic concentration, i.e. the share of a particular BHC's deposits that come from each state in which it operates.

the context of a firm's strategic decisions, managers can diversify by offering new products or entering new markets. This should reduce risk that is specific to each activity and leave only risk that is common to all activities. That is, internal diversification can eliminate a firm's idiosyncratic risk and leave only its systematic risk.

To be clear about terminology, consider the simple case of a bank's return ($R_{i,t}$) that depends linearly on a single risk factor, the market ($R_{M,t}$), and an idiosyncratic component ($\epsilon_{i,t}$) through the familiar capital asset pricing model (see Fama and French, 2004 and Roll, 1998, for reviews). In this case:

$$R_{i,t} = \alpha_i + \beta_M R_{M,t} + \epsilon_{i,t}$$

The independence of the residuals implies that the variance of returns can be decomposed as:

$$\sigma_i^2 = \hat{\beta}_M^2 \sigma_M^2 + \sigma_{\epsilon,i}^2$$

where σ^2 reflects the variance of the subscripted variable. Following convention, the variance of total returns, σ_i^2, is called 'total' risk and the variance of the residuals, $\sigma_{\epsilon,i}^2$, is called 'idiosyncratic' or 'firm-specific' risk. The part explained by the market factor, $\hat{\beta}_M^2 \sigma_M^2$, is called 'systematic' risk.

Portfolio theory suggests that internal diversification will reduce the idiosyncratic component of volatility, but not systematic risk, so firm diversification can, in principle, reduce idiosyncratic and therefore total risk. But, should firms pursue this strategy? Is it efficient to expend valuable resources in the pursuit of lower overall volatility and risk?

A natural starting-point is the perfect capital market world of Modigliani and Miller. An implication is that firms should not expend valuable resources diversifying, hedging, or on other risk management activities because investors can always buy or sell positions themselves to adjust their exposure. As pointed out by Sharpe (1964), an investor need not be concerned with a firm's idiosyncratic risk because it can be eliminated by holding a well-diversified portfolio. An implication is that investors should not price idiosyncratic risk and should only be concerned with the non-diversifiable, systematic component.

Despite these well-grounded arguments, questions remain. Cummins, Phillips, and Smith ask why 'managers of widely held corporations, acting in the interest of their stockholders, should manage risk that their shareholders could presumably manage themselves' (1998: 33). Winton (1999a) phrases the question differently: 'Should lenders diversify, as suggested by the intermediation literature, or specialize, as suggested by the corporate finance literature?' Said still differently, is it useful to reduce idiosyncratic risk through internal diversification?

A large body of research has concluded that there are sound reasons why risk management may be optimal. This is true for investors due to capital market frictions that make the textbook case inappropriate for many firms, particularly financial intermediaries. It is also true for other market participants such as

managers, supervisors, or bank counterparties, all of whom have an interest in the total risk of the financial firm. For these participants, both systematic and idiosyncratic risks impose real costs; to the extent that if diversification reduces a firm's idiosyncratic risk, then diversification is desirable.

In a discussion focused on financial institutions, Froot, Scharfstein, and Stein (1993) and Froot and Stein (1998) highlight several of these channels. One factor is that the costs of external funds may be non-linear, so the value of the firm will depend on the total volatility of returns. If the marginal cost of adjustment rises with the amount of external financing raised, then optimization requires that an adverse shock to cash flow induces both an increase in external finance and a decrease in real investment. Thus, lower variability in cash flow can impact real investment positively, raise the value of the firm, and be desirable for shareholders.

A second friction is that some risks are not marketable—for example, an investor may not be able lay off the risk associated with all idiosyncratic shocks such as the introduction of new products. Froot and Stein (1998) argue that risk associated with this type of illiquid asset is particularly relevant for financial firms—for example, a loan to a small company that is information-intensive and difficult to trade in the secondary market. While recent financial innovations such as a more-liquid loan-sale market or securitization may have reduced this as a concern, it remains relevant for many firms not actually involved in these markets.

Acharya, Hasan, and Saunders (2006) provide additional motivation for financial firms. One, banks are highly regulated entities and these regulations often provide conflicting incentives to diversify or focus activities—for example, capital restrictions tied to the risk of the loan portfolio or branching restrictions. Two, inherent agency problems within a bank, which stem from imperfect information and conflicts between bank owners and bank managers, are likely to be influenced by the risk of insolvency, what they call 'downside riskiness', and a bank's diversification strategy can affect this.

More broadly applicable to all firms, Smith and Stulz (1985) point to the convex nature of the tax code and conclude that firm value will be higher if earnings are more stable. This suggests that shareholders will prefer lower overall volatility, which can be achieved by internal diversification. Smith and Stulz (1985) also show that if there are costs of financial distress such as bankruptcy costs, loss of value during asset sales, and search costs for new management, then shareholders will care about total risk. This may be particularly relevant for financial firms where assets are relatively 'opaque' and hard to value from the outside.

Firm managers may have additional incentives to manage risk and reduce the overall volatility of returns that go beyond value-maximizing motives. Stulz (1984) and Cummins, Phillips, and Smith (1998) suggest that firms manage risks because their managers are risk-adverse and cannot completely diversify when a substantial fraction of their wealth is tied up in a firm's equity. Hughes and Mester (1998) provide evidence that bank managers behave as if they are risk-adverse. As a result, managers may prefer to diversify and reduce total volatility even if this is not in the

best interest of shareholders. Berger, Demsetz, and Strahan (1999), Milbourn, Boot, and Thakor (1999), Bliss and Rosen (2001), Houston, James, and Marcus (1997), and Aggarwal and Samwick (2003) also discuss managers' incentives related to empire-building, corporate control problems, or managerial hubris and self-interest, all of which could also lead to inefficient diversification.

Diversification and risk reduction may also be desirable from the perspective of other participants. Borrowers, for example, will care about the viability of their lenders if the intermediation process is built on private information and long-term relationships. Slovin, Sushka, and Polonchek (1993) show that borrower stock prices fell after the de facto failure of Continental Illinois and interpret this as evidence of the costs of severing intangible banking relationships that are valuable to borrowers. In this view, borrowers are bank stakeholders who care about the total risk of the institution. Similarly, Houston, James, and Marcus (1997) show that diversification of internal capital markets benefits borrowers through the efficient allocation of scarce capital resources. If banks have large variation in revenue, for example, positive net present value project may not be funded in periods of low-realized cash flows. Bank-dependent borrowers, therefore, care about the volatility of revenues and total risk. This may be particularly true for small firms that are dependent on small banks for financing (DeYoung, Hunter, and Udell, 2004).

Finally, total risk is the most relevant metric for supervisors who are concerned with the probability of default and the associated bankruptcy costs. The idea that a more geographically diversified banking system increases financial and economic stability goes back at least to Sprague (1903). From the regulator's perspective, concern is for the costs associated with failure including transaction and liquidation costs related to bankruptcy, systemic risk concerns, and direct costs to the insurance fund from the tax distortions implicit in its funding or increased supervisory resources needed to offset moral hazard. As a result, supervisory interest is not in a diversified portfolio of firms, but in the total risk of each individual institution that is supervised. This can be seen directly in Merton-type portfolio models of credit risk, developed by Merton (1974) and implemented in KMV risk models, which are driven by assumptions about total-asset-return volatility. Moreover, Haubrich (1998) emphasizes that the deposit insurance fund is likely to be more concerned with the expected value of future insurance payments rather than just the probability of a given bank failure, so if diversification benefits are gained along with size, a more diversified (and larger) bank may still be more risky from the deposit insurance fund's perspective. Thus, there are good reasons for supervisors and regulators to be concerned with the total risk of an institution.

A second factor reflects banks' supervisory ratings such as the CAMELS rating,[4] which depend on supervisors' assessment of a bank's ability to absorb future losses,

[4] The components of a bank's condition that are assessed: (C)apital adequacy, (A)sset quality, (M)anagement, (E)arnings, (L)iquidity, and (S)ensitivity to market risk.

the sensitivity of earnings to economic changes, and management's ability to measure, monitor, and control risk (Berger, Klapper, and Udell, 2001). Because the degrees of supervisory oversight and regulatory burden depend on these ratings, managers will care about total risk due to this supervisory effect.

Taken together, these arguments suggest that there are good reasons why shareholders, managers, borrowers, and supervisors are all concerned with the total risk of individual US banks. As a consequence, this suggests that internal diversification may be efficient and desirable as it can reduce idiosyncratic risk and total risk.

Two decades of deregulation and the expansion of bank activities

The summary statistics in Table 6.1 show a trend toward more-diversified banking activities in the US over the last two decades, in terms of both revenue sources and geographic exposure. Given the previous arguments for why financial firms may desire to be diversified, the next step is to consider how regulatory constraints may have hindered earlier attempts to diversify and how the massive deregulation of US financial markets over the last two decades facilitated it. This discussion is largely based on the more detailed treatment in Spong (1994), Berger, Demsetz, and Strahan (1999), and Strahan and Sufi (2000).

In the aftermath of the stock market crash of 1929 and in fear of conflicts of interest between commercial and investment banking, regulators moved to sever the link. The Banking Act of 1933 (the Glass-Steagall Act) limited financial integration by preventing any firm that accepts demand, time, or saving deposits from also engaging in most investment banking activities such as issuing, underwriting, selling, or distributing stocks, bonds, or other financial securities. This essentially prohibited the universal banking model common in other countries, particularly Europe.

The McFadden Act of 1927 granted national banks the same ability as state banks to branch within their own state, which limited interstate branching because most states had branching restrictions. Moreover, this was interpreted as preventing national bank branches on an interstate level (Spong, 1994). Several decades later, in response to attempts by some banks to circumvent existing constraints, the Bank Holding Company Act of 1956 imposed geographic restrictions—for example, the Douglass Amendment prohibited interstate acquisition unless state law specifically authorized it. Most states did not authorize it, however, so interstate banking through the holding company structure was halted. The Bank Holding Company Act also prevented bank holding companies from owning or controlling non-bank activities except under very specific circumstances.

As a result of this long history of regulation, US banks in the 1970s were highly restricted in terms of both the products that they could offer and where they could

Table 6.2. Major events in bank deregulation

Date	Event
30 Apr. 1987	Federal Reserve authorizes underwriting activity for Bankers Trust, J.P. Morgan, and Citicorp with a 5 percent revenue limit in 'ineligible' activities.
18 Jan. 1989	Federal Reserve expands Section 20 underwriting permissibility to corporate debt and equity securities, subject to revenue limit.
13 Sept. 1989	Federal Reserve raises limit on revenue from Section 20 ineligible activities from 5 to 10 percent.
16 July 1993	Court Ruling in *Independent Insurance Agents of America* v. *Ludwig* upholds OCC decision to allow national banks to sell insurance from small towns.
26 July 1994	Interstate Banking and Branching Efficienct Act (Riegle-Neal Act) passed by Joint Congressional Committee.
18 Jan. 1995	Court ruling in *Nationsbank* v. *VALIC* allows banks to sell fixed and variable annuities.
26 Mar.1996	Court ruling in *Barnett Bank* v. *Nelson* upholds Ludwig and overturns states' remaining restrictions on national bank insurance sales.
30 Oct. 1996	Federal Reserve announces the elimination of many firewalls between bank and non-bank remaining restrictions on national bank insurance sales.
20 Dec. 1996	Federal Reserve raises limit on revenue from Section 20 ineligible activities from 10 to 25 percent.
22 Aug. 1997	Federal Reserve eliminates many of the remaining firewalls between bank and non-bank subsidiaries within BHCs.
6 Apr. 1998	Citicorp and Travelers Group announce intentions to merge.
22 Oct. 1999	Administration and congressional leaders announce compromise legislation on the Financial Services Modernization Act (Gramm-Leach-Bliley Act).

Notes: Information from Strahan and Sufi, 2000: Table 1.

operate geographically, although the 1970 amendment of the Bank Holding Company Act did allow bank subsidiaries to engage in bank-related services that offered public benefits. In the 1980s, however, the regulatory environment began to loosen and banks were able to expand both in terms of the financial products they offered and their geographic footprint; Table 6.2 provides a chronology of this deregulation. The Federal Reserve in 1987, for example, allowed BHCs to underwrite certain securities on a limited basis through their Section 20 subsidiaries.[5] Revenue from these activities, however, could not exceed 5 percent of total revenue for the subsidiary. Banks also gradually obtained the power to provide investment advisory services along with securities brokerage activities.

Over the next few years, additional statutory and regulatory change further expanded the scope of activities such as broader underwriting abilities, increased revenue from non-traditional banking activities, and expansion into insurance sales in certain cases. By 1996, the Federal Reserve allowed Section 20 subsidiaries of BHCs to earn up to 25 percent of the subsidiaries' revenue from underwriting.

[5] Section 20 was the portion of the Glass-Steagall Act that split commercial and investment banking and 'Section 20 subsidiaries' was the name given to a bank holding company subsidiary that engaged in a limited amount of securities activities. See Kwan, 1998 and Cornett, Orrs, and Tehranian, 2002 for detailed descriptions.

This progression culminated in the Financial Services Modernization Act of 1999 (also known as the Gramm-Leach-Bliley Act), which effectively dismantled the Glass-Steagall restrictions and allowed the combination of banking, insurance, and securities activities within the same 'financial holding company' structure. See Furlong (2000) for an overview of the GLBA.

In terms of geographic restrictions, a similar path of gradual expansion unfolded. Several states passed laws allowing interstate entry in the 1970s—for example, Maine passed a bill in 1975 allowing interstate entry, conditional on other states granting reciprocity (Spong, 1994). Over the next decade, several state 'compacts' formed, which allowed interstate banking and some states also amended interstate branching prohibitions. Moreover, the Garn-St Germain Depository Institutions Act of 1982 authorized interstate acquisitions of certain failed banks and the Competitive Equality Banking Act of 1987 broadened this to include a wider set of troubled institutions. This steady relaxation of restrictions culminated in the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994, which allowed interstate mergers and branching after 1997 subject to concentration restrictions, CRA requirements, and all capital adequacy standards.

The cumulative impact was that two decades of deregulation allowed the creation of financial holding companies that offer a wider range of financial products and operate in broad geographic markets. This fundamental change paved the way for the type of widely diversified mega-banks that now dominate US banking markets.

A framework for interpreting broader activities

This section concludes with a brief discussion of how one can interpret the expansion of bank activities—either across products or geography—in an expected risk and return framework. Morgan and Samolyk (2005) use this approach and describe the opportunities to expand geographically as shifts of the risk/return frontier, as does Haubrich (1998). This familiar framework is useful because it allows a clear illustration of an important point—expansion of a bank's opportunity set and greater diversification need not lower observed risk.

As shown in Figure 6.1, a given set of regulatory, market, and technological constraints allows banks to earn higher expected returns only by taking on additional risk. This opportunity set is shown by line A. As is standard, the marginal expected return for increased risk declines with the level of risk. The bank owners' preferences are given by the utility curve 1, where owners trade-off risk for expected return, but increasing amounts of expected return are required as compensation as risk rises. The optimal point is given by the tangent at X_1.

Expansion in a bank's ability to produce a broader set of products or enter new markets expands the opportunity set. Assuming that these activities are less than

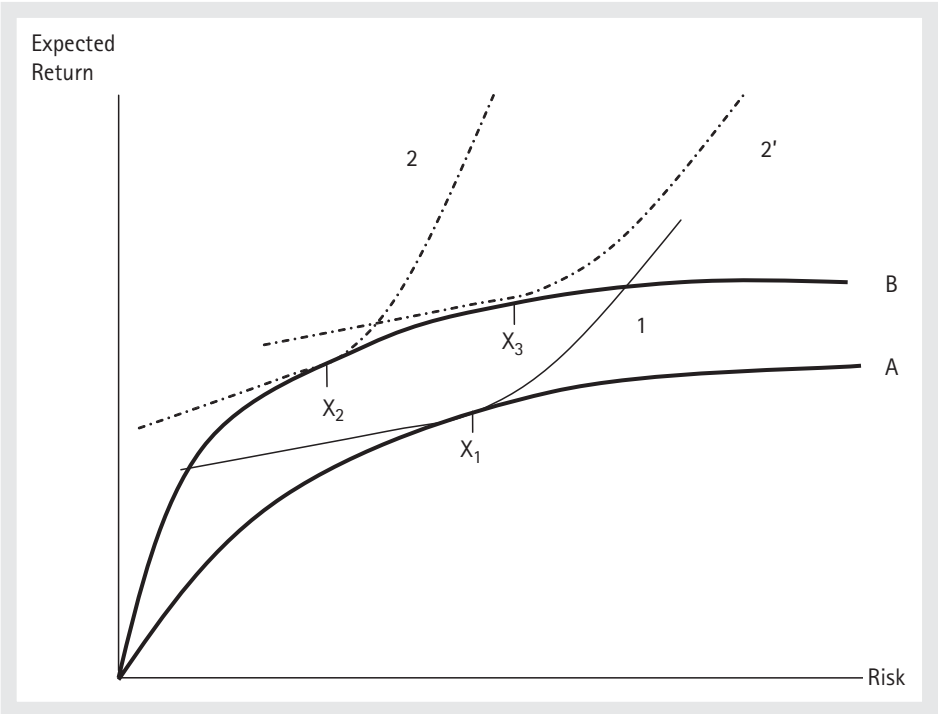

Fig. 6.1. Risk and return when the opportunity set expands

perfectly correlated with the existing set, this ability to diversify allows lower risk without surrendering expected return and effectively shifts the opportunity set up and in to B. Importantly, this expansion need not lead to lower risk taking and the actual outcome will depend on the preferences of bank owners and managers. For example, bank owners that are more risk-averse (with the relatively steep indifference curve 2) may choose to increase returns and reduce risk by shifting from X_1 to X_2. Owners who are less risk-averse (with the relatively flat indifference curve 2′) may choose a combination of higher risk and higher return and shift to X_3.[6] This observation—risk taking is endogenous and diversification need not lead to lower observed levels of risk—is a fundamental implication of standard portfolio theory.

As an example from US bank holding companies, Demsetz and Strahan (1997) showed that larger BHCs were indeed more diversified than smaller ones—for example, bank size was negatively correlated with idiosyncratic risk and

[6] Of course, indifference curves 2 and 2′ are incompatible with each other. The proper comparison is between either 2 or 2′ with 1. Both are shown here for illustrative purposes only.

positively correlated with the explanatory power of a market model regression like the one in Equation (1). These large banks, however, also held less capital and made riskier loans such as commercial and industrial loans, so size was uncorrelated with total risk. In essence, managers seemed to use up the diversification gains to take on more risk and earn higher returns. Demsetz and Strahan concluded that 'large BHCs have not used their superior diversification to reduce risk' (1997: 306).

As a second example from an earlier era, Carlson (2004) argues that branch banks in the 1930s tended to use their diversification advantage to reduce reserves and thereby increase risk, rather than reduce the risk within their loan portfolios. Again, the key point, emphasized by Hughes, Mester, and Moon (1996), is that risk taking is endogenous, chosen by bank managers, so the increased ability to diversify need not be correlated with observed declines in measured risk.

Diversification and risk

The question of whether diversified financial institutions outperform their more concentrated peers is an area of active research and researchers have examined the link between diversification and performance from a variety of perspectives. Given the previous discussion about the endogeneity of risk taking, it is perhaps not surprising that it has been difficult to find a clear and stable link between measures of diversification and measures of risk. In an early survey, for example, Saunders and Walter (1994) reviewed eighteen studies that examined whether non-bank activities reduced BHC risk and found no consensus: nine answered yes, six answered no, and three were mixed.

Both the earlier and the more recent studies approach the diversification/risk question with a variety of methods: creation of counterfactual mergers of banks with non-banks, analysis of accounting results, and analysis of equity market reactions to variation in diversification. This section reviews the literature over the last two decades on the link between diversification and risk from each of these perspectives, while Santomero and Chung (1992), Saunders and Walter (1994), Reichert and Wall (2000), and DeYoung and Roland (2001) review the earlier literature.

Note, however, that I do not cover research that examines the impact of diversification on other variables such as the cost of debt (Deng, Elyasiani, and Mao, 2007), loan pricing or interest margins (LePetit, et al., 2005; LePetit, et al., 2006; and Valverde and Ferndandez, 2007), bond returns (Penas and Unal, 2004), merger returns (Benston, Hunter, and Wall, 1995 and DeLong, 2001), market reaction to regulatory reform (Strahan and Sufi, 2000 and Yu, 2002), market values

(Klein and Saidenberg, 2005 and Laeven and Levine, 2007), or franchise values (Baele, 2007). I also don't directly discuss the broader literature on the 'diversification discount', summarized, for example, by Campa and Kedia (2002) and Laeven and Levin (2007), or the increased market power and potential antitrust challenges from more consolidated and larger institutions.

Counterfactual mergers

Owing to the highly regulated nature of US financial services in the 1980s and early 1990s, many early studies of diversification performed counterfactual mergers across industries by combining income statement and balance sheet information in a pro forma manner. The idea was that by simulating mergers and combining revenue streams one could gauge the impact on volatility from various combinations of banks, securities firms, and insurance firms. If revenue volatility from the combined entity were lower than for the stand-alone entities, then this suggests diversification benefits exist.

Boyd and Graham (1988) and Boyd, Graham, and Hewitt (1993) simulated mergers between BHCs and non-bank financial firms using data from the 1970s and 1980s. Their primary conclusion was that the biggest gains from diversification in the form of lower risk would probably come from combinations between BHCs and life insurance firms. In contrast, bank mergers with securities firms or real-estate firms would probably increase risk, they found. Saunders and Walter (1994) also found reduced risk, measured as less volatile market returns, for diversified firms through simulated mergers. Rose (1989) compared financial and non-financial firms from 1966 to 1985 and reported that the observed cash flow correlation between banking and financial service lines was positive and small, which suggested some diversification benefits.

Lown, et al. (2000) performed similar counterfactual combinations of BHCs with other financial firms and also concluded that life insurance company mergers provided the greatest potential for risk reduction. Allen and Jagtiani (2000) found evidence of significant diversification benefits in the form of lower total risk from the potential merger of a bank, a securities firm, and an insurance firm. Systematic risk, however, would probably rise as securities firms in particular bring greater market risk, which limits the combined firm's ability to diversify. This two-sided impact from the expansion of bank powers—reduced risk from diversification and increased risk from greater exposure to more volatile activities—is a common finding echoed in subsequent research.

Emmons, Gibert, and Yeager (2004) performed a counterfactual merger exercise focusing on potential diversification benefits for community banks. These banks are typically quite small and geographically focused, so they are heavily exposed to potentially diversifiable risk because of the small number of borrowers and

geographic concentration. They concluded, however, that idiosyncratic risk associated with the small number of customers is quantitatively much larger than the local market risk associated with the geographic concentration. Thus, for these community banks, the evidence pointed toward scale effects rather than diversification effects as a means to reduce failure risk.

Santomero and Chung (1992) took a different approach to the simulation strategy by employing option pricing theory to estimate the implied volatilities of asset returns from possible mergers between bank and non-bank firms. Their results indicated that consolidation between bank and non-bank businesses would lead to a lower probability of failure. In particular, they concluded that bank mergers with securities firms would not materially increase risk, while potential mergers with property/casualty insurance would increase risk but also increase expected returns. Perhaps most interestingly, they estimated that a universal bank that provided all financial activities would be the most stable of all.

Estrella (2001) extended this approach to a later time period with a somewhat different empirical method. Using option pricing techniques, he found evidence of bilateral diversification gains from mergers involving banks and insurance firms. Using arbitrage pricing theory to provide more intuition, the data showed that financial sector returns are driven by only a few factors and there is not much difference in the most important factors across industries. Estrella concluded that there was strong evidence for potential diversification gains between banks and insurance firms.

Slijkerman, Schoenmaker, and de Vries (2005) also used market data to examine the potential for diversification benefits in Europe, but their focus was on extreme events. In particular, they used extreme value theory to consider whether downside risk (the probability of a crash) is the same for European banks as for European insurance firms, and whether diversification can reduce this risk. Using market data from 1992 to 2003, they concluded that diversification benefits for financial conglomerates (banks/insurance firms) exceeded those for large, stand-alone banks owing to the relatively low dependence of returns across financial sectors.

Taken together, these counterfactual merger studies generally found evidence of diversification benefits, particularly between banks and insurance firms. An important caveat, however, is that their counterfactual nature necessarily ignores the endogeneity of risk taking and changes in behavior that managers may make in response to diversification gains. Moreover, they ignore both potential benefits such as scale and scope economies and the potential costs of mergers such as increased agency costs or culture conflicts that could impact both returns and volatility. These concerns are reasonably compelling and suggest considerable caution when interpreting the results, so I now turn to the empirical evidence on the actual performance of diversified financial firms.

Accounting studies

The second strand of research examined actual return and volatility data for banking firms involved in a broad set of financial activities using accounting data, primarily from published regulatory reports such as the Call Reports for US banks or the Y-9C data for US bank holding companies (BHCs). These data are popular because they are readily available for a large number of institutions and are reported on a relatively consistent basis over time.

Product market diversification

Several studies examined variation in performance across different types of BHC subsidiaries. Kwast (1989) studied the impact of the steady expansion of bank securities activities described in the second section above by comparing returns on securities and non-securities activities from 1976 to 1985. He concluded that there were only limited diversification benefits. Kwan (1998) performed a similar exercise and found that Section 20 subsidiaries of BHCs were riskier but not more profitable, on average, than other subsidiaries. Return correlations, however, were low, so some diversification benefits probably existed.

DeYoung and Roland (2001) compared bank profitability and volatility with revenue shares for large commercial banks from 1988 to 1995 and concluded that increased reliance on fee-based activities (revenue from all sources except loans, investment, deposit, and trading activities) did not reduce the volatility of earnings. Similarly, Stiroh (2004b) concluded that a greater reliance on non-interest income, particularly trading revenue, was associated with higher volatility and lower risk-adjusted profits in a cross-section of banks for the period 1979 to 2000.

Stiroh and Rumble (2006) performed a similar analysis with BHC data and concluded that diversification benefits existed when looking across BHCs, but these gains were more than offset by increased exposure to more volatile activities so risk-adjusted performance suffered. To be more precise, consider the return volatility on a portfolio of bank activities. If the bank can engage in two activities, *A* and *B*, then the expected return of the portfolio, $E(R_P)$, and variance, σ_P^2, are:

$$E(R_P) = wE(R_A) + (1 - w)E(R_B)$$
$$\sigma_P^2 = w^2\sigma_A^2 + (1 - w)^2\sigma_B^2 + 2w(1 - w)Cov(R_A,R_B)$$

where w is the weight, $E(R)$ and σ^2 are the expected return and variances of the subscripted variables and $Cov(R_A,R_B)$ is the covariance between returns on *A* and *B*.

Consider the impact of regulatory or technological change that induces an increase in the relative importance of activity A. If activity A offers higher and more volatile returns, then a shift toward *A* has several effects: higher expected portfolio returns because $E(R_A) > E(R_B)$, a direct increase in portfolio variance if the weighted variance of *A* exceeds the weighted variance of *B*, and an indirect

diversification benefit if there is less than perfect correlation. The results in Rosen, et al. (1989) and Stiroh and Rumble (2006) both indicated that the increased share of volatile non-interest activities outweighed the diversification benefits.

Stiroh (2004a) found consistent results for community banks in the US, although he also reported evidence of diversification gains both within the loan portfolio and within non-interest income streams. Goddard, et al. (2008) showed similar results for US credit unions where greater exposure to non-intest income was linked with more volatile accounting returns, while diversification tended to lower volatility. On net, the direct impact of more exposure to non-interest income essentially offset the diversification gains.

For Europe, Tarazi, et al. (2008) and LePetit, et al. (2005) examined European banks from 1996 to 2002 and found that increased non-interest income exposure was positively linked with measures of risk (both accounting and equity-market-based measures). This link was strongest for small banks and was driven by activities that generated commissions and fees. Similarly, Hayden, Porath, and Westernhagen (2006) examined German banks for the same period and concluded that benefits of diversification are difficult to find. Using a more detailed measure of diversification for German banks, Kamp, et al. (2007) concluded that specialized lending banks tended to have slightly higher returns and better asset quality, but also more volatile provisions and asset quality.

Mercieca, Schaeck, and Wolfe (2007) examined a set of small European credit institutions from 1997 and 2003 and found no evidence of direct diversification benefits. This result is similar to the analysis of US community banks by Stiroh (2004a) as a greater reliance on non-interest income was associated with weaker risk-adjusted performance in both studies.

DeYoung and Rice (2004b) compared a variety of banking strategies such as traditional banks, non-traditional banks, corporate banking, community banks, and diversified banks and found a clear risk/return trade-off using both accounting and equity market return data for 1993 to 2003. This suggests that many strategies are viable—for example, high risk and high return in corporate banking vs. low risk and low return in community banking, and the choice will reflect managers' preferences. This view is supported by evidence in Hirtle and Stiroh (2007), who found that retail banking activities offered a combination of relatively low returns and low volatility, and suggests that it is important to recognize that these banks will operate at different points along the risk/return frontier.

Landskroner, Ruthenberg, and Zaken (2005) examined the link between diversification and performance for universal banks in Israel from 1992 to 2001. This has an advantage relative to US studies due to the longer time period when these institutions actually engaged in a broad range of financial services such as mortgage banking, international banking, investment banking, insurance, and commerce, but the sample size is much smaller. They found strong evidence of diversification benefits and concluded that the banks appear to be operating near the efficient frontier.

Finally, Jorion (2005) focused on the diversification benefits of trading activities for large banks from 1995 to 2003 by examining trading revenues and VAR-based market risk charges. He found substantial diversification across business lines associated with trading.

Acharya, Hasan, and Saunders (2006) looked for diversification within the loan portfolio by examining the expansion of loans into new sectors for a set of Italian banks from 1993 to 1999. They concluded that loan diversification tended to reduce returns (both accounting and equity market-return data), while also producing riskier loans for high-risk banks and offering no or only modest improvements for low-risk banks. They concluded that 'diversification is not guaranteed to produce superior performance and/or greater safety for banks' (2006: 1355). Their explanation is that bank monitoring loses its effectiveness with diversification as a bank expands into areas with more competition or where it lack expertise. This can also be viewed as another example of the endogeneity of risk taking as managers use up their diversification gains by taking on more risk elsewhere, in this case by monitoring less effectively.

A general conclusion from these studies is that the growing reliance on non-interest income has not been associated with reduced volatility in earnings. Summarizing the literature, DeYoung and Rice (2004a) concluded that 'increased reliance on fee-based activities tends to increase rather than decrease the volatility of banks' earnings streams (p. 34)'. Roland and DeYoung (2001) offer three potential explanations. First, lending is typically relationship-based, so there are high switching costs for both borrowers and lenders. This tends to make the lending relationship sticky, and therefore more stable. Second, non-interest income is often associated with increased operating leverage—that is, high fixed costs relative to variable costs. As result, a given amount of revenue volatility is transformed into even more earnings volatility. Third, the activities that generate non-interest income do not typically have a substantial regulatory capital charge. This allows banks to operate with greater financial leverage, which can generate volatility. The increase in leverage is another example of the endogeneity of risk taking discussed earlier.

A second general conclusion is that expansion to new activities has two effects on volatility—a direct effect through the changing weights and an indirect effect through diversification. Both affect overall volatility and the evidence suggests that the recent expansion toward non-interest income has offsetting effects at best.

Geographic diversification

Laderman, Schmidt, and Zimmerman (1991) were among the first to examine the impact of geographic diversification in the 1980s when US states began materially to allow wider expansion. They found that the relaxation of statewide branching restrictions led rural banks to hold more non-agricultural loans and urban banks to hold more agricultural loans. While this is not a direct test of the benefits of

geographic diversification, it does show that banks responded to the changing constraints and moved to diversify their portfolios.

Rose (1996) examined US banks between 1980 and 1992 and concluded that geographic expansion generally led to higher risk, but that some diversification gains emerged when firms expanded into at least four distinct regions. Hughes, Mester, and Moon (1996) examined US banks in 1994 and searched for a link between geographic diversification and measures of insolvency risk measured as a Z-score and inefficiency. They found mixed results—more branches tended to lower insolvency risk for inefficient banks, but raise it for efficient banks, while operations across more states increased risk for efficient banks.

Pilloff and Rhoades (2000) concluded that geographically diversified banks do not have a net competitive advantage, while Morgan and Samolyk (2005) reported that a broader geographic scope would increase risk-adjusted returns. In particular, Morgan and Samolyk found a U-shaped relationship between geographic diversification and risk-adjusted returns, which implies that further broadening of the geographic footprint may be optimal.

Carlson and Mitchener (2006) studied the impact of geographic expansion in the US during the 1920s and 1930s. They emphasized that increased geographic scope through, for example, deregulation generally has two effects—increased ability to diversify and increased competitive pressures from potential entry. They concluded that the competitive effects were quantitatively more important in terms of reducing bank failures than the diversification effect. They found no evidence that diversification reduced bank failures for national banks.

Market studies

A third strand of research focused on equity market measures of risk and return. Relative to the studies that focused on accounting data, there are clear reasons to prefer this perspective. To the extent bank managers have choices in how economic activities are reported in an accounting sense, market data provide a clearer view on the risk impact. If accounting data are manipulated to generate a smoother revenue stream or if different revenue streams are subject to different accounting treatments—for example, the trading portfolio is marked-to-market on a daily basis which may induce volatility in non-interest income, then accounting returns may be misleading indicators of true risk. Second, market data provide a more forward-looking perspective on the expected returns of new activities, while accounting data are necessarily backward-looking and reflect actual performance in the past.

Product market diversification

Brewer (1989) examined the diversifying benefits of banks from 1978 to 1986 by comparing the equity market return volatility to a measure of non-bank activity

implied by the holding company's balance sheet. He found mixed results, although the risk-reducing benefits of non-bank activities seemed largest for high-risk banks. Rosen, et al. (1989) focused particularly on the real estate activities of banks from 1980 to 1985 and concluded that greater real estate investment would probably increase risk.

Templeton and Severiens (1992) examined financial market data for BHCs from 1979 to 1986 and measured diversification benefits as the share of market value not attributed to bank assets. They concluded that this measure was correlated with a lower variance of shareholder returns (total risk), but not with systematic risk, suggesting that diversification reduced only the idiosyncratic component.

Similar to the previously discussed studies of the impact of Section 20 subsidiaries on bank performance, several papers have examined the link between equity market measures of risk and return and the presence of Section 20 subsidiaries. Cornett, Ors, and Tehranian (2002) found evidence of gains from Section 20 subsidiaries as the industry-adjusted operating cash flow return on assets rises, while both total and systematic risk did not change significantly. They concluded that the improved cash flow associated with establishing a Section 20 subsidiary reflects increased revenue and decreased costs, rather than increased risk taking.

Geyfman (2005a) also examined the impact of having Section 20 subsidiaries for very large BHCs in the late 1980s and 1990s. She found that the presence of Section 20 subsidiaries was associated with lower idiosyncratic risk (evidence of diversification benefits), but higher systematic risk. Total risk tended to fall with Section 20 subsidiaries, however. Geyfman (2005b) utilized a portfolio approach and found that Section 20 subsidiaries provided strong diversification benefits. She concluded that US BHCs should reduce their commercial banking exposure and increase their securities underwriting exposure.

Other papers have examined the link between activity diversification again as measured by revenue streams and equity market returns. Stiroh (2006a) used a simple portfolio framework and found that activities that generate non-interest income do not raise average equity market returns, but are correlated with higher total risk, idiosyncratic risk, and systematic risk. These findings indicate that the higher weight on relatively volatile non-interest activities outweighs the diversification benefits, so overall volatility rises with a great non-interest exposure. Idiosyncratic risk, however, did fall with BHC assets, suggesting diversification benefits along other dimensions associated with size.

Stiroh (2006b) extended this analysis to control for variation within both the loan portfolio and the revenue stream. He found a negative link between total risk and diversification of both the loan portfolio and the sources of revenue, but that a greater reliance on non-interest income was linked to more volatile returns. Baele, De Jonghe, and Vander Vennet (2007) performed a similar exercise for European banks with supporting results; shifts of revenue into non-interest income were correlated with higher market betas and idiosyncratic risk fell with size.

Geographic diversification

Buch, Driscoll, and Ostergaard (2005) examined the investment choices of banks located in France, Germany, the UK, and the US from 1995 to 1999. Using returns to different country investments that are approximated by broad bond indices, they found that banks tended to over-invest domestically and there were considerable, unexploited gains from international diversification. As an explanation, they point to cultural, legal systems, and capital control frictions.

Finally, there is evidence from the merger literature. Delong (2001) used equity return data to gauge the reaction of investors to different types of bank mergers—that is, those that diversify versus those that specialize the bank. She found that diversifying mergers—by activity and/or by geography—do not create market value at the time of the merger announcement. In contrast to DeLong (2001), Laeven and Levine (2007), and the broader literature on the diversification discount, Elsas, Hackethal, and Holzhauser (2006) concluded that revenue diversification through both organic growth and through mergers and acquisitions leads to higher market values in a study of international banks from 1996 to 2003, which they attributed to revenue and cost economies of scope.

Mishra, et al. (2005) examined the impact of diversification for a small set of US bank mergers. They found no evidence that systematic risk changed after a merger, but significant evidence that idiosyncratic and total risk declined. They interpret this as evidence of diversification gains or a 'risk synergy benefit' from the combined entity.

Conclusions

US banks have clearly become more diversified over the last two decades as regulatory barriers fell, financial innovation progressed, and opportunities to expand into new products lines and new geographic areas opened. This diversification, however, has not provided an obvious advantage to large firms during the ongoing financial crisis as real-estate-related problems spread over a range of products and geographies.

The empirical evidence suggests that observers should not be too surprised as there is no consensus on the impact of diversification on bank risk in the US and around the world. In some sense, this is predictable. Risk taking is endogenous and optimizing managers may choose to exploit any diversification gains by increasing returns or adding risk in another dimension. Moreover, banks are shifting into precisely those activities that are relatively volatile, which can offset and obscure any diversification benefits.

Many papers, however, have found that risk-adjusted returns actually declined with the expansion of activities. This is harder to explain and requires some speculation. One possible explanation is that US bank managers may have simply got the diversification idea wrong. Managers and analysts, for example, have extolled the virtues of 'cross-selling' to lower costs, increase income, and add diversification. But, if banks are simply selling more products to the same core customers, then this might not be true diversification if business lines have simply become exposed to the same underlying shocks.

An alternative explanation may be the non-profit-maximizing motives discussed by Berger, Demsetz, and Strahan (1999), Milbourn, Boot, and Thakor (1999), Bliss and Rosen (2001), Houston, James, and Ryngaert (2001), and Aggarwal and Samwick (2003). These motives include managers' zeal for empire building, over-diversification to protect firm-specific human capital, corporate control problems, or managerial hubris and self-interest, all of which could lead to inefficient diversification.

Excess risk taking could also reflect a standard principal-agent explanation if traders, brokers, and underwriters (agents) like volatility more than shareholders (principals) do. Laeven and Levine (2007), for example, argued that the discount that the market applies to diversified financial firms is consistent with the idea of severe agency problems within financial institutions. These market failures would be exacerbated by any implicit government guarantee that reduces the incentives for debt holders to monitor and discipline managers. Many observers have raised this concern in the 'originate-to-distribute' and securitization model of mortgage finance that came to prominence in the 2000s.

Finally, the disappointing results could be a short-run phenomenon due to adjustment costs associated with the recent expansion or simply bad luck reflecting recent market conditions. Gramm-Leach-Bliley and the ability to offer full-scale financial services was passed less than a decade ago and the US economy experienced a series of financial market shocks over this period such as the Asian crisis and LTCM in 1998, the bursting of the NASDAQ bubble in 2000, the events of 9/11 2001, and corporate accounting scandals in 2002. If true, risk-adjusted performance could improve as the necessary business practices, expertise, technology, and scale are developed, and banks more successfully manage their expanded operations in a more stable environment. The performance of many larger, diversified financial firms during the current crisis, however, makes this explanation increasingly untenable.

These potential explanations are speculative, and it is critical better to understand the risk and stability of the largest financial firms. There is ample evidence that disruptions in the provision of credit can have real economic consequences (e.g., Ashcraft, 2005), so supervisors and regulators should have strong incentives to understand the motivations for and impact of increasing diversification by financial services firms.

Continuing change in financial markets, however, make this a considerable challenge. The shift toward an 'originate-to-distribute' model of credit and the increased reliance on complex securitization practices, for example, altered traditional lending practices and made them more integrated with capital markets. The long-run impact of the more recent failure of several large institutions and the fundamental restructuring of others remains unclear. This suggests that historical studies may not prove particularly insightful when assessing the potential for future diversification benefits among the largest financial institutions, but opens an exciting opportunity for continued research in this area.

References

Acharya, V. V., Hasan, I., and Saunders, A. (2006). 'Should Banks be Diversified: Evidence from Individual Bank Loan Portfolios', *Journal of Business*, 79: 1355–412.

Aggarwal, R. K. and Samwick, A. A. (2003). 'Why Do Managers Diversify Their Firms? Agency Reconsidered', *Journal of Finance*, 58: 71–118.

Allen, L. and Jagtiani, J. (2000). 'The Risk Effects of Combining Bank, Securities, and Insurance Activities', *Journal of Economics & Business*, 52: 485–97.

Ashcraft, A. B. (2005). 'Are Bank Failures Really Special? New Evidence from the FDIC-Induced Failure of Healthy Banks', *American Economic Review*, 95: December, 1712–1930.

Baele, L., De Jonghe O., and Vennet R. V. (2007). 'Does the Stock Market Value Bank Diversification?', *Journal of Banking & Finance*, 31: 1999–2023.

Barth, J. R., Brumbaugh, R. D. Jr., and Wilcox, J. A. (2000). 'The Repeal of Glass-Steagall and the Advent of Broad Banking', *Journal of Economic Perspectives*, 14: Spring, 191–204.

Berger, A. N., Demsetz, R. S., and Strahan, P. E. (1999). 'The Consolidation of the Financial Services Industry: Causes, Consequences, and Implications for the Future', *Journal of Banking & Finance*, 23: 135–94.

Bliss, R. T. and Rosen, R. J. (2001). 'CEO Compensation and Bank Mergers', *Journal of Financial Economics*, 61: 107–38.

Boyd, J. H. and Graham, S. L. (1988). 'The Profitability and Risk Effects of Allowing Bank Holding Companies to Merge with Other Financial Firms: A Simulation Study', *Quarterly Review*, Federal Reserve Bank of Minneapolis, 12: Spring, 3–20.

——— and Hewitt, R. S. (1993). 'Bank Holding Company Mergers with Nonbank Financial Firms: Effects on the Risk of Failure', *Journal of Banking & Finance*, 17: 43–63.

Brewer III, E., (1989). 'Relationships between Bank Holding Company risk and Nonbank Activity', *Journal of Economics & Business*, 41: 337–53.

Buch, C. Driscoll, J. C., and Ostergaard, C. (2005). 'Cross-Border Diversification in Bank Asset Portfolios', ECB Working Paper Series, 429.

Campa, J. M. and Kedia, S. (2002). 'Explaining the Diversification Discount', *Journal of Finance*, 57: 1731–62.

Carlson, M. (2004). 'Are Branch Banks Better Survivors: Evidence from the Depression Era', *Economic Inquiry*, 42: 111–26.

——— and Mitchener, K. J. (2006). 'Branch Banking, Bank Competition, and Financial Stability', *Journal of Money, Credit, and Banking*, 38: 1293–328.

Clark, T., Dick A., Hirtle B. J., Stiroh, K. J. and Williams, R. (2007). 'The Role of Retail Banking in the U.S. Banking Industry: Risk, Return, and Industry Structure', *Economic Policy Review*, Federal Reserve Bank of New York, 13: December, 39–56.

Cornett, M. M., Ors, E., and Tehranian, H. (2002). 'Bank Performance around the Introduction of a Section 20 Subsidiary', *The Journal of Finance*, 57: 501–21.

Cummins, J. D., Phillips, R. D., and Smith, S. D. (1998). 'The Rise of Risk Management', *Economic Review*, Federal Reserve Bank of Atlanta, First Quarter, 30–40.

DeLong, G. L. (2001). 'Stockholder Gains from Focusing versus Diversifying Bank Mergers', *Journal of Financial Economics*, 59: 221–52.

Demsetz, R. S. and Strahan, P. E. (1997). 'Diversification, Size, and Risk at Bank Holding Companies', *Journal of Money, Credit and Banking*, 29: 300–13.

Deng, S., Elyasiani, E., and Mao, C. X. (2007). 'Diversification and the Cost of Debt of Bank Holding Companies', *Journal of Banking & Finance*, 31: 2453–73.

DeYoung, R., Hunter, W. C., and Udell G. F. (2004). 'The Past, Present, and Probable Future of Community Banks', *Journal of Financial Services Research*, 25: 85–133.

—— and Rice, T. (2004a). 'How Do Banks Make Money? The Fallacies of Fee Income', *Economic Perspectives*, Federal Reserve Bank of Chicago, Q4: 34–51.

—— —— (2004b). 'How Do Banks Make Money? A Variety of Business Strategies', *Economic Perspectives*, Federal Reserve Bank of Chicago, Q4: 52–67.

—— and Roland, K. P. (2001). 'Product Mix and Earnings Volatility at Commercial Banks: Evidence from a Degree of Total Leverage Model', *Journal of Financial Intermediation*, 10: 54–84.

Elsas, R., Hackethal, A., Holzhauser, M. (2006). 'The Anatomy of Bank Diversification', Working Paper, SSRN #968214.

Emmons, W. R, Gilbert, R. A., and Yeager, T. J. (2004). 'Reducing the Risk at Small Community Banks: Is it Size or Geographic Diversification that Matters?', *Journal of Financial Services Research*, 25: 259–81.

Estrella, A. (2001). 'Mixing and Matching: Prospective Financial Sector Mergers and Market Valuation', *Journal of Banking & Finance*, 25: 2367–92.

Fama, E. F. and French, K. R. (2004). 'The Capital Asset Pricing Model: Theory and Evidence', *Journal of Economic Perspectives*, 18: 25–46.

Flannery, M. J. (1999). 'Comment on Milbourn, Boot, and Thakor', *Journal of Banking & Finance*, 23: 215–20.

Froot, K. A., Scharfstein, D. S., and Stein, J. C. (1993). 'Risk Management: Coordinating Corporate Investment and Financing Policies', *Journal of Finance*, 48: 1629–1658.

—— and Stein, J. C. (1998). 'Risk Management, Capital Budgeting, and Capital Structure Policy for Financial Institutions: An Integrated Approach', *Journal of Financial Economics*, 47: 55–82.

Furlong, F. (2000). 'The Gramm-Leach-Bliley Act and Financial Integration', *Federal Reserve Bank of San Francisco Economic Letter*, 10.

Geyfman, V. (2005a). 'Banks in the Securities Business: Market-Based Risk Implications of Section 20 Subsidiaries', Federal Reserve Bank of Philadelphia Working Paper 05–17.

—— (2005b). 'Risk-Adjusted Performance Measures at Bank Holding Companies with Section 20 Subsidiaries', Federal Reserve Bank of Philadelphia Working Paper 05–26.

Goddard, J., McKillop, D., and Wilson, J. O. S. (2008). 'The Diversification and Performance of U.S. Credit Unions', *Journal of Banking & Finance*, 32: 1836–49.

Haubrich, J. G. (1998). 'Bank Diversification: Laws and Fallacies of Large Numbers', *Economic Review*, Federal Reserve Bank of Cleveland, Q2: 2–9.

HAYDEN, E., PORATH, D., and WESTERNHAGEN, N. V. (2006). 'Does Diversification Improve the Performance of German Banks: Evidence from Individual Bank Loan Portfolios', Deutsche Bundesbank Working Paper, June.

HIRTLE, B. J. and STIROH, K. J. (2007). 'The Return to Retail and the Performance of U.S. Banks', *Journal of Banking & Finance*, 31: 1101–33.

HOUSTON, J. C., JAMES, C., and MARCUS, D. (1997). 'Capital Market Frictions and the Role of Internal Capital Markets in Banking', *Journal of Financial Economics*, 46: 135–64.

HUGHES, J. P. and MESTER, L. J. (2002). 'Bank Capitalization and Cost: Evidence of scale Economies in Risk Management and Signaling', *Review of Economics and Statistics*, 80: 314–325.

JORION, P. (2005). 'Bank Trading Risk and Systemic Risk', National Bureau of Economic Research Working Paper 11037, January.

KAMP, A., PFINGSTEN, A., MEMMEL, C., and BEHR, A. (2007). 'Diversification and the Banks' Risk–Return Characteristics: Evidence from Loan Portfolios of German Banks', SSRN #906448.

KLEIN, P. G. and SAIDENBERG, M. R. (2005). 'Organizational Structure and the Diversification Discount: Evidence from Commercial Banking', CORI Working Paper 2005–06, SSRN #721566.

KWAN, S. (1998). 'Risk and Return of Banks' Section 20 Securities Affiliates', Federal Reserve Bank of San Francisco *Economic Letter* 32.

KWAST, M. (1989). 'The Impact of Underwriting and Dealing on Bank Returns and Risk', *Journal of Banking & Finance*, 13: 101–25.

LADERMAN, E. S., SCHMIDT, R. H., and ZIMMERMAN, G. C. (1991). 'Location, Branching, and Bank Portfolio Diversification: The Case of Agricultural Lending', *Economic Review*, Federal Reserve Bank of San Francisco, Winter, 24–38.

LAEVEN, L. and LEVINE, R. (2007). 'Is There a Diversification Discount in Financial Conglomerates?', *Journal of Financial Economics*, 85: 331–67.

LANDSKRONER, Y., RUTHENBERG, D., and ZAKEN, D. (2005). 'Diversification and Performance in Banking: The Israeli Case', *Journal of Financial Services Research*, 27: 27–49.

LEPETIT, L., NYS, E., ROUS, P., and TARAZI, A. (2005). 'Product Diversification in the European Banking Industry: Risk and Loan Pricing Implications' <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=873490>.

——— ——— ——— ——— (2006). 'The Provision of Services, Interest Margins, and Loan Pricing in European Banking', <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=955852>.

MERCIECA, S., SCHAECK, K., and WOLFE, S. (2007). 'Small European Banks: Benefits from Diversification', *Journal of Banking & Finance*, 31: 1975–98.

MERTON, R. C. (1974). 'On the Pricing of Corporate Debt: The Risk Structure of Interest Rates', *Journal of Finance*, 29: May, 449–70.

MILBOURN, T. T., BOOT, A. W. A. and THAKOR, A. V. (1999). 'Megamergers and Expanded Scope: Theories of Bank Size and Activity Diversity', *Journal of Banking & Finance*, 23: 95–214.

MISHRA, S., PRAKASH, A. J., KARELS, G. V. and PETERSON, M. (2005). 'Bank Mergers and Components of Risk: An Evaluation', *Journal of Economics and Finance*, 29: 84–96.

MORGAN, D. P. and SAMOLYK, K. (2005). 'Bigger and Wider: The (Neglected) Benefits of Geographic Diversification in Banking', Working Paper, Federal Reserve Bank of New York, June.

PENAS, M. F. and UNAL, H. (2004). 'Gains in Bank Mergers: Evidence from the Bond Market', *Journal of Financial Economics*, 74: 149–79.

Pilloff, S. J. and Rhoades, S. A. (2000). 'Do Large, Diversified Banking Organizations Have Competitive Advantages?', *Review of Industrial Organization*, 16: 2873–302.

Reichert, A. K. and Wall, L. D. (2000). 'The Potential for Portfolio Diversification in Financial Services', *Economic Review*, Federal Reserve Bank of Atlanta, Third Quarter, 35–51.

Roll, R. (1988). 'R^2', *Journal of Finance*, 43: 541–66.

Rose, P. S. (1996). 'The Diversification and Cost Effects of Interstate Banking', *The Financial Review*, 31: 431–52.

Rosen, R. J., Lloyd-Davies, P. R., Kwast, M. L., Humphrey, D. B. (1989). 'A Portfolio Analysis of Bank Investment in Real Estate', *Journal of Banking & Finance*, 13: 355–66.

Santomero, A. W. and Chung E. (1992). 'Evidence in Support of Broader Banking Powers', *Financial Markets, Institutions & Instruments*, 1: 1–69.

Saunders, A. and Walter. I. (1994). *Universal Banking in the United States: What Could We Gain? What Could We Lose?*, New York, NY: Oxford University Press.

Slijkerman, J. F., Schoenmaker, D., and de Vries, C. G. (2005). 'Risk Diversification by European Financial Conglomerates', Tinbergen Institute Discussion Paper 110/2, December 7.

Slovin, M. B., Sushka, M. E., and Polonchek, J. A. (1993). 'The Value of Bank Durability: Borrowers as Bank Stakeholders', *Journal of Finance*, 48: 247–66.

Smith, R., Staikkouras, C., and Wood, G. (2003). 'Non-Interest Income and Total Income Stability', Bank of England Working Paper 198, August.

Sprague, O. M. W. (1903). 'Branch Banking in the United States', *Quarterly Journal of Economics*, 27: 242–60.

Stiroh, K. J. (2004a). 'Do Community Banks Benefit from Diversification?', *Journal of Financial Services Research*, 25: 135–60.

—— (2004b). 'Diversification in Banking: Is Noninterest Income the Answer?', *Journal of Money, Credit, and Banking*, 36: 853–82.

—— (2006a). 'A Portfolio view of Banking with Interest and Noninterest Activities', *Journal of Money, Credit, and Banking*, 38: 1351–61.

—— (2006b). 'New Evidence on the Determinants of Bank Risk', *Journal of Financial Services Research*, 30: 237–63.

—— and Rumble, A. (2006). 'The Dark Side of Diversification: The Case of U.S. Financial Holding Companies', *Journal of Banking & Finance*, 30: 2131–61.

Strahan, P. E. and Sufi A. (2000). 'The Gains from Financial Sector Modernization', Federal Reserve Bank of New York, mimeo, May.

Stulz, R. (1984). 'Optimal Hedging Policies', *Journal of Financial and Quantitative Analysis*, 19: 127–40.

Tarazi, A., LePetiti, L., Nys E., and Rous, P. (2008). 'Bank Income Structure and Risk: An Empirical Analysis of European Banks', *Journal of Banking & Finance*, 32: 1452–67.

Templeton, W. K. and Severiens, J. T. (1992). 'The Effect of Nonbank Diversification on Bank Holding Companies', *Quarterly Journal of Business and Economics*, 31: 3–16.

Valverdi, S. C. and Fernandez, F. R. (2007). 'The Determinants of Bank Margins in European Banking', *Journal of Banking & Finance*, 31: 2043–63.

Winton, A. (1999). 'Don't Put All Your Eggs in One Basket? Diversification and Specialization in Banking', SSRN #173615.

Yu, L. (2003). 'On the Wealth and Risk Effects of the Glass-Steagall Overhaul: Evidence from the Stock Market, Washington DC, American Finance Association meetings.

CHAPTER 7

UNIVERSAL BANKING

ALAN D. MORRISON[1]

Introduction

Universal banks are institutions that combine the lending and payment services of commercial banks with a wider range of financial services. In particular, universal banks underwrite securities, and hence can offer their client firms access to a broader range of sources of funds than can specialist commercial or investment banks. While universal banks dominate the financial sector in some economies, they are relatively uncommon in others. Indeed, they were outlawed in the US for the last two-thirds of the twentieth century. Understanding this international variation has been a preoccupation of economists since Schumpeter (1939) and Gerschenkron (1962); more recently, the academic literature on this topic has examined the rationale for the American decision to separate commercial from investment banking, and discussed the extent to which universal banks require special regulation. This chapter outlines Gerschenkron's ideas, and relates them to more recent discussions of the subject. It discusses the potential for conflicts of interest within universal banks, and assesses other policy debates surrounding universal banks. Finally, it concludes with a discussion of the reasons for the recent

[1] Many of the ideas concerning human capital in this chapter arose during joint work with Bill Wilhelm, to whom I am extremely grateful for numerous conversations and insights. I am also grateful to Alexander Gümbel and Dimitri Tsomocos for comments on an earlier draft.

expansion in the number and importance of universal banks, and it assesses some of the potential consequences of this expansion.

Universal banking, industrial development, and human capital formation

Numerous authors have pointed to the large-scale and well-capitalized universal banks that played an important role in the pre-World War I financing of German industry. These banks held the equity as well as debt securities of their clients; Schumpeter (1939) argues that this resulted in long-term relationship formation, which facilitated efficient resource direction. In contrast, the finance required for British industrialization was garnered partly from commercial banks, and also from stock market flotations, which were brought to market by small-scale merchant houses, which lacked the financial capital of the commercial banks.

In a famous essay, Gerschenkron (1962) explains these differences as consequences of the ways in which the respective economies developed. He argues with reference to 'economic backwardness'; this is a term that he never defines precisely, but he uses it to describe economies that were relatively late adopters of modern methods of production and distribution. He argues that development in these economies was hampered by a number of 'institutional obstacles'. First, entrepreneurs had not accumulated capital in the earlier stages of development. Second, the workforce in 'economically backward' economies had little experience of new technologies and ways of doing things. In the language of modern labor economics, they lacked human capital and, in particular, they lacked *tacit* human capital, acquired through on-the-job experience but not easily taught at arm's length, for example in a classroom.[2]

Gerschenkron's observation about human capital is central to his argument, although it has been little discussed by modern authors. Lack of labor skill renders development difficult; in some cases, as for example with the serfdom of peasants in pre-1861 Russia, he argues that it renders it impossible.[3] In general, Gerschenkron argues, industrialization in backward economies can proceed only when technology

[2] Becker (1964) discusses human capital; its importance to development has been stressed by many authors. Tacit skill is discussed by Polanyi (1966).

[3] One might expect serfdom to provide the owners of peasants with incentives to invest in human capital. Gerschenkron argues that serfdom was symptomatic of a social sclerosis that undermined any tendency toward innovation.

reaches a sufficiently advanced stage to *substitute* for human capital: it is easier to teach a worker to operate very advanced production machinery than it is to teach him the tacit production skills he needs to obtain any benefit from a less sophisticated machine.

Gerschenkron argues that nineteenth-century Germany, and also France and Russia, were 'economically backward'; he contrasts them with England, which industrialized earlier, and hence was not. According to Gerschenkron, industrialization in economically backward nations relied upon technologies that rendered hard-to-transmit tacit skills sufficiently unimportant. But these technologies operated at a very large scale. They could only function effectively if an adequate infrastructure developed to support them: factories required railways to ensure adequate levels of throughput, railways required coal, and so on.[4] Industrialization therefore had to proceed on a 'broad front', and this required capital on a scale that, by virtue of their economic backwardness, local entrepreneurs had not accumulated. Hence, appropriate institutions were required to pool capital, and to direct it toward the technologies that would underpin economic development. These institutions were the universal banks.

In sum, Gerschenkron argues that universal banks arose naturally in countries that had to play economic catch-up, because a lack of human capital generated a pressure toward bigness which could only be satisfied if dispersed sources of capital were combined. And, because capital came from dispersed sources, it had to be closely watched by the banks that directed it. Gerschenkron claims that the German banks established 'the closest possible relations with industrial enterprises', and that they 'accompanied an industrial enterprise from the cradle to the grave, from establishment to liquidation throughout all the vicissitudes of its existence'.

Gerschenkron's analysis extends to what modern writers call industrial organization, but which he calls the 'industrial structure' of the economy. He argues that concentration of power and relations in the universal banks served to reinforce the basic tendencies inherent in backward countries, so that attention was devoted to the heavy industry where large-scale financial capital was most useful. Moreover, Gerschenkron argues that the late-nineteenth-century cartelization movement in German industry was a natural result of the amalgamation of German banks which 'refused to tolerate fratricidal struggles amongst their children'. In contrast, while English banks also consolidated at this time, the process was not mirrored to the same extent in industry.

Gerschenkron does not discuss the US banking sector. America industrialized later than Britain; although it lacked some of the institutional features that Gerschenkron argues were a brake on development elsewhere, one might expect America's banking system to have developed along German lines. Indeed, while

[4] More recently, Chandler (1990) stresses the importance to the development of industrial capitalism of a large-scale infrastructure that can service expensive capital.

the US financial system owed rather more to the British system of financing than did the German one and, moreover, the extent of universal banking in the US was restricted by regulation, US banking had some features that, at least at the level of casual empiricism, were consistent with Gerschenkron's stories. At the end of the first decade of the twentieth century, Redlich (1968: 381–2) argues that not more than six banking firms were responsible for managing the organization of the American economy.[5] Lamoreaux (1985) documents the merger wave that the US experienced between 1895 and 1904: it saw 1,800 firms disappear into merged entities, and many of the firms formed at this time continued to dominate their industries for the following century: examples are US Steel, DuPont, International Harvester, Pittsburg Plate Glass, American Can, and American Smelting and Refinery. As discussed by Morrison and Wilhelm (2007: 182–4), these mergers were largely orchestrated by bankers, and in particular by J.P. Morgan and Co. In line with Gerschenkron's observations, Morgan was concerned throughout his career to avoid what he regarded as destructive competition between competitors, and this concern informed his deal-making at this time.

As far as I am aware, Gerschenkron's assertion that the rate of human capital formation affected the way in which the banking sector developed in Germany (and, arguably, in the US) has not been subjected to a formal empirical analysis. His assertion that universal banks formed closer and longer-term relationships than their counterparts elsewhere has, however, been discussed.

Calomiris (1993; 1995) examines the effect of laws that prevented US banks from consolidating and branching during the second Industrial Revolution (1870–1914). He argues that these laws increased the informational and transactions costs of issuing securities, and hence he argues that there was a lower propensity to issue equities in the US than in Germany where, he argues, universal banks were better able to extract valuable information from their borrowers. Moreover, he presents evidence that the costs of financing German industrialization were lower than in the US, precisely because there were universal banks in Germany. He argues that institutional changes that increased bank concentration in the US lowered the costs of finance there.

Calomiris' conclusions are challenged in a series of papers by Caroline Fohlin. If long-term relationships with banks eased financing conditions then one should see less contemporary evidence of credit rationing in firms with such a relationship. Fohlin (1998b) tests whether a relationship with one of the nine 'Great Banks' of nineteenth-century Germany eased access to credit by examining the cash flow sensitivity of investment for firms with and without such a relationship. Her

[5] They were J.P. Morgan and Co., First National and the National City Bank of New York, Kuhn, Loeb, and Co., and, to a lesser extent, Kidder, Peabody, and Co. and Lee, Higginson, and Co.

approach follows Fazzari, Hubbard, and Petersen (1988) and Hoshi, Kashyap, and Scharfstein (1990): firms that have easy access to capital should be less reliant upon retained earnings to finance investment, and hence, after controlling for the quality of their investment opportunities, their investment levels should be independent of the cash that their operations generate. Fohlin faces an endogeneity problem, in that association with a universal bank may be related to the quality of investment opportunities. But, even after controlling for this effect, she finds, in an apparent contradiction of the relationship hypothesis, that a bank relationship actually *increases* the sensitivity of investment to holdings of liquid assets. In another paper (Fohlin (1998a), she shows that universal bank affiliation in Italy did nothing to ameliorate the liquidity sensitivity of investment, and finds little support from performance data for the notion that universal banks provided screening services to investors.

In other papers, Fohlin presents evidence that first indicates that German banks held more liquid assets than British banks, and that, while they held a limited number of securities in their portfolios, this was often merely because they could not place new issues in their entirety (Fohlin, 2001), and second that bank affiliation in Germany was about securities issuance and stock market listings, rather than the monitoring of debt contracts and the provision of consultancy services (Fohlin, 1997).

Edwards and Ogilvie (1996) also examine the role of universal banks in German industrialization. In contrast to Gerschenkron's claims, they find that universal banking accounted for a relatively small proportion of the total assets of financial institutions in Germany before 1914. At this time, joint stock companies never accounted for more than 20 percent of the industrial capital stock; for at least 80 percent, then, the special skills of universal banks were not relevant. In most cases, internally generated funds were the most important source of finance for joint stock companies, and much of the rest came from non-universal financial intermediaries, such as savings and mortgage banks, and credit cooperatives.

The evidence to support a close monitoring interpretation of Gerschenkron's universal banking story therefore seems rather shaky. Interestingly, however, Ramirez (1995) finds evidence that supports it in the American context. He finds, in contrast to Fohlin's (1998b) German analysis, that a relationship with J.P. Morgan significantly reduced the cash flow sensitivity of investment for American firms. Whether this reflects active monitoring or skilled screening is harder to establish, but it does suggest that a universal banking relationship could ease access to the credit markets. Indeed, it was concerns that the wrong types of firms might be helped into the capital markets by their investment banks that led the American authorities to separate commercial from investment banking. Their reasoning, and the evidence concerning it, is examined in the next section.

Universal banking and conflicts of interest

Commercial banks had a significant presence in the US securities markets of 1900. Although the Comptroller of the Currency ruled in 1902 that national banks were not permitted to engage in the securities business, the First National Bank of Chicago managed in 1903 to circumvent this ruling, by creating a securities affiliate. Securities affiliates were state banks with their own capital, owned by the shareholders of the national bank in proportion with their shares in the national bank. As state banks were not the concern of the Comptroller, affiliates were able to operate in the securities markets, and consequently the national banks functioned as de facto investment banks.[6]

However, while commercial banks were able to operate via securities affiliates in the securities markets, their activities were viewed with some skepticism by populist regulators and legislators. A series of investigations into the governance of investment houses brought the state into conflict with the securities industry in the first quarter of the twentieth century:[7] the Armstrong Committee of 1905 expressed concerns regarding excessively close relations between large investment banks and insurance companies, and the Pujo committee of 1912 tried but failed to prove the existence of a 'money trust' that suppressed competition in finance. In the wake of the 1929 stock market crash, the investment banks were again in the line of fire, this time from the Pecora committee, established in 1932 by Herbert Hoover in an attempt to substantiate his belief that the stock market was being undermined by pools of short sellers.

Ferdinand Pecora was far from neutral: Morgan remarked at the time that 'Pecora has the manner and the manners of a prosecuting attorney who is trying to convict a horse thief'.[8] Nevertheless, he found some evidence of governance failures, most notably at National City Bank.[9] His findings fed a public mood that demanded changes to the regulatory framework of the investment banking industry, and which found its voice in New Deal legislation that both established a regulatory framework for the securities industry, by creating the Securities and Exchange Commission, and that also profoundly altered the industrial organization of the industry.

[6] Carosso (1970: 276) discusses at some length the operation of securities affiliates.

[7] Morrison and Wilhelm (2007: 196–215) discuss the hearings, and their consequences, in some detail.

[8] See Leuchtenburg (1963: 59).

[9] Charles E. Mitchell, the president and board chairman of National City Bank, was paid a salary of $25,000, but awarded himself bonuses of $1 million in 1927 and 1928. Seligman (1982) discusses the hearings, and their legislative consequences, in detail.

The Banking Act of June 1933, popularly known as the Glass-Steagall Act, abolished securities affiliates by requiring a total separation of investment from commercial banking. The Act had a massive impact, since at the end of the 1920s over half of all new securities issues were sponsored by security affiliates. In the wake of the Act, all issues had to be brought to market by specialist investment houses. J.P. Morgan and Co. remained in deposit banking and hence had to leave the securities industry.[10]

While some academic articles debate the point,[11] the Glass-Steagall Act appears to have been motivated by concerns that commercial banks were using their securities affiliates to place low-quality securities on the market in order to avoid taking losses on their own loan portfolios. For example, the Pecora commission uncovered evidence that when the National City Bank's securities affiliate, the National City Company, pushed Peruvian debt, it did so despite knowing that it was a poor investment. There is, however, evidence that commercial banks lost heavily on unsold stock when underwriting issues by their debtors (see Kroszner and Rajan, 1994).

The claim that securities affiliates pushed low-quality issues that benefited their parent firms at the expense of their investors went unchallenged in the academic literature for many years. But it is rather incredible: if securities affiliates were pushing low-quality issues then, if they were dealing with rational investors, the low quality should have been reflected in share prices. Hence, if the securities affiliates were pushing poor securities, either they were dealing with naive investors who failed to learn from experience, or they were making no profits from their actions. Neither story is particularly convincing. Moreover, the fact that investment banks faced conflicts of interest is not necessarily evidence of institutional failure: Morrison and Wilhelm (2007: chaps 2 and 3) argue that investment banks are economically useful precisely because, by placing their reputations at risk, they are able to manage conflicts of interest.

Conflicts of interest in pre-1933 investment banking were examined carefully in the 1990s, as pressure mounted for a repeal of the Glass-Steagall Act. Kroszner and Rajan (1994) test the 'naive investor' theory by examining the performance of affiliate-underwritten securities. They find that there were fewer defaults among affiliate-underwritten securities, which mitigates against the hypothesis that these securities were of systematically lower quality.

Kroszner and Rajan also point to evidence about the pattern of securities issuance that suggests strongly that investors were perfectly aware of the conflicts

[10] A year later, partners from Morgan and from Drexel founded the new firm of Morgan Stanley and Co. as an investment bank. See Carosso (1970) for a discussion of the industry changes that the Act caused.

[11] For example, Macey (1984) argues that the Act was intended to protect investment bankers at the expense of commercial bankers; Langevoort (1987) argues that Carter-Glass believed that his bill would encourage banks to channel money toward small companies, rather than into the securities markets.

that their investment banks faced. Precisely because they faced a potential conflict of interest, it was harder for securities affiliates credibly to signal the quality of their issues to the ratings agencies. Kroszner and Rajan support this assertion by showing that ratings were a less accurate predictor of default for affiliate-underwritten bonds than for those underwritten by specialist investment banks. I argue above that conflicts of interest have fewer adverse consequences within a bank that has significant reputational capital at stake. Hence, one would expect the informational problems to be particularly problematic for small affiliates with a lower reputational stake. A sophisticated investor should therefore be unwilling to buy complex and opaque securities that are underwritten by a small affiliate. Consistent with this argument, Kroszner and Rajan find first that affiliates in general underwrote larger issues where information asymmetry was less likely to be a problem, and second, that smaller affiliates underwrote more senior issues by less risky firms than did larger affiliates.

Kroszner and Rajan's results suggest strongly that investors were too smart to be taken in by an affiliate pushing poor-quality stock. Affiliates with less to lose could not underwrite informationally sensitive issues, and hence could not make as much from their securities business as competitors with more reputational capital. It is even possible that combining commercial with investment banking improved incentives, as commercial banks strove to build reputations which would allow them to enter the lucrative securities markets.

Ang and Richardson (1994) present evidence that is consistent with Kroszner and Rajan's. They find that bank affiliate issues had lower default rates, lower *ex ante* yields, and higher *ex post* prices than those issued by pure investment houses; moreover, they find that the relative ability of *ex ante* yields to predict *ex post* performance was no different for affiliate issues than for investment bank issues. Even issues underwritten by the National City Company and the Chase Securities Corporation, both of which were targets of the Pecora hearings, while of lesser quality than other bank affiliate issues, were no worse than those underwritten by the investment banks. Puri (1994) also presents evidence that pre-1933 bank underwritten issues defaulted less than non-bank underwritten issues.

In contrast to other papers written on this subject, Puri (1996) bases conclusions regarding the quality of affiliate issues on *ex ante* pricing, rather than on *ex post* default performance. She finds that pre-1933 investors paid higher prices for securities underwritten by banks than for those underwritten by securities houses. Puri argues that these results are indicative of a certification role for banks, which arose because banks had superior information about the firms to which they lent, and because they faced reputational risk.

In short, recent research suggests strongly that pre-1933 commercial banks in the US did not use their securities affiliates to float securities that would repay their lowest quality loans. The Glass-Steagall Act rendered it impossible to perform precisely this type of research on contemporary US firms. However, Gompers

and Lerner (1999) are able to come close, by examining the underpricing of initial public offerings (IPOs) brought to market between December 1972 and December 1992 by investment banks that held equity in the issuing firm via a venture capital subsidiary. Once again, they find no support for the 'naive investor' hypothesis; investors appear rationally to account for the quality of securities. IPOs underwritten by affiliated investment banks in their sample perform at least as well as those in which underwriters have no position. Investors demand a greater discount for investing in affiliated issues and, consistent with the evidence in Kroszner and Rajan (1994), investment bank-affiliated venture capital firms seem to invest in less information-sensitive issues.

Another opportunity to perform research on modern data was provided by a partial relaxation of the Glass-Steagall Act in 1987, under which some banks were allowed to set up subsidiaries ('Section 20 subsidiaries') to underwrite corporate securities. The subsidiaries were subject to firewalls that limited information flows, and they limited in size: initially to 5 percent of the gross revenues of the parent bank, and ultimately to 25 percent. Gande, et al. (1997) examine the operations of Section 20 subsidiaries. Their findings are in line with all of the research cited above, in that they find no evidence of malfeasance. They control in their work for the use to which the proceeds of the issue are put. When the securities are issued for purposes other than debt repayment, spreads for sub-investment grade issues are forty-two basis points lower than for investment houses; when the stated purpose is refinancing, the spreads are statistically indistinguishable. Moreover, and in contrast to some of the earlier papers cited, Gande, et al. find that Section 20 subsidiaries tend to underwrite smaller issuers than investment houses. The evidence of this work is therefore that, if anything, the informational advantage of lending banks serves to *attract* investors, rather than to repel them: Puri (1999) presents a model along these lines, in which the information that commercial banks acquire through lending allows them to obtain better prices for securities. Gande, Puri, and Saunders (1999) find, moreover, that the entry of Section 20 subsidiaries lowered fees for security underwriting, particularly among lower-rated and smaller issues, where Section 20 subsidiaries were particularly active. Evidence largely consistent with the results of these papers is presented by Roten and Mullineaux (2002), who find that Section 20 subsidiaries charged lower fees than investment bank underwriters, who were able to capitalize upon their stronger reputational capital, but that there was no significant overall difference in yield spreads between the two types of underwriters.

While a substantial body of evidence suggests first that the market accounts for conflicts of interest when US banks underwrite securities, and second that such conflicts are seldom a significant concern, there is little comparable evidence in other countries. However, a paper by Ber, Yafeh, and Yosha (2001) generates results for the modern Israeli market that are somewhat at variance with those for the US of the 1930s. The Israeli banking industry is highly concentrated, and it is universal,

with banks managing investment funds as well as controlling subsidiaries that specialize in underwriting. While most pre-Glass Steagall data is for bond issues, Ber, et. al. focus on straight equity issues. They find that the post-issue *accounting* performance of firms underwritten by their lender is significantly better than average. However, they find that the same firms exhibit negative stock excess returns in the first day and year after issuance, which suggests that these issues are systematically overpriced. If buyers are not naive, we must look elsewhere for an explanation for this persistent mis-pricing. The authors suggest that it arises because the buyers are investment funds controlled by issuers. Hence, they argue that, at least in the Israeli market, the combination of bank lending, underwriting, and investment fund management in a single institution is potentially harmful.

The findings of Ber, et al. are worrisome. They suggest that, while managed funds are controlled by entirely rational agents, they are able to find and to exploit naive retail investors; hence, financial infrastructure needs to be designed so as to ensure that fund managers' incentives are properly aligned with their investors. Arguably, then, the efficiency consequences of allowing universal banking in one economy could be different to those in another, which has different institutional and legal features. While research into the pre-Glass-Steagall US economy helped to justify the repeal of the Banking Act, it should be applied to other economies with caution. The next section examines the policy arguments that surrounded the introduction of universal banking into the US.

Universal banks and economic efficiency: The repeal of the Glass-Steagall Act

The November 1999 Gramm-Leach-Bliley Act dismantled the barriers to universal banking that had been erected in the US by the Glass-Steagall Act. The Gramm-Leach-Bliley Act responded to an increasing commercial need for universal banking that had already been recognized by the Federal Reserve Board's 1998 approval of the merger of Citicorp and Travelers. The same pressures were apparent in Europe where, in the absence of Glass-Steagall-type legislation, financial conglomeration had been taking place for at least a decade.[12] The most important pressure for the Gramm-Leach-Bliley Act was therefore commercial, but it did not pass without considerable discussion. In this section I briefly outline and comment upon some of the sources of debate.

[12] See Lown, et al. (2000) for a survey.

Universal banking could be introduced to the US only when several concerns had been assuaged. First, the conflicts of interest that had motivated the passage of the Glass-Steagall Act had to be addressed. Second, regulators and legislators had to be convinced that universal banking would not create new systemic risks that threatened the stability and efficiency of the financial sector.

As discussed in the previous section, a growing body of research in the 1990s suggested that the deleterious consequences of the conflicts of interest facing universal banks may have been more perceived than real. The systemic risks fell into several categories. First, there were concerns that the formation of large universal banks would be too central to the operation of the economy to be allowed to fail, as a result of which a moral hazard problem would exist between the shareholders and managers of these banks. The danger that some banks might be treated as 'too big to fail' was reflected in market prices after the Comptroller of the Currency acknowledged in testimony to Congress that eleven of the largest US national banks could expect to receive the sort of $1 billion bailout extended in 1984 to the insolvent Continental Illinois Bank: Avery, Belton, and Goldberg (1988) show that, subsequently, bank bond spreads were barely related to ratings, and Boyd and Gertler (1993) find that large banks took on bigger risks than smaller commercial banks.

When large banks are systemically important, could better regulation assuage the 'too-big-to-fail' (TBTF) problem? The Federal Deposit Insurance Corporation Improvement Act (FDICIA) of 1991 was landmark legislation partly intended to accomplish this goal. It requires regulators to take prompt corrective action against distressed banks, and places checks and balances upon the decision to declare a bank TBTF. Stern and Feldman (2004) argue that FDICIA did little to resolve the TBTF problem, claiming that regulators still have the incentive and the ability to bail out insolvent banks. Some evidence does indicate that FDICIA did not entirely resolve the TBTF problem: Morgan and Stiroh (2005) find that the spread-rating relationship for banks identified in the mid-1980s as TBTF was little changed by FDICIA—although they find more sensitivity than did Avery, Belton, and Goldberg (1988)—and Brewer and Jagtiani (2007) show that banks are prepared to pay a premium for acquisitions that will push them over perceived TBTF boundaries.

Nevertheless, Mishkin (2006) argues in an essay reviewing Stern and Feldman's book that the weight of evidence does not support their assertion: Ennis and Malek (2005) find no evidence in the wake of FDICIA of the excessive risk taking documented in large banks by Boyd and Gertler (1993), and Flannery and Soresco (1996) find stronger market discipline in the subordinated debt market for banks in the post-FDICIA period. Hence, even if large universal banks are systematically so important that the regulator cannot credibly commit to deny them access to the government safety net, a case can be made that the concomitant incentive problems can be counteracted by well-designed regulatory institutions. Mishkin (1999) makes this case, arguing that universal banking should be accompanied by greater

regulatory vigilance, coupled with some constructive ambiguity regarding bailout policy.

A further systemic cost of universal banking may arise if the securities arm of a universal bank is able to access the deposit insurance safety net provided to the commercial banking arm: this would be likely to result in risk shifting, as securities firms take excessive risks, for which, by virtue of the deposit insurance scheme, their depositors do not charge them. Furthermore, as Boyd, Chang, and Smith (1998) note, banks that hold equity stakes in their borrowers have strong incentives to take advantage of the deposit insurance safety net. Benston (1994) discusses this point. He argues that there is no evidence that universal banking is more risky than specialized banking. Cornett, Ors, and Tehranian (2002) support Benston's assertion, finding that bank riskiness around the introduction of a Section 20 subsidiary does not change. In any case, several authors suggested in the 1990s that the diversification that universal banking would bring would more than outweigh any risk shifting dangers. Mälkönen (2004) and Allen and Jagtiani (2000) both perform simulations using portfolios of commercial bank loans and insurance company investments, and show that combining the two generates inter-divisional diversification. This work is, however, subject to a Lucas-style critique: Freixas, Lóránth, and Morrison (2007) show that the non-bank divisions of financial conglomerates could take more risk in order to profit from the deposit insurance put option than they would have done as stand-alone firms. Whether or not the diversification effect outweighs the enhanced risk shifting incentive is context specific. With the appropriate capital adequacy policy, Freixas, Lóránth, and Morrison (2007) demonstrate that optimal regulation forces the deposit taking and non-deposit taking arms of the bank to maintain separate balance sheets: although this reduces diversification opportunities, it enhances market discipline sufficiently to compensate.

Rajan (1996) expresses concern that universal banks may use power derived from their informational monopoly to suppress competing institutions and markets. While good regulation can probably counteract this danger, he suggests that concentration of economic power in a few universal banks could act as a brake on economic progress in developing countries. As we saw in the first section above, a similar point was made by Gerschenkron, who identified a tendency within late nineteenth- and early twentieth-century universal banks to suppress competition in the real sector of the economy. Unlike Rajan, of course, Gerschenkron argued that universal banks aided development, and hence that the danger of anti-competitive behavior was worth accepting. In any case, Benston (1994) argues that modern universal banks serve such a broad constituency that they are unlikely to favor one interest group over another, and hence that they are less likely to be a source of damaging rent seeking than more specialized institutions.

Boot and Thakor (1997) identify another way in which universal banks may reduce the beneficial effects of competition. They argue that borrowers choose

between bank and market finance by weighing up the relative benefits of bank monitoring, which attenuates moral hazard, and more-informative price signals, which facilitate efficient resource allocation. Financial innovations that increase price informativeness result in a shift from bank to market finance. These innovations raise welfare, but their effect within a universal bank is to transfer revenues from one part of the business to another. Hence, Boot and Thakor argue that the incentive to innovate in a universal bank is lower than in an investment bank, which can hope to attract new customers by innovating.

A further concern is that, by combining depository institutions with other types of financial firms, universal banks may open new channels for financial contagion, so that instability outside the banking sector, for example in the insurance market, could be transmitted to banks via universal firms that encompass both sectors. The evidence on this point is mixed. In work that to some extent anticipated Kroszner and Rajan, White (1986) finds no evidence of greater instability in universal institutions at the start of the 1930s: while 26.3 percent of national banks failed between 1930 and 1933, only 6.5 percent of the sixty-three banks that had security affiliates in 1929 and 7.6 percent of the 145 banks with large-scale bond operations failed. Logit regressions on White's data confirm that the presence of a security affiliate reduced the probability of a bank failure. Colvin (2007) argues that the Netherlands experienced in the 1920s its only traditional banking crisis since 1600; he presents evidence that the relatively large difficulties that the Rotterdamsche Bankvereeniging experienced relative to its rival Amsterdamsche Bank were attributable to its universal status. Franke and Hudson (1984) find no evidence that universal banks were behind any of the major twentieth-century financial crises to affect West Germany in the twentieth century. Canals (1997) cites Cuervo (1988) on the effect of the European recessions of the late 1970s and early 1990s upon Spanish banks. In both cases, the banks that experienced the biggest losses were universal banks with major stakes in the industrial sector.

The aforementioned evidence suggests that the dangers associated with universal banking are less than was believed in the US for much of the twentieth century. Indeed, there is a body of evidence that indicates that universal banks are positively efficiency-enhancing. For example, Barth, Brumbaugh, and Wilcox (2000) point to technological advances that open new economies of scope in large banks. Berger, et al. (2000) discuss economies of scale: universal institutions can share offices, computers, information systems, investment departments, account service centers, or other operations; they can economize on the fixed costs of raising capital, and they can re-use information about a client in several business lines.[13] On the other hand, like any other organization, universal banks may experience diseconomies of scale (see Winton, 1999 for a model incorporating this effect): the extent to which universal banks *can* realize economies of scale and scope is of course an empirical

[13] For related discussions, see Milbourn, Boot, and Thakor (1999) and Dierick (2004).

question. The second Banking Co-ordination Directive of 1989 made universal banking the norm in the European Union by introducing a single banking license valid throughout the European Union, and limiting product-mix restrictions to those imposed by home regulators.[14] Hence, one would hope for evidence for or against scope efficiencies in the European market. However, the few studies that exist are rather inconclusive: Allen and Rai (1996) and Vander Vennet (1999) find only limited evidence of scope economies in European universal banks; Cyberto-Ottone and Murgia (2000) find evidence that scope-expanding mergers in European banking markets increase shareholder wealth.

Gorton and Schmid (2000) use data from 1975 to 1985 to examine the consequences of universal banking for the real economy in Germany. They account for control rights, voting restrictions, and the effects of co-determination.[15] They find that banks affect firm performance beyond the effect they would have as non-banks, and that the concentration of control rights in banks improves firm performance. A number of authors have suggested that introducing universal banking into other countries would bring benefits that mirrored the German experience. Indeed, in an analysis of sixty countries, Barth, Caprio, and Levine (1999) find that restricting securities activities reduces bank efficiency and raises the likelihood of a banking crisis. Their data contains no evidence that restricting financial firms assists financial development, or that it increases industrial competition. Nevertheless, Rajan (1996) argues that one should be careful of drawing strong conclusions from this type of work: universal banking exists within a broader institutional framework, and it need not follow that the benefits associated with the entire framework can be achieved simply by embracing only universal banking. For example, Rime and Stiroh (2003) find no evidence to suggest that any efficiency benefits are being derived from the trend toward universal banking in Switzerland.

Notwithstanding the institutional caveats expressed by Rajan, the arguments of this section provide only weak support for regulation that prevents universal banking. Financial markets appear rationally to discount conflicts of interest within universal banks. Hence, as Kanatas and Qi (1998) argue, borrowers will choose to deal with universal banks only if the costs of conflict are outweighed by the scope economies that the universal banks can realize. Only if universal banking generates an unpriced social cost is there a case for restricting it. Kanatas and Qi suggest that this cost might arise because conflicts of interest give rise to a soft budget constraint: they argue that, because borrowers from a universal bank anticipate that they will be bailed out via a stock issue in the event of poor performance, they choose lower-quality investments. However, the empirical

[14] See Berger, De Young, and Udell (2001) for a discussion of this directive, and of the consolidation of financial services in the European Union.

[15] Co-determination gives German workers a right of representation on the Board of all but the very smallest companies. Gorton and Schmid find that it worsens firm performance.

evidence of the second section above suggests that, in fact, *ex post* conflicts are relatively small. In advanced economies, one can arguably deal with other potential problems, such as anti-competitive behavior and abuse of the deposit insurance safety net, through careful regulation.

In light of the previous paragraph, the case for proscribing universal banking seems rather weak, and the decision to repeal the Glass-Steagall Act seems justifiable. But, although the academic case for repeal was strong, the Gramm-Leach-Bliley Act was also a response to intense commercial pressures. Several authors have suggested that these pressures reflected the enhanced benefits of scale and scope made possible by advances in information technology.[16] The following section discusses this point, and relates this argument to the early ideas of Gerschenkron, discussed above.

Scale and scope in twenty-first-century banking

The investment banking industry became increasingly reliant upon financial capital in the second half of the twentieth century. Morrison and Wilhelm (2008) report data for the US: on a consumer price index (CPI)-adjusted basis (1983 dollars), the combined capitalization of the top ten investment banking firms rose at an increasing rate from $821 million in 1955 to $2,314 million in 1970, $6,349 million in 1980, $31,262 million in 1990, and $194,171 million in 2000. Over the same period the industry became increasingly concentrated, with the capitalization of the eleventh to the twenty-fifth largest investment banks as a proportion of that of the top ten dropping from 80 percent to 10 percent. Moreover, it appears that the importance of financial capital significantly increased relative to human capital over this period: while the average number of employees in the largest five banks quadrupled between 1979 and 2000, the mean capitalization per employee in these banks increased by a factor of more than fifteen. I will argue in this section that the imperative for universal banking at the end of the twentieth century was created by the same economic forces that increased both concentration and capitalization in the investment banking industry.

Starting from its origins in the nineteenth-century Atlantic trade, investment banks provided services over which it was very hard to contract: while clients may be able to distinguish a well-priced IPO from a poorly priced one, good advice from bad, or a well-executed security transaction from a botched deal, making this

[16] See, e.g., Barth, Dan Brambaugh, and Yago (2001).

distinction stick in court is very hard. It is precisely for this reason that investment banks depended upon their reputations: because clients would pay a significant premium to a trustworthy bank, investment bankers would work hard to retain their reputations, so that a strong reputation could underpin agreements that were not enforceable under black-letter law. The need for a reputation created a substantial barrier to entry into the business, and, arguably, explained the very long-lived super-normal profits that the early investment bankers made.[17]

When investment bankers relied upon reputation to underpin tacit agreements with their counterparties, their business was inevitably based upon close relationships. Many of the skills that investment bankers needed were tacit: that is, they were best learned on the job, through a close mentoring relationship with a senior banker. Morrison and Wilhelm (2004) argue that partnership firms provide the strongest possible incentives to maintain these relationships, and hence the early investment banks were constituted as partnerships. While partnership status assisted in human capital formation, it limited the size and capitalization of investment banks (Morrison and Wilhelm, 2008).[18]

Starting in the early 1960s, a number of factors undermined the traditional structure of the investment banking firm. First, the advent of transistor-based mainframe computers in the early 1960s rendered cost-effective the overnight batch processing of the large-scale repetitive tasks associated with settlement. This type of processing was particularly valuable to 'retail' firms like Merrill Lynch, which performed high volumes of small-value transactions. Mainframe computing was extremely costly, but retail firms that failed to adopt it found it impossible to cope with a massive increase in trading volumes at the end of the decade: they ultimately failed, or were absorbed by larger institutions (see Morrison and Wilhelm, 2007: 235–8). The retail firms acquired the capital needed to acquire mainframe computers by floating in the early 1970s (see Morrison and Wilhelm, 2008).

Further advances in information technology were more applicable to investment banks that specialized in wholesale business and, ultimately, to universal banks. The cost of computing started to plummet in the late 1970s, as microcomputers found their way into banks, and allowed traders and relationship managers to interrogate databases and to perform complex pricing calculations in real time. For example, the ability rapidly to create spreadsheet-based financial models revolutionized the operation of the leveraged buyout (LBO) market, and made it far easier to price new offerings. At the same time, advances in financial economic

[17] De Long (1991) argues that the impossibility of matching J. P. Morgan's reputation gave the firm a strong competitive position in the nineteenth century. Morrison and Wilhelm (2007: chaps 4–8) trace the evolution of the modern investment bank.

[18] The reason for this is twofold: first, partnership capital is provided by the partners, who have limited resources; and second, the number of partners is limited by a free-rider problem among partners.

theory were transforming the financial market place. The Black-Scholes-Merton framework for financial options valuation became a practical tool rather than an academic exercise when it could be implemented with a desktop computer; risk management practices could be hard-coded into computers, rather than based upon judgment and recruitment practices; trading and hedging strategies could be driven by computer algorithms rather than by humans.[19]

Unlike mainframe computers, microcomputers were cheap, and they substituted for a great deal of human expertise. One might expect them to *lower* the minimum scale at which investment banks could operate. Indeed, Rajan (1996) makes this point, arguing that there is no a priori reason to assume that better information technology should increase the optimal scale of a bank. But better information technology not only automated tasks that previously were the province of the human expert; it also changed the nature of investment banking skill. Activities that could be expressed in the formal language of financial economics could be taught in a classroom. Trading results that could be captured with computers and analyzed using portfolio theory could be contracted upon. As a result, businesses that previously were the preserve of a few specialists operating in businesses with reserves of human capital and reputation started to be open to any firm that could hire a smart financial engineer. Precisely because information technology and the codification of skills combined to render market entry easy for any firm, large or small, financial markets became extremely competitive. In the end, precisely because the financial markets had become so contestable by small firms, they could no longer sustain small-scale trade: bid-ask spreads narrowed to such an extent that participation in the markets became cost-effective only for firms that could operate at a large scale. The consequence was the massive increase in investment bank capitalization and concentration that I highlighted in the opening paragraph of this section.[20]

In short, distributed microcomputers and advances in financial economics lowered the value of tacit skill relative to technical, codifiable skill in many investment banking activities. It also facilitated entry, and hence lowered the minimum scale at which these activities were economically viable. Morrison and Wilhelm (2008) argue that these effects combined to cause the demise of the traditional investment banking partnership. They also opened the door to commercial bank entry into investment banking. Commercial banks had greater reserves of capital than the investment banks. Where they were legally allowed to underwrite, they could bundle their services with lending business in a way that investment banks could not. Particularly when underwriting bond issues, whose prices are most susceptible to codification, commercial banks therefore had

[19] Morrison and Wilhelm (2007: 238–49) discuss the phenomena outlined in this paragraph.

[20] This argument is given in greater detail, and with more supporting statistics, in Morrison and Wilhelm, 2008.

advantages that were denied to investment banks. Similarly, commercial banks were playing to their strengths when they invested in derivatives trading partnerships in the late 1980s: derivatives trading was a technical, computer-oriented activity that required capital on a huge scale.[21]

Gerschenkron argued in the 1960s that 'economically backward' economies relied for development upon codified knowledge that was embedded in large-scale and capital-intensive production technologies; it was for this reason that he believed that universal banking was common in economies that historically had developed from a backward state. The arguments of this section suggest that something similar is afoot in the modern banking sector, where production techniques have been revolutionized by new computer-based technologies that formalize many formerly tacit skills. As in Gerschenkron's work, the new technologies require very high levels of capital investment, which arise in this case because they generate competitive pressures that significantly raise the minimum operating scale in banking. The commercial pressures for universal banking seem unsurprising in the light of this argument; the steady erosion of the Glass-Steagall Act, starting in 1986 with the Fed's approval of an application by Banker's Trust to underwrite commercial paper and culminating in the passage of the Gramm-Leach-Bliley Act, was perhaps inevitable.

The immense scale and scope of the modern universal bank does not come without challenges, however. It may be very hard for an institution to run large-scale codified businesses side by side with those that rely upon more traditional tacit skills. When universal banks build systems and procedures around 'hard' codifiable information that can be fed into a computer, their decision-making becomes increasingly remote from the loan officers who forge relationships with their customers. As a result it becomes hard for them to accommodate lending based upon 'soft' relationship-based information that cannot easily be computerized. Stein (2002) argues that, as a result, loan officers in banks that rely upon formal systems to make decisions may be less inclined to gather information at all. Berger, et al. (2005) find evidence consistent with Stein's hypothesis, stating that 'large banks are less willing to lend to informationally "difficult" credits, such as firms with no financial records'. Of course, whether or not information is hard is to some extent a decision variable: Petersen (2004) argues that ratings agencies emerged in the nineteenth century as ways of hardening previously soft information about borrowers. But there are presumably limits to this process, and it may prove difficult in general to reconcile small-scale relationship lending with the needs of the universal bank.

[21] The most prominent derivatives trading partnerships were O'Connor, CRT, and Cooper Neff, which were acquired by Swiss Bank, Nations Bank, and BNP, respectively: see Morrison and Wilhelm (2007: 279).

The crisis of 2008

The causes of the financial crisis of 2008 are still a matter for discussion, and its consequences are still unfolding. It is too early to draw categorical conclusions from the financial crisis about banking in general, and about universal banking in particular. Nevertheless, it is possible to make some tentative observations.

First, and notwithstanding the remarks of section four above, it has become clear that, even in the most complex derivatives markets, tacit knowledge and non-codifiable information still have important roles in finance. The crisis appears to have started in the market for securitized subprime mortgage debt,[22] and its ramifications have been keenly felt in the so-called 'shadow banking sector', which relies upon complex debt securitizations. The mathematical models upon which these markets rested have been shown to be rather less precise than was previously believed. The consequence has been a loss in confidence in ratings agencies, and in the most complex securities that they rated: this loss of confidence is having profound effects upon real economic activitity.

The loss of confidence highlighted in the previous paragraph has highlighted the importance in complex markets of reputation. Many users of credit ratings relied upon them to assess the quality of assets that they did not fully understand. Similarly, purchasers of complex securities underwritten by blue-chip investment banks relied to some extent upon the reputation of their underwriter for quality certification. I have argued in this essay that reputation is most effectively fostered in small, focused institutions, where conflict-of-interest problems are least likely to impair incentives to maintain reputation. If reputation and the non-codifiable skills upon which it rests are more important that we thought at the start of 2008, then the challenges that universal banks face in providing hard-to-quantify services are greater than appeared in the last decade.

Universal banks have come in for more direct criticism, too. Some commentators have suggested that they are able to extend the reach of the deposit insurance fund; others have argued that universal banks create systemic problems because their businesses are so complex that no one, least of all their regulators, can understand them. The latter point is closely related to the reputational one, of course: reputation should substitute for transparency but, as argued in the previous paragraph, it appears not to have done in recent years.

If universal bank scale does indeed present a problem to service provision, what is the appropriate response? Some commentators have suggested that the right approach would be to reintroduce something like the Glass-Steagall Act.[23]

[22] See Brunnermeier (2009).

[23] For example, both Will Hutton and Jon Moulton argued in evidence to the (UK) House of Commons Treasury Select Committee on 13 January 2009 that the repeal of Glass-Steagall fanned the

This, they argue, would improve investment banker incentives by taking away any access to the deposit insurance safety net, and would ensure that they could fail. At the same time, it would reduce bank complexity, and so enhance systemic stability.

It is too early to say whether bank scope restrictions will be enacted. But such restrictions would come at a cost. The economies of scale and scope that have been discussed in this article remain important. Universal banks achieve diversification that ought to increase financial stability. And, finally, if reputation is an important basis for profitable business, one would expect participants in a free market to evolve institutions and procedures for maintaining it. This process may already have started: Citigroup announced at the start of 2009 that it planned to divest itself of its consumer finance operations, private-label credit card businesses, and parts of its investment banking business—thus partially reversing the move into universal banking that began with its 1998 acquisition of Travelers.[24] Nevertheless, a properly focused universal bank can generate significant economies of scale and scope. The universal banking model seems highly unlikely to vanish.

Conclusion

Historically, universal banking was common in some economies, but not in others. Gerschenkron (1962) argued that this variation could be explained with reference to the way in which development occurred: economies that had to play 'catch-up' did so by adopting technologies on a broad front that could compensate for the lack of a deep human capital pool, and the institution that collected and directed capital into these technologies was the universal bank.

Notwithstanding the success of universal banking, it was regarded with suspicion in the US for much of the twentieth century, where it was outlawed by the 1933 Banking Act. Contemporary evidence suggests that this suspicion was largely misplaced. Moreover, commercial pressures in the final decades of the twentieth century were for large, complex financial intermediaries that offered services that encompassed security market business as well as traditional commercial banking. I have argued that these pressures contained an echo of the forces studied by Gerschenkron: simultaneous advances in information technology and financial

flames of the credit boom and, ultimately, that it was a contributing factor to the credit crunch. Geoffrey Wood made a similar point in evidence to the House of Lords Economic Affairs Committee on 20 January 2009, stating that complexity allows for rapid bank failure, which in turn creates systemic problems.

[24] See Greg Farell, 'Universal Model Dies as Bank Goes Back to Basics,' *Financial Times* (14 January 2009).

economies codified traditional knowledge and created massive pressure for scale. These pressures resulted ultimately in the 1999 repeal of the Banking Act.

Universal banking creates challenges for some traditional commercial banking activities. As banks adopt hierarchical structures and rely increasingly upon hard quantifiable data, it will be harder for them to act upon the tacit knowledge and skills of their relationship managers. If large universal banks struggle to supply credit to small opaque businesses then a long-term role for small, specialist commercial lenders will remain.

References

Allen, L. and J. Jagtiani (2000). 'The Risk Effects of Combining Banking, Securities, and Insurance Activities'. *Journal of Economics & Business*, 52: 485–97.

—— and A. Rai (1996). 'Operational Efficiency in Banking: An International Comparison'. *Journal of Banking & Finance*, 20: 655–772.

Ang, J. S. and T. Richardson (1994). 'The Underwriting Experience of Commercial Bank Affiliates Prior to the Glass-Steagall Act: A Re-Examination of Evidence of Passage of the Act'. *Journal of Banking & Finance*, 18: 351–95.

Avery, R., T. Belton, and M. Goldberg (1988). 'Market Discipline in Regulating Bank Risk: New Evidence from the Capital Markets'. *Journal of Money, Credit, and Banking*, 3: 597–619.

Barth, J. R., G. Caprio, Jr., and R. Levine (1999). 'Banking Systems around the Globe: Do Regulation and Ownership Affect Performance and Stability?' Policy Research Working Paper 2325, World Bank, Washington, DC.

—— R. D. Brumbaugh, Jr., and J. A. Wilcox (2000). 'The Repeal of Glass-Steagall and the Advent of Broad Banking'. *Journal of Economic Perspectives*, 14: 191–204.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago: University of Chicago Press.

Benston, G. J. (1994). 'Universal Banking'. *Journal of Economic Perspectives*, 8: 121–43.

Ber, H., Y. Yafeh, and O. Yosha (2001). 'Conflict of Interest in Universal Banking: Bank Lending, Stock Underwriting, and Fund Management'. *Journal of Monetary Economics*, 47: 189–218.

Berger, A. N., N. H. Miller, M. A. Petersen, R. G. Rajan, and J. C. Stein (2005). 'Does Function Follow Organizational Form? Evidence From the Lending Practices of Large and Small Banks'. *Journal of Financial Economics*, 76: 237–69.

—— R. De Young, H. Genay, and G. F. Udell (2000). 'Globalization of Financial Institutions: Evidence from Cross-Border Banking Performance'. *Brookings-Wharton Papers on Financial Services*, 3: 23–125.

—— —— and G. F. Udell (2001). 'Efficiency Barriers to the Consolidation of the European Financial Services Industry'. *European Financial Management*, 7: 117–30.

Black, F. and M. Scholes (1973). 'The Pricing of Options and Corporate Liabilities'. *Journal of Political Economy*, 81: 637–54.

Boot, A. W. A. and A. V. Thakor (1997). 'Banking Scope and Financial Innovation'. *Review of Financial Studies*, 10: 1099–131.

Boyd, J. H., C. Chang, and B. D. Smith (1998). 'Moral Hazard Under Commercial and Universal Banking'. *Journal of Money, Credit and Banking*, 30: 426–68.

—— and M. Gertler (1993). 'U.S Commercial Banking: Trends, Cycles, and Policy', in O. J. Blanchard and S. Fisher (eds.), *National Bureau of Economic Research Macroeconomics Annual 1993*, Cambridge, MA: MIT Press, 319–68.

Brewer III, E. and J. Jagtiani (2007). 'How Much Would Banks be Willing to Pay to Become "Too-Big-to-Fail" and to Capture Other Benefits?', Research Working Paper 07-05, Federal Reserve Bank of Kansas City.

Brunnemeier, M. (2009). 'Deciphering the Liquidity and Credit Crunch of 2007–08'. *Journal of Economic Perspectives*, 23: 77–100.

Calomiris, C. W. (1993). 'Corporate Finance Benefits from Universal Banking: Germany and the United States, 1870–1914'. National Bureau of Economic Research Working Paper 4408, Cambridge, MA.

—— (1995). 'Universal Banking and the Financing of Industrial Development'. Policy Research Working Paper 1533, World Bank, Washington, DC.

Canals, J. (1997). *Universal Banking: International Comparisons and Theoretical Perspectives*. New York, NY: Oxford University Press.

Carosso, V. P. (1970). *Investment Banking in America: A History*. Cambridge, MA: Harvard University Press.

Chandler, A. D. (1990). *Scale and Scope; The Dynamics of Industrial Capitalism*. Cambridge, MA: Harvard University Press.

Colvin, C. L. (2007). 'Universal Banking Failure? An Analysis of the Contrasting Responses of the Amsterdamsche Bank and the Rotterdamsche Bank to the Dutch Financial Crisis of the 1920s'. Economic History Working Paper 98, London School of Economics, London.

Cornett, M. M., E. Ors, and H. Tehranian (2002). 'Bank Performance around the Introduction of a Section 20 Subsidiary'. *Journal of Finance*, 57: 501–21.

Cuervo, Á. (1988). *La Crisis Bancaria en España, 1977–1985*, Barcelona: Ariel.

Cyberto-Ottone, A. and M. Murgia (2000). 'Mergers and Shareholder Wealth in European Banking'. *Journal of Banking & Finance*, 24: 831–59.

De Long, J. B. (1991). 'Did J.P. Morgan's Men Add Value? An Economist's Perspective on Financial Capitalism', in P. Temin (ed.), *Inside the Business Enterprise: Historical Perspectives on the Use of Information*, Chicago, IL: University of Chicago Press, 205–36.

Dierick, F. (2004). 'The Supervision of Mixed Financial Services Groups in Europe'. Occasional Paper 20, European Central Bank, Frankfurt.

Edwards, J. and S. Ogilvie (1996). 'Universal Banks and German Industrialization: A Reappraisal'. *Economic History Review*, 49: 427–46.

Fazzari, S. M., R. G. Hubbard, and B. C. Petersen (1988). 'Financing Constraints and Corporate Investment'. *Brookings Papers on Economic Activity*, 141–95.

Flannery, M. J. and S. M. Sorescu (1996). 'Evidence of Bank Market Discipline in Subordinated Debenture Yields', *Journal of Finance*, 51: 1347–77.

Fohlin, C. (1997). 'Universal Banking Networks in Pre-War Germany: New Evidence From Company Financial Data'. *Research in Economics*, 51: 201–25.

—— (1998a). '*Fiduciari* and Firm Liquidity Constraints: The Italian Experience with German-Style Universal Banking'. *Explorations in Economic History*, 35: 83–107.

—— (1998b). 'Relationship Banking, Liquidity, and Investment in the German Industrialization'. *Journal of Finance*, 53: 1737–58.

——— (2001). 'The Balancing Act of German Universal Banks and English Deposit Banks, 1880–1913'. *Business History*, 43: 1–24.

Franke, H.-H. and M. Hudson (1984). *Banking and Finance in West Germany*, New York, NY: St. Martin's Press.

Freixas, X., G. Lóránth, and A. D. Morrison (2007). 'Regulating Financial Conglomerates'. *Journal of Financial Intermediation*, 16: 479–514.

Gande, A., M. Puri, and A. Saunders (1999). 'Bank Entry, Competition, and the Market for Corporate Securities Underwriting: A Survey of the Evidence'. *Journal of Financial Economics*, 54: 165–95.

——— ——— ——— and I. Walter (1997). 'Bank Underwriting of Debt Securities: Modern Evidence'. *Review of Financial Studies*, 10: 1175–202.

Gerschenkron, A. (1962). *Economic Backwardness in Historical Perspective*, Cambridge, MA: Harvard University Press.

Gompers, P. and J. Lerner (1999). 'Conflict of Interest in the Issuance of Public Securities: Evidence from Venture Capital'. *Journal of Law & Economics*, 42: 1–28.

Gorton, G. and F. A. Schmid (2000). 'Universal Banking and the Performance of German Firms'. *Journal of Financial Economics*, 58: 29–80.

Hoshi, T., A. Kashyap, and D. Scharfstein (1990). 'Corporate Structure, Liquidity and Investment: Evidence from Japanese Industrial Groups'. *Quarterly Journal of Economics*, 106: 33–60.

Kanatas, G. and J. Qi (1998). 'Underwriting by Commercial Banks: Incentive Conflicts, Scope Economies, and Project Quality'. *Journal of Money, Credit and Banking*, 30: 119–33.

Kroszner, R. S. and R. G. Rajan (1994). 'Is the Glass-Steagall Act Justified? A Study of the U.S. Experience with Universal Banking Before 1933'. *American Economic Review*, 84: 810–32.

Lamoreaux, N. R. (1985). *The Great Merger Movement in American Business, 1895–1904*. Cambridge, UK: Cambridge University Press.

Langevoort, D. C. (1987). 'Statutory Obsolescence and the Judicial Process: The Revisionist Role of the Courts in Federal Banking Regulation'. *Michigan Law Review*, 85: 672–733.

Leuchtenburg, W. E. (1963). *Franklin D. Roosevelt and the New Deal, 1932–1940*. New York, NY: Harper and Row.

Lown, C. S., C. L. Osler, P. E. Strahan, and A. Sufi (2000). 'The Changing Landscape of the Financial Services Industry: What Lies Ahead?'. *Federal Reserve Bank of New York Economic Policy Review*, 6: 39–55.

Macey, J. R. (1984). 'Special Interest Groups Legislation and the Judicial Function: The Dilemma of Glass-Steagall'. *Emory Law Journal*, 33: 1–40.

Mälkönen, V. (2004). 'Capital Adequacy Regulations and Financial Conglomerates'. Discussion Paper 10, Bank of Finland, Helsinki.

Merton, R. C. (1973). 'Theory of Rational Option Pricing'. *Bell Journal of Economics and Management Science*, 4: 141–83.

Milbourn, T. T., A. W. A. Boot, and A. V. Thakor (1999). 'Megamergers and Expanded Scope: Theories of Bank Size and Activity Diversity'. *Journal of Banking & Finance*, 23: 195–214.

Mishkin, F. S. (1999). 'Financial Consolidation: Dangers and Opportunities'. *Journal of Banking & Finance*, 23: 675–91.

——— (2006). 'How Big a Problem is Too Big to Fail? A Review of Gary Stern and Ron Feldman's *Too Big to Fail: The Hazards of Bank Bailouts*'. *Journal of Economic Literature*, 91: 988–1004.

Morgan, D. and K. J. Stiroh (2005). 'Too Big to Fail After All These Years'. Staff Report 220, Federal Reserve Bank of New York.

Morrison, A. D. and W. J. Wilhelm, Jr. (2004). 'Partnership Firms, Reputation and Human Capital'. *American Economic Review*, 94: 1682–92.

—— —— (2007). *Investment Banking: Institutions, Politics and Law*, Oxford: Oxford University Press.

—— —— (2008). 'The Demise of Investment Banking Partnerships: Theory and Evidence'. *Journal of Finance*, 63: 311–50.

Petersen, M. A. (2004). 'Information: Hard and Soft'. Working paper, Kellogg School of Management, Northwestern University, IL.

Polanyi, M. (1966). *The Tacit Dimension*. Garden City, NY: Doubleday.

Puri, M. (1994). 'The Long-Term Default Performance of Bank Underwritten Security Issues'. *Journal of Banking & Finance*, 18: 397–418.

—— (1996). 'Commercial Banks in Investment Banking. Conflict of Interest or Certification Role?' *Journal of Financial Economics*, 40: 373–401.

—— (1999). Commercial Banks as Underwriters: Implications for the Going Public Process. *Journal of Financial Economics*, 54: 133–63.

Rajan, R. G. (1996). 'The Entry of Commercial Banks Into the Securities Business: A Selective Survey of Theories and Evidence', in A. Saunders and I. Walter (eds.), *Universal Banking: Financial System Design Reconsidered*, Chicago: Richard D. Irwin, 282–302.

Ramirez, C. D. (1995). 'Did J.P. Morgan's Men Add Liquidity? Corporate Investment, Cash Flow and Financial Structure at the Turn of the Twentieth Century'. *Journal of Finance*, 50: 661–78.

Redlich, F. (1968). *The Molding of American Banking: Men and Ideas*. New York, NY: Johnson Reprint Corporation.

Rime, B. and K. J. Stiroh (2003). 'The Performance of Universal Banks: Evidence from Switzerland'. *Journal of Banking & Finance*, 27: 2121–50.

Roten, I. C. and D. J. Mullineaux (2002). 'Debt Underwriting by Commercial Bank-Affiliated Firms and Investment Banks: More Evidence'. *Journal of Banking & Finance*, 26: 689–718.

Schumpeter, J. A. (1939). *Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process*. New York: McGraw-Hill.

Seligman, J. (1982). *The Transformation of Wall Street: A History of the Securities and Exchange Commission and Modern Corporate Finance*. Boston, MA: Houghton Mifflin Company.

Stein, J. C. (2002). 'Information Production and Capital Allocation: Decentralized versus Hierarchical Firm'. *Journal of Finance*, 57: 1891–21.

Stern, G. H. and R. J. Feldman (2004). *Too Big To Fail: The Hazards of Bank Bailouts*. Washington, DC: Brookings Institution Press.

Vander Vennet, R. (1999). 'The Effect of Mergers and Acquisitions on the Efficiency and Profitability of EC Credit Institutions'. *Journal of Banking & Finance*, 20: 1531–58.

White, E. N. (1986). 'Before the Glass-Steagall Act: An Analysis of the Investment Banking Activities of National Banks'. *Explorations in Economic History*, 23: 33–55.

Winton, A. (1999). 'Don't Put All Your Eggs in One Basket? Diversification and Specialization in Lending'. Working Paper, Carlson School of Management, University of Minnesota, MN.

CHAPTER 8

THE CORPORATE STRUCTURE OF INTERNATIONAL FINANCIAL CONGLOMERATES

COMPLEXITY AND ITS IMPLICATIONS FOR SAFETY AND SOUNDNESS

RICHARD HERRING
JACOPO CARMASSI[1]

Introduction

International financial conglomerates have become an increasingly important feature of the financial landscape. Universal banking countries have long integrated the securities business with traditional commercial banking, but over the last

[1] The authors are grateful to the editors of this volume and Robert Eisenbeis for helpful comments on an earlier draft. Jacopo Carmassi would like to acknowledge that the opinions expressed are only those of the author and do not necessarily coincide with those of the Institution he is affiliated with.

decade the US and Japan, which formerly required strict separation of commercial banking from the securities business, have permitted banks to combine these two activities subject to some limitations. Increasingly combinations of banking and securities business have expanded to include insurance operations as well. Allianz in Germany, ING and Fortis in the Netherlands, Credit Suisse in Switzerland, and Citi in the US have all made important cross-sector acquisitions in recent years to combine banking and insurance activities (although both Credit Suisse and Citi have subsequently divested some of their insurance acquisitions). Indeed, virtually all of the large, international financial institutions are to some extent financial conglomerates combining at least two of the three formerly distinct functions of banks, securities firms, or insurance companies.

This consolidation and conglomeration appears to be motivated by hopes for cost savings and revenue enhancements from large, lumpy expenditures on information technology (Group of Ten, 2001). Economies of scope in production[2] may be important whenever a significant fixed cost can be shared across several different products. In addition to investments in information technology, several other kinds of fixed costs may be important—the costs of distribution channels, managing a client relationship, or establishing and maintaining a sound reputation and brand image. But diseconomies of scope may also be important (Herring and Santomero, 1990). In any event it is difficult to find evidence of significant economies of scope in the data. Indeed, Laeven and Levine (2007) find evidence of a diversification discount applied to financial conglomerates.

The trend toward consolidation and conglomeration may also be motivated by the hope of achieving greater market power. By controlling the full range of substitutes for a financial product, a financial conglomerate may be able to raise prices above marginal costs. In order to sustain such market power, the financial conglomerate would also need to be able to limit entry and enforce mandatory joint product sales. Of course, antitrust policy is intended to prevent such abuses. Moreover, intensified cross-border competition and technological advances that render all major markets for financial products highly contestable make it unlikely that any financial conglomerate could sustain market power should antitrust policy prove ineffectual.[3]

More than thirty countries have restructured and unified their regulatory and supervisory systems to deal with financial conglomerates in a more integrated

[2] Economies of scope in consumption may also be important. But they could be exploited by using the distribution network of one institution to sell packages of financial services produced by other firms and thus cannot explain the formation of institutions such as LCFIs that produce and distribute several different kinds of financial services (Herring and Santomero, 1990).

[3] See Berger, Demsetz, and Strahan (1999) for a review of the literature on consolidation and market power. They note that market power is most likely to be of concern with regard to in-market rather than market extension mergers and that retail customers are more likely to be adversely affected than wholesale customers. Berger (1995) makes a careful distinction between the market power and efficient structure hypotheses taking account of both X-efficiency and scale efficiency. He concludes that neither the market power nor efficient scale hypotheses 'are of great importance in explaining bank profits'.

fashion (Herring and Carmassi, 2008). Many of these international financial conglomerates have achieved a scale of operation and centrality in the functioning of the international financial system that render them systemically important. Traditionally, systemic concerns have been the preoccupation of bank regulators, but these concerns do not diminish when a bank becomes part of a group that includes insurance and securities activities as well. Although it is possible that larger, more diversified international financial conglomerates will be less likely to fail, if a failure should occur the spillover effects on the rest of the financial system are bound to be greater. Moreover, the heavy involvement of these firms in trading activities around the clock, around the globe means that the authorities would have very little time to react if one should experience extreme financial distress.

Our central premise is that the complexity of the corporate structures that most international financial conglomerates have developed is itself a significant source of systemic risk. In the event of bankruptcy, hundreds of legal entities would need to be resolved. Since most of these firms are managed in an integrated fashion along lines of business with only minimal regard for legal entities, national borders, or functional regulatory domains, and with substantial and complex intragroup relationships, simply mapping an institution's business activities into its legal entities presents a formidable challenge. Moreover, these legal entities would be subject to scores of different national regulatory and bankruptcy procedures, many of which conflict.

The corporate complexity of international financial conglomerates is likely to impede timely regulatory intervention and disposition. This exacerbates the moral hazard implicit in the financial safety net and diminishes market discipline on some of the most systemically important institutions, while at the same time constraining the ability of the supervisory authorities to substitute regulatory discipline for market discipline. In effect, several of these institutions may have become 'too complex to fail'.

We will begin with a consideration of the corporate structure that international financial conglomerates might prefer in the absence of regulatory and tax distortions. Then we will examine some of the (largely unintended) consequences for corporate structure of tax and regulatory policies. We will conclude with an analysis of some of the challenges this corporate complexity poses to an orderly winding down of an international financial conglomerate. But first we present an overview of the large complex financial institutions that we use to illustrate several aspects of the problem.

Large complex financial institutions

The regulatory authorities have identified sixteen financial conglomerates as large complex financial institutions (LCFIs) that are of crucial importance to the

functioning of the international financial system.[4] LCFIs 'include the world's largest banks, securities houses and other financial intermediaries that carry out a diverse and complex range of activities in major financial centers' (Bank of England, 2007b: 29). These firms are key intermediators of risk through their market-making activities and principal risk taking, as well as their provision of liquidity to capital markets. The concept was given empirical content in the Bank of England's *Financial Stability Review* (Bank of England, 2001) and since that time both the *Financial Stability Review* and the International Monetary Fund's *Global Financial Stability Report* have tracked developments among this group of financial institutions. The Bank of England (2007a: 7) has expressed concern about the rising systemic importance of LCFIs: 'Given their scale and their pivotal position in most markets, distress at an LCFI could have a large, unanticipated, impact on other financial markets participants. This could arise from losses on direct exposures to an LCFI that failed or from the wider market implications of actions taken by an LCFI to manage problems.'

Like the Holy Roman Empire, which was not holy, nor Roman, nor an empire, the term 'large and complex financial institutions' is imprecise. It does not include some of the largest financial institutions, nor some of the institutions that pursue the most diverse lines of business. Criteria for inclusion in the group require that an institution achieve a position as one of the ten largest participants in two or more of the following activities: book runners of international bond issues, book runners of international equity issues, book runners of global syndicated loans, notional interest rate derivatives outstanding, foreign exchange revenue, or worldwide assets under custody (Hawkesby, Marsch, and Stevens, 2003). LCFIs are completely dominant in some of these activities. For example, just two LCFIs act as custodians for around three-quarters of all assets in value terms (Bank of England, 2007a: 30) and three LCFIs are the dominant intermediaries in the market for credit derivatives (Bank of England, 2007a: 35).

Table 8.1 displays the sixteen institutions that are currently classified as LCFIs by the Bank of England (Bank of England, 2007: 29) and the IMF. At year end 2006, all of these institutions (except Lehman Brothers) ranked among the world's twenty-five largest banking groups in terms of total assets. Although these institutions differ with regard to the diversity of their activities (see column 6, the HHI for revenues for individual lines of business) and the extent of their international engagement (see column 4, the percentage of foreign subsidiaries, and column 5, the percentage of net foreign income), they are all major participants in international capital markets. LCFIs have had a greater than 70 percent market share as

[4] The term LCFI was introduced by a task force of the Financial Stability Forum, the G10 Ministers and Governors and the Basel Committee formed in 2000 to review the issues likely to arise in winding down an LCFI (Hüpkes, 2005). The Group of Ten Report (2001) on consolidation also considered a number of problems that might arise as consequence of the growth of large and complex financial organizations.

Table 8.1. Overview of Large Complex Financial Institutions

1 LCFIs	2 Total assets (billions of $, year end 2006)[1]	3 Total subsidiaries[1]	4 % of foreign subsidiaries	5 % of net foreign income before taxes (2006)[2]	6 HHI—business lines revenues (2006)[3]	7 Number of countries[4]	8 Subsidiaries in OFCs, number[5]	9 Subsidiaries in OFCs, %[5]
UBS AG	1,964	417	96%	62%	2,903	41	38	9%
Barclays Plc	1,957	1,003	43%	44%	2,179	73	145	14%
BNP Paribas	1,897	1,170	61%	51%	1,843	58	62	5%
Citi	1,884	2,435	50%	44%	4,122	84	309	13%
HSBC Holdings Plc	1,861	1,234	61%	78%	3,945	47	161	13%
The Royal Bank of Scotland Group Plc	1,711	1,161	11%	34%	1,966	16	73	6%
Deutsche Bank AG	1,483	1,954	77%	80%	3,931	56	391	20%
Bank of America Corporation	1,460	1,407	28%	12%	4,256	29	118	8%
JPMorgan Chase & Co.	1,352	804	51%	26%	2,086	36	54	7%
ABN AMRO Holding NV*	1,300	670	63%	77%	1,381	43	37	6%
Société Générale	1,260	844	56%	46%	4,128	60	64	8%
Morgan Stanley	1,121	1,052	47%	42%	4,476	46	203	19%
Credit Suisse Group	1,029	290	93%	71%	3,868	31	53	18%
Merrill Lynch & Co., Inc.	841	267	64%	35%	4,089	25	23	9%
Goldman Sachs Group, Inc.	838	371	51%	48%	5,391	21	29	8%
Lehman Brothers Holdings Inc.	504	433	45%	37%	7,807	20	41	9%

Note: Year end 2007 (unless otherwise specified).

*After the most recent list of LCFIs (Bank of England, 2007b) was published, a consortium of three banks (RBS, Fortis, and Santander) acquired ABN AMRO.

[1] Bankscope. Data on subsidiaries refer to majority-owned subsidiaries for which the LFCI is the ultimate owner with a minimum control path of 50.01%.

[2] Annual reports for each LCFI. Net income before taxes with five exceptions: net income after taxes for Citi, and net revenues for Barclays Plc, BNP Paribas, Lehman Brothers Holdings Inc., Merrill Lynch & Co., Inc.

[3] Oliver Wyman. The Herfindahl-Hirschman Index ranges from 0 to 10,000 and it is calculated on the percentage of revenues per business line. Higher values indicate a higher degree of specialization. Lower values imply a higher degree of diversification.

[4] Number of countries in which the LCFI has at least one majority-owned subsidiary.

[5] Offshore Financial Centers identified by the Financial Stability Forum (2000). We exclude Swiss subsidiaries for Credit Suisse and UBS and Hong Kong subsidiaries for HSBC. Four subsidiaries were allocated to OFCs on the basis of locations designated in their names even though Bankscope did not specify a home country.

lead arrangers and book runners of issues of residential mortgage-backed securities, leveraged syndicated loans, corporate debt, and asset-backed securities during the first three quarters of 2007 (Bank of England, 2007b: 38). LCFIs have experienced remarkable growth since the turn of the century, with total assets more than doubling in size from 2000 to 2006. In 2006, trading assets constituted more than one-third of the total (Bank of England, 2007a: 9).

LCFIs have also developed a remarkable degree of corporate complexity. In what follows we focus on the number of majority-owned subsidiaries as an indicator of corporate complexity. Of course, this is a somewhat arbitrary measure. The Federal Reserve Board, for example, takes a more expansive view of control in bank holding companies, establishing a 25 percent ownership level as the threshold. Moreover, it is a regrettably superficial measure of corporate complexity. Unfortunately, the Bankscope[5] data do not permit us to identify shell corporations or other inconsequential subsidiaries. Although it would be useful to supplement this simple quantitative measure with an indication of each entity's importance in the overall financial group, cross-guarantees, and role in the overall business structure, such information is not publicly available for many subsidiaries. Nonetheless, the number of majority-owned subsidiaries is an indication of the magnitude of the legal challenge that would confront the authorities in taking an LCFI through bankruptcy. All of the LCFIs have several hundred subsidiaries. Eight have more than 1,000 subsidiaries and one (Citi) has nearly 2,500 subsidiaries.

In the absence of tax and regulatory constraints, how much corporate complexity would LCFIs choose to adopt? The formation of subsidiaries can be costly. In addition to the start-up costs of obtaining a charter and creating a governance structure, there are ongoing costs for accounting, financial reporting, and tax filings. Nonetheless, LCFIs have adopted a considerable amount of corporate complexity even within some countries where they are under no regulatory obligation to do so. Germany, for example, has followed a universal banking model that permits banking and securities activities to be conducted within a single legal entity. Only investment funds, building societies, and insurance companies require the establishment of a separate legal entity. Nonetheless, Deutsche Bank, the leading German bank, has over 300 fully owned domestic subsidiaries (Bankscope, October 2007). What are the perceived, compensating benefits that justify the formation of corporate subsidiaries?

In the frictionless world of Modigliani and Miller (1958), a firm's choice of capital structure and, by extension, its corporate structure, cannot affect its value. But financial institutions lack any rationale in such a world. As Berger, et al. (1995: 394) note, most '[R]esearch on financial institutions has begun with a set of assumed imperfections', which includes asymmetric information and

[5] Bankscope is a global database containing information on public and private banks <http://www.bvdep.com/en/bankscope.html>.

transactions costs, costs of financial distress, taxes, and regulation. Each of these imperfections may influence a financial institution's choice of corporate structure.

Asymmetric information and transactions costs

Asymmetric information problems appear to afflict financial institutions more seriously than many other kinds of firms. Morgan (2002) presents evidence that financial institutions are inherently more opaque than other firms based on disagreements among bond rating agencies. Because many financial institutions specialize in lending to opaque borrowers and their trading positions can be easily and almost instantaneously changed, they are hard to monitor. Morgan finds that insurance companies may be even more opaque than banks since their primary assets are privately placed, long-term loans and the indemnity risks they underwrite may be even more uncertain to outsiders than bank liabilities.

Asymmetric information problems arise when one party to a transaction or relationship has information that the other does not, and it is too costly to write, monitor, and enforce a contract that would compensate adequately for the imbalance in information. When the objectives of the parties conflict, firms incur agency costs because of concerns about adverse selection—the fear the better-informed party will take advantage of the less-informed party by misrepresenting the quality of the product or service—or moral hazard—the fear that, once the transaction takes place, one party will covertly shift risk to the other's disadvantage. Financial firms have devised many different ways of mitigating these costs, including, sometimes, the creation of separate subsidiaries. Asymmetric information exacerbates conflicts of interest, which may arise between shareholders and creditors, between shareholders and managers, and between the firm and its customers. We will consider each in turn.

Asymmetric information: shareholders vs. creditors

The fundamental conflict of interest between shareholders and creditors springs from differences in their payoff functions. After debt-servicing costs have been paid, shareholders reap all the upside returns. They participate in the downside losses, however, only to the extent of their equity stake. In contrast, the upside return of creditors is limited to the promised return, while they may lose all that they have lent. Creditors will, thus, generally prefer safer investments than shareholders. With asymmetric information, creditors will be concerned that shareholders may

engage in risk shifting after the terms of a loan have been set by substituting riskier assets for the safer assets. To safeguard against this possibility creditors may charge a higher premium and attempt to constrain the firm in a number of ways, perhaps even refusing to lend. Kahn and Winton (2004) have shown that the choice of a corporate structure can ease this problem. By forming a risky subsidiary, the firm provides a commitment that limits its incentive to engage in risk shifting. Placing safer assets in a separate subsidiary increases the safe subsidiary's net returns in bad states of the world and reduces its incentives to engage in risk shifting. It may also improve terms on which the safe subsidiary can obtain external financing. Although the firm may still have an incentive to engage in risk shifting in the riskier subsidiary, Kahn and Winton (2004) argue that this limits the amount of risk shifting that can take place within the conglomerate. (For an opposing view, see Merton and Perold, 1993).

In support of their theory, Kahn and Winton (2004) note the tendency of commercial banks to form separate subsidiaries for their finance companies—for insurance companies to form separate subsidiaries for riskier policy lines, and for investment banks to form separate subsidiaries for their riskier private equity investments. Their theory also provides a rationale for good-bank/bad-bank restructurings such as the regulatory restructuring of the Continental Illinois National Bank and Mellon Bank's creation of Grant Street Bank in 1988. In fact, Kahn and Winton (2004: 2532) emphasize that several of the commonly advanced rationales for 'bad' bank structures are not convincing unless the implications for incentives to engage in asset shifting are taken into account.

Asymmetric information: shareholders vs. managers and internal agency problems

International financial conglomerates generally have broadly dispersed shareholders with no one dominant owner. This separation of ownership from managerial control means that shareholders face an asymmetric information problem *vis-à-vis* the managers of a firm. This is a classic principal–agent problem in which managers may be tempted to pursue their own objectives, such as empire building or the enjoyment of lavish corporate perquisites, rather than serving the interests of shareholders. This may lead to several different kinds of resource misallocations that diminish share values. Managers may be excessively risk-averse and seek to protect their entrenched positions by underinvesting in risky, positive net present value projects (Smith and Stulz, 1985). Or managers may take advantage of free cash flows to overinvest in value-destroying, negative net present value projects (Jensen, 1986). More broadly, managers may shirk.

Senior managers face similar issues with regard to managers lower down the corporate hierarchy. These internal agency costs include managerial entrenchment,

misallocations of resources, and rent-seeking behavior (Fulghieri and Hodrick, 2005). Although a number of corporate governance mechanisms deal with these problems, the choice of organizational form can also be used as an instrument to control the behavior of multiple agents and better align the incentives of owners and managers.

If a particular line of business has compensation practices or a culture that is very different from other lines of business in the conglomerate, segregation of that line of business into a separate entity may facilitate oversight and control. For example, it has often proven difficult to manage traders or deal-oriented investment bankers within the same compensation structure as relationship-oriented commercial bankers. Corporate separateness provides greater flexibility to better align incentives with the interests of shareholders and tailor employment contracts to prevailing market standards without destroying the culture necessary to make the business segment successful. As Aron (1991: 506) observes, the normal practice of tying the compensation of the manager of a business unit to the overall stock value of the group may not provide efficient incentives: 'When a division is part of a multiproduct corporation, the stock value of the firm is a noisy signal of the market's evaluation of any one divisional manager's productivity. Loosely speaking, the more noise there is in the signal, the costlier it is to properly motivate the manager.'

Despite massive investments in management information systems, integrated financial conglomerates may find it difficult to track and evaluate the performance of individual lines of business. Informal, internal capital markets sometimes contribute to the blurring of performance and result in unintended cross subsidies (Rajan, et al., 2000).[6] A degree of corporate separateness may be introduced to sharpen strategic focus and improve monitoring. For example, some groups have established separate units to handle client transaction processing with the intention of clarifying the performance of other risk taking units and giving senior managers better control over costs, pricing, product design, and delivery of transactions services.[7] This organizational innovation also facilitates benchmarking the transaction processing business against publicly traded, stand-alone businesses that provide similar services.

Occasionally a firm may take the additional step of partially spinning-off a subsidiary so that it has a separate listing and can be publicly traded. As Habib, Johnsen, and Naik (1997) observe, this enlists the help of capital markets in generating information that should improve the quality of investment decision.

[6] Holod and Peek (2006), however, provide evidence that internal capital markets in multi-bank holding companies enhance the efficiency of capital allocation. In particular, internal secondary loan markets avoid the asymmetric information problems faced by participants in the external secondary loan market and thus mitigate financial constraints faced by individual subsidiaries.

[7] For a description of the formation of PROFITCO at Bankers Trust, which was the first bank to restructure its processing services in this way, see Guil (2008).

It may also reduce the uncertainty of uninformed investors regarding the value of the subsidiary. Both effects should increase the value of the firm.

Firms may achieve some of the incentive benefits by simply forming a separate entity even though the spin-off never actually occurs. Aron (1991: 505) notes that 'The possibility of a future spinoff induces the divisional manager to act as if he were being monitored and evaluated by the capital market, even though the capital market's evaluation is observed only if a spinoff actually occurs'.

Information asymmetry: Customer concerns about conflicts of interest

Conflicts of interest are ubiquitous even in specialized financial institutions, but, as Walter (2003: 21) notes, '[A] matrix approach to mapping conflicts of interest demonstrates that the broader the range of clients and products, the more numerous are the potential conflicts of interest and the more difficult is the task of keeping them under control—and avoiding even larger franchise losses'. Customers fear that a firm may use its informational advantage to their detriment. Firms invest substantial resources to reassure clients and potential customers that they will not be disadvantaged *vis-à-vis* the firm or other clients. Such efforts include the erection of 'Chinese walls' restricting the flow of information across lines of business, the adoption of codes of conduct reinforced with compliance audits, and disclosures of potential conflicts (for a detailed study on conflicts of interest in the financial industry, see Walter, 2004).

Sometimes firms take the additional step of segregating activities in separate subsidiaries. For example, investment advisory services may be provided by a separate entity from underwriter and broker/dealer. Or, management consulting services may be offered through a separate entity in a separate location from the parent to reassure customers that confidential information would not be used in lending decisions or to aid other firms in which the parent might have an ownership position. Equally, corporate separateness may provide greater flexibility for operating units that would otherwise be constrained by conflict-of-interest concerns or burdensome reporting requirements. For example, Cox and Curry (2007: C12) reported that Goldman Sachs moved some of its proprietary trading desks from its investment bank into a separate, asset-management-unit. They speculate that one of the advantages may be that '[T]he stock-arbitrage desk may find it has more freedom to invest in companies involved in mergers or acquisitions that were once off limits because of the investment bank's activities as the world's top M&A adviser'.

Krozner and Rajan (1997) found evidence of this behavior in the way in which US banks organized their investment banking operations before the 1933 Glass-Steagall Act forced a separation between commercial and investment banking. During this period, some banks organized their investment banking operations as an internal department within the bank, while others formed separately incorporated affiliates with separate boards of directors. They found that the market attached a higher risk premium to issues underwritten by internal departments. Krozner and Rajan (1997: 475) conclude that this is consistent with 'investors discounting for the greater likelihood of conflicts of interest when lending and underwriting are within the same structure' and conclude that a separate affiliate structure is 'an effective commitment mechanism' to reassure customers that the underwriter will not abuse its information advantage.

Costs of financial distress: Protecting the group from a risky subsidiary

Financial distress occurs when a financial institution is expected to have difficulty in honoring its commitments. Costs of financial distress include not only costs of bankruptcy, but also the loss in value that may occur as a result of the perception that bankruptcy may be imminent even though it may ultimately be avoided. Talented employees may leave, suppliers may demand payment on delivery, revenues from credit-risk-sensitive products may decline, and conflicts of interest between shareholders and creditors may degrade the quality of operating, investment, and financial decisions. As Berger, et al. (1995: 396) note, 'Financial distress should be distinguished from economic distress. The cost of financial distress may be measured as the additional loss from economic distress for a leveraged bank versus an identical bank that is unleveraged. When asset quality deteriorates, both banks will experience economic distress, but the leveraged bank experiences a greater loss of value'.

When costs of financial distress are substantial, firms may prefer to segregate risky activities in separately incorporated subsidiaries even though information is shared equally between corporate insiders and capital markets. A holding company structure, in which subsidiaries are separately funded, can limit the damage to the rest of the group from financial distress in one of its affiliates. Corporate separateness provides the option of partial liquidation when losses in one of the subsidiaries would otherwise jeopardize the solvency of the rest of the group.

Bianco and Nicodano (2002) show that both shareholders of the financial group and the rest of society are better off when external debt is raised through separately incorporated subsidiaries rather than through the holding company and then downstreamed to the subsidiaries. In either case, gains from co-insurance could be realized: the holding company may choose to rescue a faltering subsidiary with profits from the rest of the group. But, if funding is primarily from the holding company, a group-threatening loss that hits a subsidiary will certainly inflict the costs of financial distress on the rest of the group. In contrast, if subsidiaries are separately funded in external capital markets, the loss could be stopped at the subsidiary directly affected, reducing the costs of financial distress to the rest of the group. Of course, the providers of debt will charge a higher risk premium when they lend to the subsidiary. But, as long as the premium does not include a substantial, adverse-selection premium, both shareholders and society should be better off. (Of course, this depends crucially on the authors' assumption of full information. If lenders are concerned that they are less-well-informed about risk, then the Kahn and Winton model discussed above is more relevant.)

It is sometimes asserted that a financial group could not afford to walk away from a faltering subsidiary because it would undermine confidence in the rest of the group. For example, Baxter and Sommer (2005: 187) argue that 'it is unlikely that limited liability is a strong argument for complex affiliate structures . . . [I]f limited liability aids an entity within the group, it is only at the expense of other entities in the group'. And Walter Wriston (1981), former Chairman of the predecessor of Citi, testified before Congress that '[I]t is inconceivable that any major bank would walk away from any subsidiary of its holding company'. While it is true that a loss of reputation may be more costly to financial firms than to other, less leveraged firms, limited liability has option value. In some instances, banks have walked away from insolvent subsidiaries without notable detrimental impact on the rest of their business. For example, ING cut loose a failing insurance subsidiary in London without substantial repercussions (Herring and Schuermann, 2005) and Bank of Nova Scotia and Crédit Agricole abandoned insolvent subsidiaries in Argentina (Dermine, 2006).

Moreover, banks sometimes appear to isolate riskier activities in separate subsidiaries. Dermine (2006) and Cerutti, Dell'Ariccia, and Martinez-Peria (2005), for example, have observed that banks tend to prefer to organize as subsidiaries (rather than branches) in riskier countries. Herring and Santomero (1990) reported that some banks chose to join clearing and settlement schemes that had open-ended loss-sharing agreements with separately capitalized subsidiaries in order to limit potential losses. The panic that swept through Asian securities markets after the collapse of Barings stemmed, in part, from the fear that a number of institutions would abandon their subsidiaries if losses should exceed their capital investments in memberships in some of the exchanges (Herring, 2003). But, in other cases—for

example in dealing with troubled SIVs, financial institutions have provided additional funds to protect their reputations even though they were under no legal obligation to do so.

In some jurisdictions, moreover, the limited liability option is constrained by regulation. The Federal Reserve Board has long held that the failure of a parent bank holding company to act as a source of strength to a troubled banking subsidiary would be considered 'an unsafe and unsound banking practice' (Ashcraft, 2004). The source-of-strength doctrine is intended to enhance the position of the bank within a holding company. It implies that during periods of financial stress, the regulatory authorities should be permitted to use the resources of the holding company and its subsidiaries to support the bank. In essence, the source-of-strength doctrine would give the regulatory authorities an option on the assets of the rest of the holding company to prevent the default of the bank. Nonetheless, the Fed's attempt to enforce this doctrine in the Mcorp case was thwarted by the courts and the Federal Deposit Insurance Corporation settled two cases where the parent of a failed bank sued the receivership to recover funds and assets that were downstreamed by the holding company to a faltering bank subsidiary. But, subsequently, Congress enacted two laws that enhanced the ability of the regulatory authorities to force bank holding companies to act as a source of strength in some circumstance. First, the Financial Institutions Reform, Recovery and Enforcement Act (FIRREA) of 1989 contained a cross-guarantee provision that permitted the FDIC to charge off any expected losses from a failing banking subsidiary to the capital of non-failing affiliate banks. Second, under the prompt corrective action section of the Federal Deposit Insurance Corporation Improvement Act (FDICIA) of 1991, the Federal Reserve Board was given authority to force a parent bank holding company to guarantee the performance of a troubled affiliate as part of a capital restoration plan.

Ashcraft (2004) has presented evidence that the ability of the FDIC to claim the capital in a non-failing banking subsidiary increased the incentives of bank holding companies to bail out a subsidiary before it fails and diminished the attractiveness of walking away from a distressed subsidiary. He concludes (Ashcraft, 2004: 19) that, 'In contrast to the historical experience before FIRREA, bank holding companies now appear to be a source of strength to their subsidiaries. Distressed affiliate banks are more likely to receive injections of capital than stand-alone banks, and recover from distress more quickly'.

In addition, financial groups sometimes voluntarily choose to forego the potential advantages of limited liability by explicitly guaranteeing the external debt of some subsidiaries, presumably to achieve more favorable borrowing terms. For example, Citigroup (2007: 156) provides explicit guarantees for external debt of four of its wholly owned subsidiaries: Citigroup Global Markets Holdings Inc., Citigroup Funding Inc., CitiFinancial Credit Company (CCC), and Associates First Capital Corporation.

Costs of financial distress: Protecting a subsidiary from the rest of the group

The growth of securitization has led to a proliferation of special purpose vehicles (SPVs),[8] which are designed to be financially insulated from the rest of the group. An SPV is a legal entity set up by a corporate sponsor for a specific, limited purpose. It buys pools of assets, usually originated by the sponsor, and issues debt to be repaid by cash flows from that pool of assets. It is tightly bound by a set of contractual obligations that ensure the activities of the entity are essentially predetermined at the inception of the vehicle. SPVs tend to be thinly capitalized, lack independent management or employees, and have all administrative functions performed by a trustee who receives and distributes cash according to detailed contracts. Most SPVs involved in securitization are organized as trusts, although they may also be organized as limited-liability companies, limited partnerships, or corporations. For some kinds of transactions substantial tax benefits can be achieved if an SPV is domiciled offshore—usually in Bermuda, the Cayman Islands, or the British Virgin Islands (Gorton and Souleles, 2006).

LCFIs (Bank of England, 2007b: 50) 'have been at the heart of the growth of the structured credit markets' and have dominant shares in arranging residential mortgage-backed and other asset-backed securitizations that rely heavily on SPVs. It is evident from Table 8.2 that trusts may represent a very substantial number of subsidiaries for each of the LCFIs. Some of these trusts are SPVs, but most securitization vehicles are unlikely to be included in our count of majority-owned subsidiaries because sponsors generally seek to avoid the appearance of voting control.

SPVs are constructed to be bankruptcy remote. The objective is to reassure investors in the SPV that their rights to the promised cash flows will not be compromised by financial distress or insolvency in the sponsor or its affiliates. Similarly, the SPV itself is structured so that it cannot be taken through bankruptcy. Typically, any shortfall of cash that would otherwise cause an event of default will trigger, instead, an early amortization of the pool of assets. The benefit of this structure is that it should avoid the deadweight costs of financial distress and so the debt issued by the SPV should not be subject to a bankruptcy premium. By separating the control rights over assets from the financing of these assets, the SPV reduces the costs of financial distress and thus the cost of debt financing (Gorton and Souleles, 2006).

Although the desire to avoid the deadweight costs of financial distress may be the primary motive for securitizing assets, Tufano (2006) notes that other factors may also be important. For example, SPVs may be formed to achieve more favorable

[8] The term 'special purpose entity' (SPE) is used more or less interchangeably. For an analysis of SPVs see also Strahan (Chap. 5 in this volume). For an overview of Structured Investment Vehicles (SIVs) see Allen and Saunders (Chap. 4 in this volume).

Table 8.2. Breakdown by industry of subsidiaries of Large Complex Financial Institutions

LCFIs	Banks	Insurance companies	Mutual & pension funds/nominees/ trusts/trustees	Other financial subsidiaries[1]	Non-financial subsidiaries[2]
ABN AMRO Holding NV	50	7	129	204	280
Bank of America Corporation	32	24	396	282	673
Barclays Plc	49	21	309	239	385
BNP Paribas	88	74	102	433	473
Citi	101	35	706	584	1,009
Credit Suisse Group	31	4	91	63	101
Deutsche Bank AG	54	9	458	526	907
Goldman Sachs Group, Inc.	7	4	48	151	161
HSBC Holdings Plc	85	37	246	381	485
JPMorgan Chase & Co.	38	17	229	145	375
Lehman Brothers Holdings Inc.	9	3	84	210	127
Merrill Lynch & Co. Inc.	16	9	85	89	68
Morgan Stanley	19	22	225	170	616
The Royal Bank of Scotland Group Plc	31	29	168	450	483
Société Générale	81	13	93	270	387
UBS AG	29	2	121	66	199
Total By Industry	**720**	**310**	**3,490**	**4,263**	**6,729**
% by industry	**5%**	**2%**	**22%**	**27%**	**43%**

Note: Year end 2007.

Source: Bankscope. Majority-owned subsidiaries.

[1]'Other financial subsidiaries' include private equity subsidiaries.

[2]'Non-financial subsidiaries' include all companies that are neither banks nor insurance companies nor financial companies. They can be involved in manufacturing activities but also in trading activities (wholesalers, retailers, brokers, etc. We have allocated foundations and research institutes to this category as well.

accounting treatment for the sponsor, to increase tax efficiency, to avoid regulatory capital requirements, to tap new pools of capital through changing the risk characteristics of a pool of assets, or to reduce the deadweight costs of information asymmetry by separating the funding of a more transparent pool of assets from the rest of the sponsor's balance sheet.

Protection of the bankruptcy-remote status of SPVs requires that the sponsor refrain from making any commitment to support the SPV. The concern is that a legal commitment might undo the bankruptcy-remote structure. If a sponsor should enter bankruptcy proceeding, the judge might recharacterize the sale of assets to the SPV as a secured financing, which would bring the assets back onto the sponsor's balance sheet. Attempts to minimize this possibility account for a considerable amount of the complexity of securitization vehicles. For example, sponsors often employ a two-tiered SPV structure to provide an extra layer of

insulation between the claims of the investors and the sponsor (Gorton and Souleles, 2006: 558).

The requirements for a true sale are set out in Financial Accounting Standard No. 140. The sponsor must surrender control of the assets sold to the SPV and the SPV must be a qualifying SPV (QSPV). QSPVs must be demonstrably distinct from their sponsors, as evidenced by the fact that the sponsor cannot unilaterally dissolve the SPV, and at least 10 percent of the fair value of its beneficial interests must be held by unrelated third parties (Gorton and Souleles, 2006: 556). QSPVs need not be consolidated in their sponsor's financial statements. Some variable interest entities (SPVs that do not meet the requirements for QSPVs) must be consolidated. Other VIEs, in which the sponsor is unlikely to absorb a majority of the expected losses or receive the majority of the expected residual returns, need not be consolidated (Soroosh and Ciesielski, 2004). Thus, at best, our measure of corporate complexity is likely to capture consolidated VIEs. For many of the LCFIs, this is a relatively small fraction of the total securitization activity. For example, JPMorgan Chase & Co. reports that in 2006 its revenue from qualifying special purpose entities (QSPEs) was almost fifteen times greater than the combined revenues of its consolidated and significant unconsolidated VIEs (JPMorgan Chase & Co., 2007: 59).

If SPVs are, in fact, bankruptcy-remote, would they complicate the unwinding of an LCFI? Perhaps not, but Gorton and Souleles (2006) present evidence that sponsors have supported their SPVs and, based on the pricing of debt issued by SPVs and the credit rating of the sponsoring institution, conclude that investors rely on this implicit support. Gorton and Souleles (2006) argue that this implicit commitment is essential to deal with moral hazard and adverse selection problems implicit in the asymmetric information between the originator of the assets and investors in the SPV. Nonetheless, the efforts by several LCFIs to support their SIVs and asset-backed commercial paper (ABCP) conduits during the turmoil in financial markets in the latter half of 2007 appear to have surprised shareholders and some regulators. In any event, this disconnect between explicit and implicit contracts complicates any analysis of how the existence of SPVs might affect the resolution of an LCFI experiencing extreme financial distress. Moreover, many of the innovative securitization structures have not been tested in a bankruptcy proceeding. Although these bankruptcy-remote structures may well turn out to be 'bulletproof', they are likely to complicate the resolution of a faltering LCFI, nonetheless.

The legacy of mergers and acquisitions

Mergers and acquisitions may have a significant impact on the degree of complexity of corporate structure. Relative to a firm of equal size that has grown

organically, an acquisitive financial conglomerate is likely to have many more subsidiaries, if only because it may be costly to close or consolidate them. Most of the LCFIs have engaged in a remarkable amount of merger activity. LCFIs have engaged in a large number of mergers, some of them exceptionally large. For example, since 1990, Bank of America, Deutsche Bank, JPMorgan Chase, and UBS have implemented mergers in which the target institution was larger than 10 percent of the acquiring firm's total assets (Thomson Securities Data Company). The acquiring firm may choose to retain a considerable amount of corporate separateness in the target firm for two reasons. First, it may perceive value in the brand and hope to retain the reputational capital of the target firm. Second, the willingness to retain the existing corporate structure may facilitate acceptance of the merger. As Dermine (2006) notes, by committing to keep in place a local structure and staff, local shareholders and the board of directors of the target may be reassured about the future of the target firm. Also, as we discuss below, host country regulatory authorities sometimes require that the acquiring bank maintain the target bank as a separate, locally chartered corporation. Dermine (2006) observes, however, that this decision to maintain a separate entity is often tactical rather than strategic. Over time, LCFIs generally decide to build a global brand identity, which may be inconsistent with the retention of separate subsidiaries bearing legacy names. Based on his interviews with ING and Nordea, Dermine (2006) found that even though both firms initially left many legacy organizations intact, they were also committed to building a global brand over time.

JPMorgan Chase provides a good example of how mergers may increase corporate complexity. The current organization is the result of a series of mergers of very large banks that began in 1991 with the merger of Chemical Bank Corporation and Manufacturers Hanover Corporation. This merger resulted in a near doubling of the size of the surviving institution, Chemical Bank, and, in 1996, was followed by the merger of Chemical Bank with The Chase Manhattan Corporation. The resulting institution merged with JPMorgan & Co., incorporated in 2000 forming JPMorgan Chase & Co. (JPMC). This series of mergers finally culminated in July 2004 with the merger of JPMC and Bank One Corporation (BOC). At year end 2003 JPMC had 248 wholly owned subsidiaries and BOC had 239. After the merger, at year end 2004, the surviving organization had 360 wholly owned subsidiaries (SEC Info database). (Note that the data in Tables 8.1 and 8.2 reflect majority-owned subsidiaries and are not directly comparable.) Although this represents nearly a 30 percent reduction relative to the combined total number of subsidiaries of the predecessor institutions, the result of the merger was, nonetheless, a much larger institution of considerably greater corporate complexity.

The efforts of JPMC to reduce its corporate complexity are consistent with evidence presented by Klein and Saidenberg (2005) that bank holding companies

with many bank subsidiaries are valued at a discount relative to similar bank holding companies with fewer bank subsidiaries. Although this conglomerate discount has sometimes been attributed to inefficient internal capital markets, they find that affiliated banks benefit from access to internal capital markets by lending more and holding less capital than comparable unaffiliated banks. Since activity and geographic diversification is broadly similar for their sample of affiliated and unaffiliated banks, they infer that the valuation discount is attributable mainly to greater complexity of organizational structure rather than diversification (but Laeven and Levine, 2007 adopt a different approach and find a diversification discount in financial conglomerates; they identify agency problems and insufficient economies of scope as probable causes). This finding may help explain why several large banks have attempted to simplify their corporate structures. Rosengren (2003: 111) presents evidence that from 1993 to 2002 eight large US bank holding companies reduced their number of subsidiaries relative to the number of subsidiaries in their predecessor organizations. Also, Citigroup (2007: 97) reported a consolidation project to merge twelve of its US-insured depository institutions into four. These efforts notwithstanding, continuing merger activity undoubtedly adds to corporate complexity.

Tax frictions

Taxes can have a major impact on the choice of corporate structure for all firms, especially international financial firms, because they tend to have more flexibility to shift profits from one entity to another. Demirgüç-Kunt and Huizinga (2001: 430) observe that '[M]ultinational banks, perhaps even more than other multinational firms, have opportunities for reducing their tax burdens in high-tax countries by way of intrafirm transfer pricing'. The choice of corporate structure may be influenced by income taxes (and the details of permissible deductions and credits), capital gains taxes, taxes on interest and dividends, value-added taxes, withholding taxes, transactions taxes and stamp duties.[9] It is difficult to generalize about the influence of taxes on corporate structure because tax codes differ markedly across countries, even among the relatively homogeneous members of the European Union. Moreover, the application of tax laws often depends on complex interpretations and rulings by the tax authorities.

[9] Banks are often subject to a number of implicit taxes as well, which may include the obligation to hold required reserves at the central bank at less than the market rate of interest or deposit insurance premiums that exceed the fair value of insurance.

Nonetheless, tax considerations appear to play a central role in a number of choices regarding corporate structure, including the location and organizational form of SPEs for leasing, real estate holdings, investment management, and private equity. In the US, specific tax code provisions make it advantageous to organize real estate mortgage investment conduits (REMICs), financial asset securitization investment trusts (FASITs), regulated investment companies (RICs), and real estate investment trusts (REITs) (Gorton and Souleles, 2006: 550). In general, SPEs are structured so that profits are not taxed in order to avoid double taxation that would otherwise occur if both the income of the sponsor and distributions from the SPE are taxed. Tax motives have also led to the creation of trusts for issuance of trust-preferred securities that are taxed like debt obligations, so that interest payments are deductible, yet are treated as Tier 1 capital by the bank regulatory authorities. Citi alone had established nineteen of these subsidiary trusts by the end of 2006 (Citigroup, 2007: 141).

Tax considerations are especially important for internationally active financial groups. Because home countries often tax groups on their consolidated world-wide income and, at the same time, most host countries tax locally generated income as well, cross-border transactions are usually subject to double taxation. Without some sort of relief, multiple taxes could stifle cross-border transactions completely.

Governments have devised a number of ways to alleviate double taxation, such as exempting foreign source income from the computation of taxable income or negotiating tax treaties to reduce or eliminate withholding taxes among pairs of countries. Some countries have also negotiated tax-sparing conventions to preserve tax concessions granted by less-developed countries. These conventions attempt to preserve the benefit of host-country tax incentives (such as tax holidays, credits, deductions, or exemptions) through tax sparing. In the absence of such tax-sparing arrangements these incentives may be reduced or eliminated by the home country, particularly when the home country provides recognition for taxes paid to the host country under the credit system. Tax-sparing treaties generally grant home country tax credit for taxes that were not actually collected by the home country. The rationale for such arrangements is that host country tax concessions are economically equivalent to grants or subsidies. Proponents of such treaties argue that, just as it would be inappropriate for the home country to insist on repayment by the parent company for grants or subsidies received by its foreign subsidiaries, it is inappropriate to recoup the value of tax incentives.

More broadly, when foreign source income is not exempt from taxation in the home country, firms are often permitted to credit foreign taxes paid against domestic tax owed. Generally, the foreign tax credit is limited by the amount of taxes that the firm would have paid if the income had been earned at home. Thus, firms have a strong incentive to reduce the average tax rate on foreign source income by shifting profits from relatively high-tax countries to tax havens (permissible

foreign tax credits may be constrained in other ways as well; see Demirgüç-Kunt and Huizinga (2001) for restrictions imposed on profit-shifting by the US).

A crude indication of the extent to which tax issues may have contributed to the corporate complexity of LCFIs may be seen in the number of entities located in tax havens. Our list of tax havens is based on the forty-two countries/territories/jurisdictions classified by the Financial Stability Forum as Offshore Financial Centers (Financial Stability Forum, 2000; and International Monetary Fund, 2000). The list includes countries/territories/jurisdictions which provide low or zero taxation, moderate or light financial regulation, and/or banking secrecy and anonymity. Of course, the impact of tax issues on organizational complexity is much more pervasive and complex than can be represented by a count of the number of subsidiaries in these centers. Nonetheless, even this number is substantial for some of the LCFIs (see Table 8.1). Six of our LCFIs each have more than 100 subsidiaries located in these booking centers. Moreover, three of the LCFIs have located nearly 20 percent of their subsidiaries in tax havens.

Regulatory constraints

All of the preceding rationales for corporate separateness—asymmetric information problems, insulation against risk, the legacy of mergers and acquisitions, and taxes—apply to large corporations in general, not just financial groups. But financial groups are subject to an additional source of constraints that complicates their corporate structures—regulation. This may help explain, at least in part, why they have a substantially greater number of subsidiaries than non-financial groups of comparable size. On average, the sixteen LCFIs have nearly two and a half times as many majority-owned subsidiaries as the sixteen largest non-financial firms ranked by market capitalization at year end 2007 (Bankscope and Osiris data).

Banks are among the most regulated institutions in every country, although countries differ with regard to the constraints imposed on banks' expansion into other lines of business. Broadly, three different regulatory models can be discerned: (1) complete integration; (2) parent bank with non-bank operating subsidiaries; and (3) holding company parent with bank and non-bank affiliates.[10] Universal banking countries (see Chapter 7 in this volume) tend to follow the first model, with only minimal corporate separateness imposed for regulatory reasons. For

[10] Seven of the 16 LCFIs have bank holding companies (source: Bankscope). See Herring and Santomero (1990) for a more detailed discussion of these models and their variations.

example, Germany allows the combination of bank and securities businesses in a single legal entity. The US Comptroller of the Currency (which regulates banks, but not bank holding companies) has long argued for the second model and has permitted the national banks, which it supervises, to create subsidiaries to conduct some non-bank activities. The dominant model in the US, however, is the third. Moreover, the corporate separateness imposed on bank holding companies and financial services holding companies is reinforced by restrictions on the flows of credit between different functional units and the bank. Sections 23A and 23B of the Federal Reserve Act limit the amount of credit from banks to their affiliates and require that such transactions be collateralized and made at market prices. The Gramm-Leach-Bliley Act, which authorized financial services holding companies, extended these provisions to credit flows between banks and their own financial subsidiaries and, to some extent, to flows between holding companies and the financial subsidiaries of banks.

In a survey of 143 countries Barth, Caprio, and Levine (2007) find that of the majority of 127 countries that permit banks to engage in some securities activities, fifty-nine impose some form of corporate separateness on these activities. Of the eighty-seven countries that permit banks to engage in the insurance business, eighty-five impose some form of corporate separateness. Finally, of the sixty-two countries that permit banks to engage in the real estate business, forty-five require some form of corporate separateness.

In countries that have not adopted the single or integrated regulator model, different functional regulators often require that the activities which they regulate be conducted in separate legal entities. This not only facilitates oversight, but makes it easier to ring-fence those activities should it become necessary to intervene.[11] Thus, even without consideration of the complexities introduced by international expansion, financial conglomerates may be required to adopt a certain amount of corporate separateness for regulatory purposes.

LCFIs have established subsidiaries in numerous countries (see column 7 in Table 8.1) and international expansion may require substantial additional corporate complexity for two reasons. First, host countries that apply some variation of model three to domestic financial conglomerates generally impose the same restrictions on foreign firms to maintain a level playing field. The fact that the US, the largest market in the world for financial services, applies model three to domestic and foreign firms can account for a significant amount of the complexity of the corporate structure of LCFIs headquartered outside the US.

Second, even if the host country has not adopted a variation of model three for domestic firms, it may require that foreign-owned firms incorporate locally to

[11] In some jurisdictions it is possible to ring-fence entities that are not separately incorporated; for example, the US regulatory authorities can ring-fence a foreign branch.

ensure that the domestic authorities can intervene to protect domestic residents. New Zealand provides, perhaps, the most extreme example of the second rationale. More than 85 percent of the banking system is controlled by foreign-owned banks and the New Zealand authorities have been uncomfortable accepting the passive role often associated with host country oversight of resident foreign branches (Woolford and Orr, 1995). They have insisted that systemically important foreign entities be organized as subsidiaries. Moreover, they have buttressed this corporate separateness by additional measures that ensure that a subsidiary could continue operation without interruption (and without its previous owners) should it become necessary.

Barth, Caprio, and Levine (2007) find that in their sample of 143 countries only three countries prohibit entry by foreign subsidiaries, but twenty-eight countries prohibit entry by foreign branches. Moreover, even if foreign branch entry is not prohibited, host countries often impose stricter regulatory requirements on foreign branches that make the formation of a separate subsidiary relatively attractive. For example, of the nineteen Latin American and Central European countries surveyed by Cerutti, Dell'Ariccia, and Martinez-Peria (2005), seven restrict foreign branches more heavily than foreign subsidiaries.

Functional and national regulators frequently employ corporate separateness as a means of regulating, supervising, and monitoring the part of a financial conglomerate that falls in their bailiwick. While this may enhance local regulatory oversight, an unintended consequence may be that international financial conglomerates may have significantly more complex corporate structures than domestic firms of comparable size.

More broadly, LCFIs often respond to new regulations with still more corporate complexity. Kane (1977; 1981; and 1984) has characterized this dynamic as a regulatory dialectic, in which regulators impose a rule (or implicit tax) and the regulated firms react within their constrained environment to minimize the implicit tax burden. The regulators in turn react to perception of regulatory avoidance with still more regulations. Robert Eisenbeis, in correspondence with the authors, described how the regulatory dialectic evolved under the Bank Holding Company (BHC) Act:

> From the very beginning, financial conglomerates exploited the BHC loopholes to expand geographically as well as into new activities. Finance companies were acquired to expand across state lines. Credit card special purpose banks were designed to get around usury ceilings. SPEs and off-balance sheet activities were designed to avoid capital constraints. Mortgage banking subsidiaries were established to avoid having to pay taxes for doing business across state lines.

This kind of dynamic has undoubtedly increased the corporate complexity of LCFIs. In the event of financial distress, however, this complexity could impede an effective regulatory response.

Implications of corporate complexity for safety and soundness of the financial system

Despite their corporate complexity, LCFIs tend to be managed in an integrated fashion along lines of business with only minimal regard for legal entities, national borders, or functional regulatory authorities. Moreover, there are often substantial interconnections among the separate entities within the financial group. Baxter and Sommer (2005) note that, in addition to their shared (although possibly varying) ownership structure, the entities are likely to be linked by cross-affiliate credit relationships, cross-affiliate business relationships, and reputational relationships.

What would happen should one of these LCFIs experience extreme financial distress? Quite apart from the difficulty of disentangling operating subsidiaries that provide critical services to other affiliates and mapping an integrated firm's activities into the entities that would need to be taken through a bankruptcy process, the corporate complexity of such institutions would present significant challenges. The fundamental problem stems from conflicting approaches to bankruptcy across regulators, across countries, and, sometimes, even within countries. There are likely to be disputes over which law and which set of bankruptcy procedures should apply. Some authorities may attempt to ring-fence the parts of an LCFI within their reach to satisfy their regulatory objectives without necessarily taking into account some broader objective such as the preservation of going concern value or financial stability. At a minimum, authorities will face formidable challenges in coordination and information sharing across and among jurisdictions. Yet, experience has shown that in times of stress information-sharing agreements are likely to fray (see Eisenbeis and Kaufman, 2008 and Herring, 2007, for examples).

Bad news tends to be guarded as long as possible. Managers of a regulated entity are often reluctant to share bad news with their regulators because they fear they will lose discretion for dealing with the problem (and, indeed, may lose their jobs). Similarly, the primary supervisor of the regulated entity is likely to be reluctant to share bad news with other supervisory authorities out of concern that the leakage of bad news could precipitate a liquidity crisis or that other supervisory authorities might take action—or threaten to take action—that would constrain the primary supervisor's discretion for dealing with the problem or cause it to take action rather than forbear. As Baxter, Hansen, and Sommer (2004: 79) note, 'Once the bank's condition degrades, supervisors think less about monitoring and more about protecting their creditors. This creates a conflict among supervisors' (see Kane, 1989 for a thorough analysis of the incentives to forbear).

Generally, the primary supervisor will use its discretion to forbear so long as there is a possibility that the regulated entity's condition may be self-correcting,

particularly if the alternative is closure. A closure decision is sure to be challenged and so supervisors will tend to forbear until losses are so large that there can be no reasonable doubt that the entity is insolvent. Losses that spill across national borders, however, will intensify conflicts between home and host authorities and make it difficult to achieve a cooperative resolution of an insolvent financial group (see Eisenbeis and Kaufman, 2006 for an analysis of differences in resolution policies and procedures among member countries of the European Union). Freixas (2003) has argued that disagreements regarding the causes of losses and metrics for allocating losses across countries would lead to the underprovision of recapitalizations of international banks even when the social benefits of recapitalization exceed the cost.

Within the relatively homogeneous banking sector, despite thirty years of harmonization initiatives by the Basel Committee on Banking Supervision, approaches to bank resolution differ substantially across countries. For example, countries differ with regard to the point at which a weak bank requires resolution. In many countries, intervention is required when a bank's net worth (which may be defined in a number of different ways) declines to zero, but in the US, which has adopted a Structured Early Intervention and Resolution policy, action must be taken when the ratio of tangible equity to total assets is equal to or less than 2 percent. In Switzerland, the authorities may intervene even earlier if they perceive a threat to depositors' interests. Countries also differ with regard to what entity initiates the resolution process. The supervisory authorities? The courts? Or the bank itself? Barth, Caprio, and Levine (2007) find significant differences across the 143 countries they survey. The bank supervisor can legally declare that a bank is insolvent in sixty-six countries. Courts have this prerogative in ninety-seven countries and the deposit insurance agency only in four, while in twenty-six other countries this function is exercised by other agencies or the bank itself. In many countries more than one entity can declare insolvency. Clearly cross-border differences in regard to how and when the resolution process is initiated can cause delays that may be costly in a crisis.

In the event that an entity is declared insolvent, which jurisdiction will oversee the insolvency? The place where the bank was chartered? Where the management resides? The principal place of business? The domain of the largest concentration of assets? Or where the largest concentration of creditors resides? The collapse of BCCI revealed that each of these questions may have a different answer. Baxter, Hansen, and Sommer (2004: 61) observe that it is difficult to devise a good jurisdictional rule that 'would be both *ex ante* predictable (to defeat forum shopping or subsequent jurisdictional squabbling) and sensible in application (to discourage name-plate incorporations or prevent unseemly jurisdictional choices)'.

The choice of jurisdiction, however, may have important implications for the outcome of the insolvency proceedings. Most countries have adopted a universal approach to insolvency in which one jurisdiction conducts the main insolvency

proceedings and makes the distribution of assets, while other jurisdictions collect assets to be distributed in the main proceedings. But the US follows a more territorial approach with regard to US branches of foreign banks and will conduct its own insolvency proceedings based on local assets and liabilities. Assets are transferred to the home country only after (and if) all local claims are satisfied.

The choice of jurisdiction will also determine a creditor's right to set off claims on the insolvent bank against amounts that it owes the bank. The BCCI case revealed striking differences across members of the Basel Committee (Basel Committee, 1992). In the US, the right of set off can be exercised only with regard to claims denominated in the same currency at the same branch. Claims denominated in different currencies or at different branches may not be set off. In contrast, in the UK, the right to set off may be exercised even when the claims are not denominated in the same currency, at the same branch or even at branches in the same country. And in Luxembourg the right to set off may not be exercised after a liquidation order and may be exercised before a liquidation order only when the claims are fixed in amount, liquid, and mature.

Similarly, the ability to exercise close-out netting provisions under the International Swap Dealers Association (ISDA) Master Contracts may vary from jurisdiction to jurisdiction. In principle, in the event of a default, the non-defaulting counterparty can close out all existing transactions under the Master Agreement, which may include many different kinds of derivative contracts with many different affiliates of the defaulting entity, making them immediately due and payable. The non-defaulting counterparty can then offset the amount it owes the defaulting entity against the amount it is owed to arrive at a net amount. In effect, close-out netting permits the non-defaulting counterparty to jump the bankruptcy queue for all but the net value of its claims. But the ability to apply close-out netting and the extent to which it may be applied may depend on whether the country in which the insolvency proceeding is conducted has enacted legislation to ensure that all outstanding transactions under a master netting agreement can be terminated upon the occurrence of an insolvency and that close-out netting will be respected by the bankruptcy trustee.

The outcome of insolvency proceedings will also depend on the powers and obligations of the resolution authority, which may differ from country to country. For example, does the resolution authority have the authority to impose 'haircuts' on the claims of creditors without a lengthy judicial proceeding? Does the resolution authority have the power (and access to the necessary resources) to provide a capital injection? With regard to banks, is the resolution authority constrained to choose the least costly resolution method, as in the US?[12] Or is the resolution

[12] The US resolution authority can choose a resolution method that is more costly to the FDIC only if the systemic risk exception is invoked. This requires agreement by two-thirds of the Federal Reserve Board, two-thirds of the FDIC Board, and the Secretary of the Treasury in consultation with the President that the implementation of the least costly resolution method would have serious adverse effects on economic conditions or financial activity.

authority obliged to give preference to domestic depositors as the law requires in Australia and the US?

More fundamentally, what is the objective of the supervisory intervention and the resolution process? Is it to protect the domestic financial services industry? Or to safeguard the domestic financial system? Or to protect domestic employment? Or to protect the deposit insurance fund? Or to minimize the fiscal costs of the insolvency to domestic taxpayers? Or to minimize the spillover costs in all countries in which the insolvent bank conducts business? Only the last of these alternatives is implausible. The priority that supervisors will inevitably place on domestic objectives in the event of insolvency is the essential source of conflict between home and host authorities.

Three asymmetries between the home and host country may create additional problems even if procedures could be harmonized. First is asymmetry of resources. Although international agreements among sovereigns are, necessarily, based on the polite fiction that all sovereigns are equal, this is demonstrably not the case. Supervisory authorities may differ greatly in terms of human capital—the number and quality of employees—and financial resources. This means that even if the fundamental conflicts of interest could be set aside, the home supervisory authority may not be able to rely on the host supervisory authority (or vice versa) simply because it may lack the capacity to conduct effective oversight.

Second, asymmetries of financial infrastructure may give rise to discrepancies in the quality of supervision across countries. Weaknesses in accounting standards and the quality of external audits may impede the efforts of supervisors just as informed, institutional creditors and an aggressive and responsible financial press may aid them. The legal infrastructure matters as well. Inefficient or corrupt judicial procedures may undermine even the highest quality supervisory efforts.

Perhaps the most important conflict, however, arises from asymmetries of exposures: what are the consequences if the entity should fail? Perspectives may differ with regard to whether a specific entity jeopardizes financial stability. This will depend on whether the entity is systemically important in either or both countries and whether the foreign entity is economically significant within the parent group.

A number of proposals have been advanced recently to enhance the oversight of LCFIs and safeguard their solvency. For example, Čihák and Decressin (2007) propose the creation of a European Banking Charter, to improve and harmonize supervision of LCFIs with systemic cross-border exposures. Nieto and Schinasi (2007) focus on decentralization and cooperation issues which arise from the nature of public good of the European Union financial stability. Garcia and Nieto (2007) question the effectiveness of decentralization and voluntary cooperation in safeguarding financial stability in the European Union and support the enhancement of market discipline and the adoption of prompt corrective action and least-cost resolution. Mayes, Nieto, and Wall (2007) propose a US-style

prompt-corrective-action framework for preventing cross-border banking crises in the European Union. Hüpkes (2005) advocates adoption of a functional approach to regulation and supervision. She favors a tighter alignment between legal entities and the functions they perform so that systemically important functions could be more easily protected in the event of a crisis either by insulating them from problems in the rest of the LCFI or detaching them from the LCFI. While these proposals to enhance supervision have many attractive features, none can be relied upon to prevent insolvencies. Thus it is also important to consider ways to improve the resolution of insolvent institutions.

Concluding comments

The corporate complexity of LCFIs is likely to defy efficient resolution in the event of bankruptcy. It seems doubtful that going-concern value could be protected adequately and, worse still, the unwind is likely to spill-over to damage other institutions and market participants if counterparties attempt to liquidate positions at once, driving down prices and causing problems for other investors with similar positions. In the absence of workable procedures to unwind the affairs of a failing LCFI in an orderly manner, the result is likely to be a chaotic scramble for assets that could infect other markets and institutions, with potential disruption of the real economy.

Despite *ex ante* protestations to the contrary, the authorities are unlikely to risk such an outcome and so the result is likely to be a bailout that will prop up the failing group. The continuation of recent trends toward globalization, conglomeration, consolidation, and increasing reliance on trading of over-the-counter (OTC) derivatives implies that we may be confronted with a growing category of firms that are 'too complex to fail'. This, of course, has ominous implications for moral hazard. A market perception that such firms will benefit from official support in times of stress gives them a competitive advantage completely unrelated to their ability to add value to the financial system. It dulls the incentives for creditors to demand disclosure of risky positions and monitor such exposures. Weakened market discipline will enable such institutions to take larger, riskier positions without paying appropriately higher risk premiums to their creditors. The result may be larger potential insolvencies that require still larger bailouts to forestall systemic risk.

For market discipline to operate effectively in constraining risk taking by LCFIs, the regulatory authorities need a credible procedure to unwind the affairs of an LCFI in an orderly manner, without systemic spillovers. Simplification of the

corporate structure of large complex financial institutions would be a good place to start. Since regulatory and tax policies have contributed significantly to the problem, they need to be part of the solution.

Postscript on the bankruptcy of Lehman Brothers

The preceding was completed early in 2008 before the actual collapse of an LCFI. The editors asked us to reflect on what we had learned from the costly experiences of 2008 about the implications of corporate complexity for systemic risk. The first observation to be made is that the list of sixteen LCFIs was not nearly long enough to reflect the perceptions of regulators when they were confronted with the prospect of collapse of institutions not on the LCFI list. Bear Stearns may be the most obvious case. Although Bear Stearns was one of the five largest investment banks in the US, it was little more than half the size of the fourth-largest investment bank, Lehman Brothers (LB). Nonetheless, when Bear Stearns was about to collapse,

Table 8.3. Corporate structure of Lehman Brothers Holdings Inc.

Country	Number of majority-owned subsidiaries
USA	238
United Kingdom	120
Cayman Islands	18
Australia	9
Luxembourg	6
Ireland	5
Netherlands	5
Bermuda	4
France	4
Hong Kong	4
Japan	4
Korea (Republic of)	4
Germany	3
Singapore	2
Thailand	2
Argentina	1
Canada	1
Switzerland	1
India	1
Mauritius	1
Total	**433**
Number of countries	**20**

Note: Year end 2007.

Source: Bankscope. Majority-owned subsidiaries.

the US regulatory authorities subsidized a merger of Bear Stearns with JPMorgan Chase out of concern for the 'interconnectedness' of Bear Stearns with the rest of the financial system. The enormous subsidy to AIG is another case in which intervention was justified on similar grounds. Other countries took similar measures to support other institutions that were not large enough or complex enough to make the official list.

There was one significant exception, however, to the general trend of hastily improvised bailouts. After trying to broker a merger of LB with other, stronger institutions, the US authorities declined to bail it out and sent the holding company, Lehman Brothers Holdings International (LBHI), to the bankruptcy courts for protection under Chapter 11 of the US bankruptcy code, the largest bankruptcy in US history. Although LB was by far the smallest and one of the least complex institutions on the list of LCFIs, it was nonetheless of sufficient systemic importance that its collapse led to substantial spillovers on global capital markets. Credit risk spreads rose to record highs, equity prices fell by 4 percent worldwide when the bankruptcy was announced and government bond yields declined sharply as foreign exchange carry trades were unwound.

Lehman's total reported assets were roughly $700 billion. Table 8.3 shows its corporate structure at the end of 2007. It included 433 subsidiaries in twenty countries. This corporate complexity greatly impeded the orderly resolution of the firm, created significant spillovers to other institutions and markets, and led the Group of 7 finance ministers to pledge (Guha, 2008) 'to do everything in their power to prevent any more Lehman Brothers-style failures of systemically important financial institutions'.[13]

Understandably, after the US government had subsidized the merger of Bear Stearns, a much smaller, less complex investment bank, the market expected that Lehman Brothers would receive similar treatment. Why then was LB permitted to fail? The Fed and the Treasury claimed they lacked authority to bail it out. It is also likely that they wished to limit moral hazard by engaging in a bit of 'constructive ambiguity', a dubious remedy at a time when a consistent policy framework might have helped stabilize expectations. Moreover, since they had a team of examiners in LB ever since the collapse of Bear Stearns, they knew much more about the condition of LB and may have believed they could predict and control the spillover costs. They may have thought that counterparties and creditors had sufficient warning about LB's weakening condition to take precautionary measures. But, of course, in a complex and integrated financial system, regulatory action or inaction can have unintended consequences through indirect exposures and linkages that are apparent only after the fact.

[13] Observers said that it came close to a Group of seven-wide temporary implicit guarantee for many or all of the liabilities of systemically important financial firms. See Guha (2008).

One of the major concerns was that LB was the sixth largest counterparty in OTC derivatives markets. But, back offices succeeded in processing billions of dollars of contracts and the International Swap Dealers Association organized an auction to determine settlement prices. Because derivatives contracts in which LB was a counterparty were usually marked to market daily and collateral was adjusted each evening to reflect changes in market prices, losses were relatively light. Losses were much greater, however, with regard to credit default swap contracts written on LB. Those selling protection on LB are in a similar position to bondholders and received a similar price. Buyers of $100 of default protection will receive $91.375, a substantial loss for sellers of protection.

A second major concern was LB's key role in the Repo market, which totals roughly $11 trillion and is the short-term, collateralized lending market that banks, broker/dealers, and hedge funds use to finance securities positions. The Fed attempted to address the risk that the market would seize up by allowing broader use of the Primary Dealer Credit Facility through expanding the list of eligible securities. In addition, a group of global banks announced plans to use their own capital to establish a $70 billion private sector credit facility for those securities not eligible for the Fed facility. The Fed also announced an increase in its Treasury Securities Lending Facility to $200 billion.

What turned out to be more disruptive were the traditional exposures to LB's outstanding debt. Among the largest unsecured creditors were the US federal government's Pension Benefit Guaranty Corp., the German government's deposit insurance arm (McCracken, 2008), and money market mutual funds. The last proved to be one of the most important channels of contagion. One of the oldest money market funds, the Reserve Primary fund, was forced to write off $785 million of short- and medium-term notes and became the first money market mutual fund to 'break the buck' in fourteen years. This triggered $184 billion in money market mutual fund redemptions and forced fund managers to sell assets into illiquid markets. This spilled over into commercial paper markets including not only asset-backed commercial paper, but also non-asset-backed commercial paper that had held up reasonably well and was a key means of financing corporations and banks. The interbank market seized up entirely with the almost complete collapse of confidence in counterparties in money markets. Spreads between the euro-dollar interbank rate and the comparable US Treasury rate rose to nearly 450 basis points, more than double the already high spreads that prevailed before the LB bankruptcy.

In addition, failed trades proved particularly disruptive. Prior to LB's bankruptcy, portfolio managers placed thousands of trades with LB's broker dealer, Lehman Brothers International, many of which were subsequently transferred for settlement to LBI affiliates throughout the world. After the bankruptcy, these failed to settle and this has led to civil proceedings on three continents. The UK administrator said that about 43,000 trading deals were still 'live' in the London subsidiaries alone and would need to be negotiated with each counterparty (Hughes, 2008a).

But, the fundamental problem was that LB was managed as an integrated entity with minimal regard for the legal entities that would need to be taken through the bankruptcy process. LBHI issued the vast majority of unsecured debt and invested the funds in most of its regulated and unregulated subsidiaries. This is a common approach to managing a global corporation, designed to facilitate control over global operations, while reducing funding, capital, and tax costs. LBHI, in effect, served as banker for its affiliates, running a zero-balance cash-management system. LBHI lent to its operating subsidiaries at the beginning of each day and then swept the cash back to LBHI at the end of each day. The bankruptcy petition was filed before most of the subsidiaries had been funded on 15 September and so most of the cash was tied up in court proceedings in the US.

Lehman also centralized its information technology so that data for different products and different subsidiaries were co-mingled. This was an efficient way of running the business as a going concern, but presents an enormous challenge in global bankruptcy proceedings. LB stored data in 26,666 servers, 20,000 of which contained accumulated emails, files, voicemail messages, instant messages, and recorded calls. The largest data centers were in New York, London, Tokyo, Hong Kong, and Mumbai. Moreover, LB used approximately 2,700 proprietary, third-party, and off-the-shelf programs, each of which interacted with or created transactions data. The bankruptcy administrators must preserve, extract, store, and analyze data relevant to the entities they are dealing with. This problem was made more difficult by the success of the administrators of LBHI in selling two important entities that were rapidly declining in value because of loss of human capital: its investment banking operations and its asset management business.

Most of the US investment banking operations—the assets, not the legal entities—were sold to Barclays. This necessitated bringing a Securities Investor Protection Corporation (SIPC) proceeding, which put all LBI accounts under the control of the SIPC Trustee and permitted the broker-dealer to be liquidated. Nomura bought most of the investment banking business in Asia and continental Europe and LB's asset managementbusiness was sold in a management buyout. But this meant that the data were owned by Barclays, Nomura, and the now independent asset management division and so bankruptcy administrators are dependent on the new owners for access to data to determine the assets and liabilities of each legal entity. The administrator of the four London subsidiaries complained that nine weeks after the bankruptcy, he has yet to receive a confirmation of the assets owned by these subsidiaries.

The US administrators expressed the optimistic view that they would be able to complete the resolution within eighteen to twenty-four months, but the presiding judge reminded the administrator that the biggest impediments to a timely completion of the administration are the timetables of the other insolvency fiduciaries around the world. The administrators in London warned that it may take years for creditors to get their money back, noting that they were continuing to work on

Enron, which failed seven years ago, which was about one-tenth the size and complexity of Lehman (Hughes, 2008b).

The conclusion we draw from the LB experiment is not that all systemically important institutions should be bailed out, but rather than regulators and supervisors should focus on devising orderly resolution plans that will enable them to unwind even the largest, most complex institution with minimal spillover to the rest of the financial system. A useful first step would be to require that each institution create and maintain a plan for winding down the institution just as they now maintain plans for business continuity. The bankruptcy administrator of LBHI has claimed that the hastily prepared bankruptcy filing has cost as much as $75 billion in lost value (McCracken, 2008). If the regulators deem the plan unworkable, the institution may be required to reduce its complexity or set aside a higher capital charge. An institution that is 'too complex to fail' is simply too complex and presents too great a threat to the rest of the financial system.

References

Aron, D. (1991). 'Using the Capital Market as a Monitor: Corporate Spinoffs in an Agency Framework', *RAND Journal of Economics*, 22: 505–18.

Ashcraft, A. (2004). 'Are Bank Holding Companies a Source of Strength to Their Banking Subsidiaries?', Federal Reserve Bank of New York Staff Report 189, June.

Bank of England (2001). 'The Financial Stability Conjuncture and Outlook', *Financial Stability Review*, December.

—— (2007a). 'Overview', *Financial Stability Report*, 21, April.

—— (2007b). 'Structure of the Financial System', *Financial Stability Report*, 22, October.

Basel Committee (1992). 'The Insolvency Liquidation of a Multinational Bank', *BCBS Compendium of Documents, International Supervisory Issues*, III (May 2001).

Baxter, T., Hansen, J., and Sommer, J. (2004). 'Two Cheers for Territoriality: An Essay on International Bank Insolvency Law', *American Bankruptcy Law Journal*, 78: 57–91.

—— and Sommer, J. (2005). 'Breaking Up is Hard to Do: An Essay on Cross-Border Challenges in Resolving Financial Groups', in Evanoff, D. and Kaufman G. (eds.), *Systemic Financial Crises, Resolving Large Bank Insolvencies*, Singapore: World Scientific, 175–91.

Berger, A. (1995). 'The Profit-Structure Relationship in Banking—Tests of Market-Power and Efficient Structure Hypotheses', *Journal of Money, Credit and Banking*, 27: 404–31.

—— Demsetz, R. S., and Strahan, P. E. (1999). 'The Consolidation of the Financial Services Industry: Causes, Consequences, and Implications for the Future', *Journal of Banking & Finance*, 23: 135–94.

—— Herring, R. and Szegö, G. P. (1995). 'The Role of Capital in Financial Institutions', *Journal of Banking & Finance*, 19: 393–430.

Bianco, M. and Nicodano, G. (2002). 'Business Groups and Debt', Working Paper, Universitá di Torino, January 29.

Caprio, G., Levine, R. E., and Barth, J. R. (2007). 'Database on Bank Regulation and Supervision', World Bank.

Cerutti, E., Dell'Ariccia, G., and Martinez Peria, M. S. (2005). 'How Banks Go Abroad: Branches or Subsidiaries?', World Bank Policy Research Working Paper 3753, October 2005.

Čihák, M. and Decressin, J. (2007) 'The Case for a European Banking Charter', IMF Working Paper, WP/07/173.

Citigroup Inc. (2007). '2006 Financials and Form 10-K'. < http://www.citigroup.com/citi/fin/data/k07cu.pdf>.

Cox, R. and Currie, A. (2007). 'Demystifying Goldman, Some of Trading Complex is Moved Out of Shadows; Bid for Investor Respect?', *Wall Street Journal*, October 4, C12.

Demirgüç-Kunt, A. and Huizinga, H. (2001). 'The Taxation of Domestic and Foreign Banking', *Journal of Public Economics*, 79: 429–53.

Dermine, J. (2006). 'European Banking Integration: Don't Put the Cart before the Horse', *Financial Markets, Institutions & Instruments*, 15: 57–106.

Eisenbeis, R. and Kaufman, G. (2006). 'Cross-Border Banking: Challenges for Deposit Insurance and Financial Stability in the European Union', Federal Reserve Bank of Atlanta Working Paper 15.

Financial Stability Forum (2000). Press Release, 26 May 2000.

Freixas, X. (2003). 'Crisis Management in Europe', in Kremers, J., Schoenmaker, D., and Wierts, P. (eds.), *Financial Supervision in Europe*, Cheltenham: Edward Elgar, 102–19.

Fulghieri, P. and Hodrick, L. (2005). 'Synergies and Internal Agency Conflicts: The Double-Edge Sword of Mergers', Working Paper, Kenan-Flagler Business School.

Garcia, G. and Nieto, M. J. (2007). 'Preserving Financial Stability: A Dilemma for the European Union', *Contemporary Economic Policy*, 25: 444–58.

Gorton, G. and Souleles, N. (2006). 'Special Purpose Vehicles and Securitization', in Carey, M. and Stulz, R. (eds.), *The Risks of Financial Institutions*, Chicago: National Bureau of Economic Research, 549–97.

Group of Ten. (2001). Report on Consolidation in the Financial Sector (Ferguson Report) < http://www.bis.org/publ/gten05.htm>

Guil, G. (2008). 'Bankers Trust and the Birth of Modern Risk Management', Case Study, Wharton Financial Institutions Center.

Guha, K. (2008). 'G7 Pledge "No More Lehmans".' *Financial Times*, October 12.

Habib, M., Johnsen, B. and Naik, N. (1997). 'Spinoffs and Information', *Journal of Financial Intermediation*, 6: 153–76.

Hawkesby, C., Marsh, I., and Stevens, I. (2003). 'Large Complex Financial Institutions: Common Influences on Asset Price Behaviour?', *Financial Stability Review*, Bank of England, December, 91–101.

Herring, R. (2003). 'International Financial Conglomerates: Implications for National Insolvency Regimes', in Kaufman G. (ed.), *Market Discipline and Banking: Theory and Evidence*, Elsevier, 99–129.

—— (2007). 'Conflicts between Home and Host Country Prudential Supervisors', *International Financial Instability, Global Banking and National Regulation*, Evanoff, D., Kaufman, G. and LaBrosse, J. (eds.), World Scientific, 201–19.

—— and Carmassi, J. (2008). 'The Structure of Cross-Sector Financial Supervision', *Financial Markets, Institutions & Instruments*, 17: February.

—— and Santomero, A. (1990). 'The Corporate Structure of Financial Conglomerates', *Journal of Financial Services Research*, December, 471–97.

Herring, R. and T. Schuermann. (2005). 'The Regulation of Position Risk in Banks, Securities Firms and Insurance Companies', in Scott, H. (ed.), *Capital Adequacy Beyond Basel, Banking, Securities and Insurance*, Oxford: Oxford University Press, 15–86.

Holod, D. and Peek, J. (2006). 'Capital Constraints, Asymmetric Information and Internal Capital Markets in Banking: New Evidence', Working Paper available on SSRN website, December.

Huertas, T. (2006). 'Dealing with Distress in Financial Conglomerates', paper presented at conference, 'Prompt Corrective Action and Cross-Border Supervisory Issues in Europe', Financial Markets Group, LSE, November 20.

Hughes, J. (2008a). 'Lehman Highlights Need for Plan', *Financial Times*, November 8.

—— (2008b), 'Lehman Creditors to Face Years of Waiting', *Financial Times,* November 14.

Hüpkes, E. H. G. (2005). 'Too Big to Save—Toward a Functional Approach to Resolving Crises in Global Financial Institutions', in Evanoff D. and Kaufman, G. (eds.), *Systemic Financial Crises, Resolving Large Bank Insolvencies*, Singapore: World Scientific, 193–215.

International Monetary Fund (2000). 'Offshore Financial Centers—IMF Background Paper', June 23.

Jensen, M. (1986). 'Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers', *The American Economic Review*, 76: 323–9.

JPMorgan Chase & Co. (2007). Annual Report 2006.

Kane, E. (1977). 'Good Intentions and Unintended Evil: The Case against Selective Credit Allocation', *Journal of Money Credit and Banking,* 9: 55–69.

—— (1981). 'Accelerating Inflation, Technological Innovation, and the Decreasing Effectiveness of Banking Regulation', *Journal of Finance*, 36: 355–67.

—— (1984). 'Technological and Regulatory Forces in the Developing Fusion of Financial Services Competition', *Journal of Finance*, 39: 759–72.

—— (1989). 'Changing Incentives Facing Financial Services Regulators', *Journal of Financial Services Research*, 2: 265–74.

Klein, P. and Saidenberg, M. (2005). 'Organizational Structure and the Diversification Discount: Evidence from Commercial Banking', CORI Working Paper 2005–06, University of Missouri, April.

Kroszner, R. and Rajan, R. (1997). 'Organization Structure and Credibility: Evidence from Commercial Banks Securities Activities before the Glass-Steagall Act', *Journal of Monetary Economics*, 39: 475–516.

Laeven, L. and Levine, R. (2007). 'Is There a Diversification Discount in Financial Conglomerates?' *Journal of Financial Economics*, 85: 331–67.

Mayes, D. G., Nieto, M. J., and Wall, L. (2007). 'Multiple Safety net Regulators and Agency Problems in the EU: Is Prompt Corrective Action a Partial Solution?' Bank of Finland Discussion Paper, 7/2007.

Merton, R. and Perold, A. (1993). Management of Risk Capital in Financial Firms, in Hayes, S. (ed.), *Financial Services: Perspectives and Challenges,* Harvard Business School Press, 215–45.

McCracken, J. (2008). 'Lehman's Chaotic Bankruptcy Filing Destroyed Billions in Value', *Wall Street Journal*, December 29.

Modigliani, F. and Miller, M. (1958). 'The Cost of Capital, Corporation Finance and the Theory of Investment', *American Economic Review*, June, 261–97.

Morgan, D. P. (2002). 'Rating Banks: Risk and Uncertainty in an Opaque Industry', *American Economic Review*, 92: 874–88.

NIETO, M. J. and SCHINASI, G. J. (2007). 'EU Framework for Safeguarding Financial Stability: Towards An Analytical Benchmark for Assessing its Effectiveness', IMF Working Paper, WP/07/260.

RAJAN, R., SERVAES, H., and ZINGALES, L. (2000). 'The Cost of Diversity: The Diversification Discount and Inefficient Investment', *Journal of Finance*, 55: 35–80.

ROSENGREN, E. (2002). 'Comment', in Gaspar, V., Hartmann, P., and Sleijpen, O. (eds.), *The Transformation of the European Financial System,* Second ECB Central Banking Conference, October, 109–15.

SMITH, C. and STULZ, R. (1985). 'The Determinants of Firms' Hedging Policies', *The Journal of Financial and Quantitative Analysis*, 20: 391–405.

SOROOSH, J. and CIESIELSKI, J. (2004). 'Accounting for Special Purpose Entities Revised: FASB Interpretation 46(R)', *The CPA Journal*, 74: 30–7.

TUFANO, P. (2006). 'Comment', in Carey, M. and Stulz R., (eds.), *The Risks of Financial Institutions*, Chicago: National Bureau of Economic Research, 597–602.

WALTER, I. (2003). 'Strategies in Financial Services, the Shareholders, and the System: Is Bigger and Broader Bettter?', *Brookings/Wharton Papers on Financial Services 2003*, 1–36.

—— (2004). 'Conflicts of Interest and Market Discipline among Financial Service Firms', *European Management Journal*, 22: 361–76.

WOOLFORD, I. and ORR, A. (2005). 'The Limits to Hospitality', *The Financial Regulator*, 10, June, 41–6.

WRISTON, W. (1981). 'Testimony', Hearings before the Committee on Banking, Housing and Urban Affairs, US Senate, Part II, 97th Congress, 1st session, October 29, 589–90.

This page intentionally left blank

PART II

REGULATORY AND POLICY PERSPECTIVES

This page intentionally left blank

CHAPTER 9

CENTRAL BANKING

MICHEL AGLIETTA
BENOIT MOJON[1]

Introduction

Central banks perform several tasks. They provide settlement services to large-value payments, oversee banks for the sake of financial stability, act as lenders of last resort, and implement monetary policy. These tasks and their mode of operations have been repeatedly redefined in order to resolve specific monetary and financial crises. Actually, all major stages in the shaping of central bank functions have been responses to monetary or financial crises. The analysis of these crises and their competing interpretations is necessary to understand the functions that central banks need to incarnate and implement for a monetary economy to prosper.

The genesis of central banks as bankers' banks took place in nineteenth-century England. Two opposing theoretical conceptions of money: the currency principle and the banking principle, implied radically different roles for central banks. For the former, strict convertibility of money into a *special commodity* of which the

[1] We would like to thank Doug Evanoff, Philippe Moutot, and the editors for comments and suggestions. All remaining errors are ours. Benoît Mojon worked on this text mainly while working at the Chicago Fed. The text reflects the authors' opinion alone and neither the opinion of Banque de France, nor of the European Central Bank.

supply is independent of the government is an insurance against the secular manipulation of the unit of account to raise an inflation tax. The role of the central bank is to enforce convertibility of its bills into gold, as the Bank of England was assigned to in 1844. However, a repetition of liquidity crises in 1847, 1857, and 1866 demonstrated the need of flexibility in the supply of money.

Such flexibility is more consistent with the banking principle, whereby money is a *debt* that financial intermediaries endogenously issue as the counterpart to their credit operations. Yet, if the means of payments are debts issued by competing banks, payments between banks call for high-powered money to settle interbank transactions. This is precisely the role of the money issued by the central bank, money that the sovereign designates as legal tender for all debts.

However, the trust of the economic agents in central bank money cannot be imposed by law. It is essentially to preserve this trust that central banks have developed their functions. They supervise the banks to ensure the integrity of the payment system and prevent liquidity crises. In the event of liquidity crises, they stand ready to lend in last resort. They conduct monetary policies to stabilize the unit of account and thereby provide a nominal anchor to the economy.

The second section of this chapter explores the evolution of central bank attributes as a bankers' bank, relating them with the centralization of payments. The third section outlines how central banks have been committed to different monetary regimes from the gold standard to the present inflation targeting. It also points out the monetary doctrines underlying the different practices to deliver the nominal anchor that preserves trust in money. Finally, section four raises some prospective issues that may require further evolution of central banks in the twenty-first century.

The centralization of payments and the emergence of central banks as bankers' banks

The idea of a central bank—that is, a bankers' bank—was slow to emerge in monetary thinking. It was an offshoot of devastating financial crises that had become international in scope in the mid-nineteenth century. The expansion of industrial capitalism had intertwined credit networks, making contagion more virulent. In the crises of 1847, 1857, and 1866 the Bank Act of 1844, which had split the Bank of England in two departments—the issue department and the banking department—had to be de facto suspended, though no provision had been enacted to do so *de jure*. On the contrary, the Banque de France had

acknowledged responsibility to the financial system earlier. However, in 1868, when Crédit Mobilier was still recovering from the 1866 crisis, the Banque de France refused to discount its paper. The French central bank was involved in a rivalry between the Rothschilds and the Pereire Brothers, owners of the Crédit Mobilier. By taking the side of the Rothschilds to preserve the privilege of the *Haute Banque* against the emergence of modern commercial banking, the Banque de France failed to pursue the stability of the financial system as a whole in the circumstances, since it would have implied rescuing Crédit Mobilier.

The financial centers of the main European countries initially resented that the banks created to manage the public debt and to regulate currency should have superior status relative to other banks. Because bank money is a debt, it is the counterpart of credit. Because debts must be settled in other forms of debts, there is a hierarchy of debts and for that matter of the institutions that issue them. The central bank is the bank that issues the debt in which all other debts are settled. The hierarchy of the banking system, whereby the central bank issues the high-powered money that can be used for the settlement of interbank debts, appeared as a necessary condition for the integrity of the system of payments. The latter can allow means of payments issued as liabilities by competing commercial banks to coexist if three general rules are respected.

The first and foremost rule is the institution of *the unit of account*. In a decentralized market economy, market participants discover relative prices through nominal prices denominated in the unit of account. As long as it can be trusted over time, the monetary standard reduces transaction costs efficiently in avoiding offer prices to be announced in incompatible numeraires.

The second rule is that *issued debts can be made eligible to means of payments*—e.g. debts can circulate among third parties to redeem other debts and buy commodities. In a developed market economy, producers necessarily incur debts because they must buy resources, not least the human resource (pay wages), before they can sell their products. The quality of being accepted by third parties is *the liquidity of debts*, which in turn depends on the financial strength and the reputation of the issuer. The selection and verification of the liquidity of debts make the financial system hierarchical. Banks are financial agents specialized in the issuance of the most liquid debts.

The third rule is the one that makes the verification of debt liquidity a social process: *the settlement*. It is the process by which payments are made final—for example, by which any kind of debt used as a means of payment in any private transaction proves that it can be transferred against a unanimously accepted form of money. Depending on the definition of the unit of account, the ultimate liquidity in a payment system can be a commodity minted by the sovereign (or a foreign currency), or it can be the liability of a financial institution empowered by society as a whole or by its highest political authority, the sovereign. This institution is a central bank. It has become the paramount monetary institution.

How do the rules make a system? If the ultimate liquidity is the liability of the central bank, the unit of account is purely abstract.[2] It is the name given to the numerical unit of liability issued by the central bank. Such are the 'US dollar', the 'euro', the 'pound sterling', the 'yen', and the 'yuan' nowadays, to quote the most important world currencies. In that case all forms of money are issued as liabilities of a financial agent. The banking system is *hierarchical*. The central bank is the bankers' bank because its liability is the means of settlement of all commercial bank debts.

The law of reflux, multilateral clearing systems, and the emergence of central banks

Under free banking—that is, when there were no central banks—commercial banks could issue notes and open deposit accounts against their assets over and above their reserves of species. Notes and deposits were convertible on demand in gold coins. Convertibility (into gold or silver) was the rule that validated bank money. The *law of reflux* was the settlement mechanism whereby convertibility limited the issuance of competing bank monies. It saved the use of species, while, at the same time, it vetted the quality of bank issued notes. A *free banking payment system* without central bank was conceivable as long as settlement in gold coins could be made viable. The law of reflux effectively centralized the relationships between interbank correspondents within multilateral clearing systems. The question arises of the *modus operandi* of the law of reflux, both historically and theoretically.

Historically, such a system was successful in Scotland in the late eighteenth century, because banks were few and highly capitalized. It also operated for a long time in the US in the nineteenth century. Nevertheless, in the latter country, free banking was evolving under acute tensions. Solving settlement crises induced the development of institutions called clearing houses. They were incomplete surrogates of central banks, as far as settlement issues were concerned.

Banks were competing to increase their market share in the discounting of trade bills, as a counterpart to which they issued bank notes. The law of reflux implied that a bank that issued too many notes would face a liquidity shortage when the redemption of excess notes at par would be exercised. Theoretically, three mechanisms can be at work. First, the excess notes are immediately canceled when customers demand redemption in species and the bank loses an equal amount of reserves. Second, the note owners prefer to buy other banks' notes. Therefore, those banks increase their claims on the issuing bank in the interbank market—for example, the exchange market for bank notes, which leads to a net settlement

[2] It was not so when the ultimate liquidity was minted into a commodity, let us say gold. Debts used as means of payments had to be redeemed in gold coins.

demand in species after clearing. The destruction of the excess notes arises via correspondent banks. Third, the notes remain active balances for future expenditures. In the latter case, a multiplier effect of payments arises and diffuses the excess issue on more banks, as long as more individuals buy the bank notes they prefer. However, the pressure to redeem the notes will eventually be exerted via the former mechanisms, either public demand for species or bank clearing.

The law of reflux was the process emphasized by the Banking School to advocate regulation of the *quality* of money counterparts rather than the *quantity* of money. Since the law of reflux restrained automatically the banks from their loss of ultimate liquidity, the proper oversight was to check that banks discounted good assets. Only solvency might be problematic, while the interplay of the banks would assume the regulation of liquidity. In their own interest, banks would maintain a ratio of liquidity that balances the marginal cost of relative illiquidity and the marginal gain of asset building. Reconstituting the optimal ratio of reserves to notes triggered an adjustment toward balance sheet equilibrium. The payment system was theoretically viable. Macroeconomically, the optimal amount of bank money was determined by the desired demand for species and the general level of prices in terms of gold, which was actually largely independent from the banking system.

Such a system had functioned without a central bank. However, central banks progressively emerged as banks for banks in order to remedy two essential limits of self-organized non-hierarchical payment systems. First, the law of reflux with 100 percent redemption in specie on a bilateral basis is a straitjacket that becomes incompatible with the needs of trade in a growing market economy. The monetary system was under various form of stress when the law of reflux failed to discipline a bank.

In the case where bank notes dominate, the failing banks suffer a discount in the value the notes they issued. The consequence is a fragmented payment system levying crippling information costs on the economy. People always wondered in which bank money they should have their income paid, with self-fulfilling prophecies often destroying many banks in time of stress. The payment system was plagued by recurring runs on weak banks with contagion making up bank panics.

The case where checkable deposit accounts dominate is different because the means of payments (the checks) are dissociated from the liquidity registered on the account. This form of money became prevalent in the second half of the nineteenth century, when banking business went beyond discounting trade bills to meet the financing needs of industrial capitalism. Banks became intermediaries. On the asset side, they made illiquid credit requiring investment in specific information whose quality depositors were not able to assess. On the liability side, they offered non-marketable deposits combined with the provision of payment services. This asymmetric information structure, coupled with network effects in the payment system, implied that valuing deposits at par in unit of accounts and securing their convertibility at par into the base currency (e.g., into gold coins) made the most efficient contract.

However, in contrast with the mechanisms whereby bank notes dominated payment systems, pricing bank money in a crisis became impossible. The law of reflux could not work without an innovation in the technology of payments that allows a drastic saving of species. This innovation was *the multilateral clearing system*. It *ipso facto* differentiated the community of banks because some banks become clearing houses.

A clearing house is a centralized organization that introduces collective rationality into payment systems. Clearing houses appeared in the leading business centers of the US in the midst of the free banking era (1838–63). Free banking developed after the Second Bank of the US lost its federal charter in 1836.[3] It ended in 1863 while the Civil War was raging, when the huge increase of liquidity needs induced the creation of the Comptroller of the Currency to regulate the quality of money. Clearing and settlement of the banks that were members of the system were made on the books of clearing houses. Clearing houses issued settlement certificates on behalf of their members who deposited reserves with them. Multilateral clearing and net settlement saved a huge amount of specie and reduced the cost of check collection.

Moreover, clearing houses managed to insulate the execution of payments at the time of liquidity stringency by suspending convertibility into species. As Goodfriend (1988) pointed out, they acted as de facto surrogates of central banks. The higher status of settlement money gave the clearing houses hierarchical authority over their members. The responsibility of preserving the integrity of payments among the club of retail banks led the clearing houses to guarantee the irrevocability of payments in return for the banks' compliance with restrictive obligations. The system of the Suffolk Bank in Boston, Massachusetts, was the first to reach such a degree of centralization while it was evolving from 1825 to 1861.

The second limit that required the emergence of central banks appeared in the incapacity the of the large banks that acted as clearing houses to separate their profit maximization objectives from the interest of the community of banks participating in the clearing houses. In addition, in the case of the US, the clearing houses did not protect non-member banks that were located outside large financial centers from liquidity crises.

The National Bank Act of 1863 aimed at homogenizing the quality of money by instituting a Comptroller of the Currency and granting national charter to banks that respected stipulated obligations. Despite the extension of regional clearing houses, the US payment system remained vulnerable to bank panics until the paroxysm of the 1907 crisis. The financial debacle persuaded Congress to undertake a radical reform to provide the country with a single institution capable of securing

[3] The Second Bank was chartered in 1818 after the expiration of the First Bank of the United States. It was closed by President Andrew Jackson, who accused the Bank of political corruption and fraud. Jacksonian democracy, which leaned toward libertarian policies, favored free banking.

the payment system under a dual mandate: supplying elastically unique settlement money and assuming the responsibility of a lender of last resort. In countries where a government-sponsored bank already existed for a long time or had been enacted in the wake of political unification (Germany and Italy), they took over the role of clearing houses more smoothly. In England and in France, the former bank of the sovereign already had a special role in the banking system and this role evolved, in spite of the financial community's resistance, to the one of banker's bank.

Central banks and the regulation of payment systems in tranquil time

Since its emergence as bankers' bank, the central bank is the institution at the center of payment systems. It safeguards settlement, prevents systemic failures against operational risks ($50 billion of liquidity injected a single day in November 1985 to offset the destruction of liquidity due to a computer break at the Bank of New York) and imposes safety rules on banks that are members of large-value payment systems.

Commercial banks may pertain to different clearing systems linked to retail payments or payments of securities and other financial transactions (including the national currency leg of foreign exchange transactions). But the balances, resulting from those payment services, contribute to build interbank positions. They must register on the books of the central bank for multilateral clearing and settlement in central bank money. To honor their settlement obligations, banks with net negative positions must secure central bank money by all means available before settlement time, depending on the technique of settlement (net end-of-day, or gross in real time). They can draw on their reserve account at the central bank, borrow overnight money from surplus banks, or use repo facilities with the central bank against eligible collateral. All those devices connect the central bank to the channels of liquidity provision to the whole banking system. From this unique posture, the central bank can draw superior information on the situations of banks regarding liquidity ease or stringency. They can detect anomalies in the interbank market, as they did on 9 August 2007, when they observed a spike in overnight interest rates.

In the last thirty years or so, payment flows generated by financial transactions have swelled ominously. High-value payments concentrate risks that have a strong likelihood of becoming systemic. As a result, central banks have reformed interbank payment systems to deal with credit and liquidity risks in order to guarantee final settlement—that is, irrevocability of the payment for the beneficiaries whatever the situation of the payers. The central bank, which is the only player that can do so, guarantees unconditionally the payments it settles.

The path toward centralization of payments goes on unabated with the creation of a wider range of private means of payments. Contrary to a popular opinion,

innovation in payment systems does not undermine central bank leadership in the control of money. It reinforces central bank power because more complex payments, ever larger volumes and shorter lead times increase systemic risk. Such risk in payment systems threatens trust in money at its most fundamental level. Therefore the integrity of payment systems is much more than a merely technical issue. It must be seen as part of the broader objective of financial stability and the sustainability of the financial system.

The central bank in stressful time: the lender of last resort

Lending in last resort has been pinpointed the gist of the art of central banking. It was nonetheless long before being recognized as such. As early as 1802, Thornton had highlighted the responsibility of the Bank of England in supplying liquidity to sound banks in time of panic. But there was no follow-up to his advice. Recurring panics arose with ever more devastating outcomes. The Banking Act of 1844 worsened financial crises in making harder for the Bank of England to issue notes while losing gold reserves. De facto, the Bank Act was overruled with the suspension of convertibility in 1847, 1857, and 1866 without any complementary legislation. It permitted the Bank's banking department to expand its loans over the limit fixed by the gold stock in the issue department. But the easing in money markets was too late, too little. This unsatisfactory state of affairs prompted Bagehot (1873) to elaborate his doctrine that is defined below.

Indeed, lending in last resort is an extraordinary operation that violates market rules. It is a unilateral and discretionary decision to provide potentially unlimited amounts of the ultimate means of payment/settlement to the money markets. It escapes market contracts and is therefore an act of sovereignty that keeps afloat debtors who otherwise would have failed to settle their debts. This operation allows other perfectly sound liabilities to perpetuate, whereas they would have been destroyed by the spillover of the failed debts. Therefore the economic impact of LOLR interventions is ambiguous. On the one hand, it forestalls systemic risk, because the social cost of letting insolvency spread is higher than the private cost of the original failure. On the other hand, it can induce moral hazard if it strengthens reckless behavior against which it provides collective insurance. The purpose of Bagehot's principles and prescriptions was to stop contagion while keeping moral hazard at bay.

According to Bagehot, the lender of last resort is concerned about the overall stability of the financial system, not by the fate of any particular financial firm. It must lend without limit to solvent but illiquid firms, who cannot borrow in the market because the mistrust of potential lenders dries up liquidity. Insolvent institutions must be sold to new owners for what they are worth. However, these objectives require operational principles to distinguish intrinsic insolvency from threats of failure due to liquidity stringency.

Bagehot (1873) proposed as a distinctive criterion the quality of the collateral presented by borrowers. The central bank should accept the collateral and evaluate the solvency of banks at pre-crisis value. Furthermore, to better safeguard against moral hazard, Bagehot insisted that the central bank should lend at punitive rates. This disposition would be both a risk premium for the central bank and a deterrent for borrowers. Finally central bank interventions in last resort should be made unpredictable. This is the constructive ambiguity that central bankers are fond of. It is an attribute of the radical discretion that is the essence of sovereignty.

Nowadays, banks have to comply with capital regulation and to accept the ongoing supervision of activity by either the central bank or an independent supervision authority. These requirements are counterparts of the 'insurance' that the central bank would provide, as LOLR, to insulate banks that have not taken excessive risks from crises that threaten the integrity of payments.

Failure to lend in last resort can have the most dramatic consequences. A prominent example is the Great Depression. The Wall Street Crash of October 1929 had led to a scramble for liquidity. By year end the deflation in equity prices had been communicated to primary commodities and durable goods industries. The Federal Reserve lowered the discount rate from 6 percent in August 1929 to 2.5 percent in June 1930. But the money stock continued shrinking unabated. According to Friedman and Schwartz (1963), it should have undertaken blanket open-market operations to avoid the seizure of the credit markets. After mid-1930, the financial crisis changed in nature and in magnitude. Three waves of extended bank failures, one every year, wrecked the banking system completely, leading to the Bank Holiday in March 1933. A drastic change in regulation severed commercial banks from financial markets.

With the comeback of financial crises in deregulated financial systems, the LOLR has returned to fashion since the Penn Central failure in the US in 1970 and the secondary banking crisis in the UK in 1972. Since that time there have been innumerable banking and financial market crises worldwide that have solicited the interventions of central banks in last resort. There has been a diversification of such central bank interventions, ranging from securing the payment systems to, in recent history, spectacular interventions by the Federal Reserve System to restore confidence or attempt to limit the chances of financial crises before they occur.

The first type of intervention is illustrated by the response to the breakup of communication lines in interbank payment systems after the events of 9/11 2001. The Fed injected liquidity massively both through the Fed Funds market and the discount window. Without this emergency supply of overnight money the interest rates would have gone through the roof. On the contrary, the overnight rate of interest fell almost to zero, which is an indication that the intervention was indeed unlimited. Each day for a whole week the Fed injected between $36 billion and $81 billion against a daily average of $5 billion in normal times. Other central banks

acted in concert and emergency swap agreements were concluded between central banks in the main financial centers of the world.

Restoring confidence in distressed financial markets is a preoccupation well illustrated by the circumstances following the Long-Term Capital Management (LTCM) episode in September and October 1998. Since the end of August, the shock wave of the Russian financial crisis had disoriented credit markets. Spreads had spiked unexpectedly, taking unawares hedge funds that had betted a reduction in spreads. Losses surged and the huge hedge fund LTCM, which was heavily leveraged, became insolvent. In an already stressed environment, confidence broke down in all segments of credit markets. The flight to quality concentrated on Treasury Bills. By the end of September, private borrowers could not find any credit (Scholes, 2000).

The central bank was confronted with a dual problem: the direct impact of the LTCM debacle on the bank creditors, on the one hand, and the general flight to liquidity, on the other hand. To solve the first, it was necessary to consolidate LTCM debt. The second was a problem of mass psychology: how to re-establish trust in the midst of universal mistrust? The New York Fed was the coordinator in LTCM's rescue. It organized a bank consortium that took charge of the Fund's management to pilot an orderly reduction of its indebtedness and proceeded with a $3.5 billion debt-equity swap. To restore confidence, the Fed decided on three interest rate cuts, each one of twenty-five basis points on 29 September, 15 October, and 17 November. Fully anticipated, the first one had no impact. It even deepened the crisis which reached the foreign exchange market with the surge in the yen on 8 October. The second one was crucial. Taken outside scheduled Federal Open Market Committee meetings, it was a complete surprise: an act of sheer sovereignty to handle an extraordinary situation. It struck market sentiment and had a dramatic effect. As long as they remained caught in the psychology of contagion, market participants were unable to price assets. They were obsessed with the immediate liquid value of their claims. Liquidity evaporated under the one-way selling pressure of asset holders who no longer had any confidence in the floor price of their securities. The sovereign decision of the central bank anchored the floor price of short-term securities. It reinstated the benchmarks necessary for market evaluation of differentiated risks. The third intervention was a message of confirmation. It reassured the financial community in the conviction that the central bank was determined to provide all the liquidity necessary to permit the well-functioning of financial intermediation.

Innovation in last resort lending was spectacularly pursued in since August 2007 after a global credit crisis in securitized market struck, following and propagating a surge of insolvency in the US subprime mortgage market. At the time of writing this text, it would be very presumptuous to assess the effectiveness of the 2007–9 financial crisis management. The crisis and the scale of public interventions have escalated in parallel since August 2007. Moreover, it goes beyond the scope of this

chapter to give a detailed account of the crisis escalation and the series of central banks and fiscal authorities to avoid the collapse of the banking system. Box 9.1 lists the major steps taken by the Eurosystem and the Federal Reserve System from the beginning of the crisis to December 2008 when this text was completed. Central banks themselves have recorded and will continue to publish comprehensive descriptions of their initiatives and innovations in handling the crisis.[4]

Our inquiry into central bank policy as a bankers' bank leads to monetary policy. It highlights the consistency in all aspects of central bank behavior due to its pivotal position in the monetary system. The roundabout dynamic between credit expansion and asset price appreciation in global financial systems entails a macroeconomic risk that is both endogenous and pro-cyclical. The central bank is the sole institution able to handle it.

Nominal anchor and monetary policy

The most well-known function of central banks is to conduct monetary policy. The objectives of monetary policy can include price stability, sustainable growth of output, full employment, and financial stability. The monetary policy legal mandates of central banks (Maastricht Treaty for the European Central Bank) may specify that price stability is the primary objective or that sustainable growth is a side benefit of price stability, but they don't always do so (as in the case of the US Federal Reserve Bank Act).

The monetary policy function of central banks relates directly to the *unit of account* attribute of monies. Economic agents engage in current and inter-temporal exchanges at nominal prices—that is, at prices expressed in terms of the unit of account. Monetary policy, therefore, consists of issuing money in a quantity that would stabilize the value of the unit of account, hence avoiding both inflation and deflation of the general price level. The central banks aim at providing a nominal anchor for economic agents to set prices in their current and in their planned transactions.

Monetary policy doctrines

Monetary policy, concerned with the money supply and the value of the unit of account, has always been the object of intense debates among economists and

[4] See in particular Committee on the Global Financial System, paper 31: 'Central Bank Operations in Response to the Financial Turmoil'.

Box 9.1. The Major Steps taken by the Eurosystem and the Federal Reserve System from the Beginning of the Credit Crisis to December 2008

The central banks have taken a number of unprecedented steps to tame the financial turmoil. Central banks worldwide have extended their lending facilities and widened the range of collateral they accept. They modified their lending facility to fixed rate tenders with full allotment. They substituted the decentralized money markets that ceased to function by providing liquidity to banks on a bilateral basis. In these processes, they took on their balance sheet larger and larger amounts of risky assets. The Fed balance sheet has increased from US $891 billion in December 2007 to US $2,311 billion in December 2008. Moreover, the credit risk of the financial instruments piled in the Fed's assets has deteriorated drastically. Treasuries, which made up to 86 percent of the Fed's assets at the end of 2007 now represent only 20 per cent of it.[5] The other 80 percent is composed mainly of term auction credit, commercial paper, and foreign currency counterparts to the dollars lent by foreign central banks in agreement with the Fed as initiated when the TAF was introduced on 12 December 2007. This unprecedented asset structure implies a credit risk that may make the central bank technically insolvent, and therefore at the mercy of the fiscal authority.

Central banks also have repeatedly coordinated their crisis management actions (lowering interest rates together on 8 October 2008, extending the maturity of their liquidity provision, setting up currency swaps to extend the provision of dollars outside the US). This coordination of 'LOLR' operations is, however, not new. Financial markets are typically integrated internationally and more typically so in times of financial stress. In the current crisis, the banks' distress has happened or spread across major financial markets at each and every stage of the crisis. This echoes, for instance, the 1907 crisis resolution when the Banque de France lent gold to the Bank of England to allow her to provide enough emergency liquidity to the US banking system. While not new, this stresses once more that the contagious nature of financial crises eventually requires the coordination of central banks in crisis management.

Last but not least, they slashed interest rates in order to facilitate the de-leveraging process now engaged in by financial intermediaries, households, and firms. It is meant to avoid a 'credit crunch' in the downward stages of the financial cycle. It was spectacular both after the turning point of the real estate bubble in 1991–3 and the stock market bubble in 2001–3. In both cases, US short-term interest rates were driven much lower and for much longer than the easing in monetary policy that would have been warranted by the arbitrage between the medium-run objectives of anchoring inflation expectations and keeping economic activity close to potential. The stance of monetary policy was motivated by concern about the macroeconomic impact of financial distress. Risk management was then the primary objective of US monetary policy. Turning to the 2007–9 crisis, it is too early to assess the effectiveness of the changes in the monetary policy stance on economic activity.

[5] Effectively some of the US $680 billions distributed in the context of the Term Auction Facility by other central banks who have an SWAP agreement with the Fed may also use Sovereign debt as collateral.

commentators of public policies. Throughout history, the main axis of division in this debate opposes those who consider that money is essentially exogenous and those who consider that money is essentially endogenous. For the former, *discretion* in the supply of money should be avoided because it tends to be used by the political power to raise an inflation tax that spoils the people and destabilizes the economy. For the latter, *rules* of money supply, such as strict convertibility into species, lack the flexibility to accommodate changes in the money needed for the economy's growth. Interestingly, the doctrine underlying the current consensus that dominates the conceptualization of monetary policy and of which a prominent example is inflation targeting, claims that it strikes a balance between rules and discretion (Goodfriend, 2007; Woodford, 2003; and Bernanke and Mishkin, 1997).

The *quantity theory of money* is a milestone of the debate on monetary policy. First expressed by Cantillon (1755) and Hume (1752), the quantity theory implies that increases in the supply of money are eventually reflected in higher prices with no effects on output. An important consequence of the quantity theory is that monetary policy should strictly focus on price stabilization because manipulating money supply can only affect prices. In line with the quantity theory, the Currency School in nineteenth-century England considered that only strict convertibility of bank notes into gold would prevent over-issuance of notes, and inflation. A century later, while the straitjacket of strict convertibility into gold was no more an essential feature of the monetary system, Milton Friedman and other monetarists argued that the supply of money should follow strict rules—for example, increase the money supply at a pre-announced, k-percent rate. Authorities should not use monetary policy to fine tune the business cycle, because, although money was not neutral in the short run, the transmission from changes in the money stock to output and prices 'took long and variable lags'. As a result, fine-tuning policies paradoxically risked introducing volatility, in complete opposition to their objective.

These views have been opposed by economists who believed that the money stock would evolve endogenously in response to changes in the liquidity needs of the economy. An important consequence of the endogenous character of money implies that strict rules of money supply can abruptly curtail transactions and growth. Monetary policy authorities should therefore have *discretion* in supplying liquidity.

To start with, the Banking School advocates considered that the bank notes in circulation were secured by their counterparts on the asset side of a bank's balance sheet. The law of reflux would warrant that, as credit (at the time, Merchant's bills) was reimbursed to banks with bank notes, over-issuance was prevented. In addition, the major flaw of strict convertibility lay in the mismatch between the amount of money needed for economic growth and the stock of metal available for minting money. With neither flexible nominal prices nor a rapidly adjusting velocity of money, the nineteenth century saw a repetition of monetary crises where

the shortage of metal limited the expansion of output. These crises pointed to the need to free money supply from the corset of strict convertibility.

Wicksell (1907; 1935) was the first to introduce the notion that the supply of money should depend on the rhythm at which production capacity grows. The money interest rate, set by the monetary authority, ought to be equal to the real interest rate, which itself reflects the expected return on newly produced capital goods. This approach of monetary policy can avoid both over-expansion of money, credit and, henceforth, inflation, and a contraction of credit and deflation. The central bank should accommodate the private sector demand for liquidity at the chosen level of money interest rate. It is easy to understand that, in this conceptual framework, non- contingent rules of money supply turn out to be destabilizing because the real rate can change over time with economic circumstances.

However, it took another two major monetary crises, in the 1930s and in the 1970s before the Wiksellian approach became the dominant monetary policy doctrine. The notion that monetary policy (as well as fiscal policy) should be articulated in order to stabilize the business cycle became popular largely because of the trauma of the 1930s Great Depression, and its interpretation by Keynes.[6] From World War II to the mid-1970s the common wisdom had been that monetary policy (and fiscal policy) may have to stimulate demand and let inflation increase so that real wages would permit full employment. In the event of a slowdown of the business cycle, monetary and fiscal policies should stabilize output by exploiting the trade-off between inflation and unemployment (what became known as 'the Philips curve').

However, the experience of the 1970s Great Inflation discarded the fine-tuning policies of Keynesian inspiration. Besides, in spite of their success in stabilizing inflation in Germany and Switzerland, monetarist approaches (e.g., targeting a fixed rate of growth for money aggregates) appeared difficult to generalize because of the instability of money demand.

On the conceptual front, the increasing importance of expectation formation in the analysis of macroeconomic policies led the Rational Expectation School to argue that monetary policy may actually be neutral even in the short run. Kydland and Prescott (1977) demonstrated that, as long as agents believe that the central bank would try to exploit an inflation–unemployment trade-off, the economy would converge to higher inflation because the expectation of eventual stimulating monetary policy would induce higher wages and prices, in order for agents to preserve their purchasing power. Issues with the credibility of the central bank's anti-inflation commitment led economists and policymakers to consider that independence of central banks from governments could be desirable. However,

[6] Keynes' interpretation of the Great Depression was exactly opposite to the one of the monetarist. He stressed in particular that, beyond a certain threshold, increasing the money supply would have no effect on the level of interest rates (the liquidity trap). His policy recommendation was instead to manage demand through active fiscal policy.

this independence, which freed central banks from the influence of elected governments, also called for the development of monetary policy frameworks that would enhance their accountability.

These considerations led to a new consensus of monetary policymaking, which ingredients include a credible commitment to low inflation, flexibility of money supply in the pursuit of this objective, and an effort of transparency in the communication of monetary policy decisions. This consensus is exemplified by, but not exclusive to, *inflation targeting*.

An inflation targeting central bank announces a target level for inflation and engineers the monetary policy that would drive inflation near this level. The inflation target is either a point or a range that sets a low and positive level of inflation for a given consumer price index, and the horizons vary, across countries, from a couple of years to the business cycle or indefinite. This pre-announcement helps anchor inflation expectations and provides a benchmark against which the central bank can be held accountable.

Inflation targeting has been portrayed as a compromise between rules and discretion. Bernanke and Mishkin (1997) actually used the term '*constrained discretion*' to describe the monetary policy strategy of inflation targeting. They argue that commitment of the central bank to keep inflation near a pre-announced inflation target provides a nominal anchor for economic agents. The latter can therefore engage in nominal contracts even at relatively long horizons—for example, a mortgage interest rate, with a fair assessment of the real purchasing power of future flows of income and payments. The inflation target is also a benchmark to evaluate the monetary policy performance of the central bank. It is a discipline device that should prevent—and so far that has prevented—the inflation bias allegedly inherent to unconstrained discretion in the conduct of monetary policy.

However, an inflation target does not provide prescriptions for money supply. The latter has to be decided by the central bank with the aim of keeping inflation close to its target. A good framework to analyze such monetary policy decisions is to consider a benchmark interest rate rule, whereby the central bank increases the real interest rate proportionally to deviations of inflation from the inflation target in order to weigh on demand and bring back inflation to the target. Likewise, one can conceive that the central bank should increase the interest rate when it observes tensions on prices, as measured, for instance, by an output gap—that is, the demand-supply gap. Taylor (1993) has actually showed that a contingent interest rate rule, such that the real interest rate increases equally to inflation and to the output gap, provided a good model of effective monetary policy in the US. This framework is also fully consistent to a Wicksellian approach to monetary policy, where the central bank sets its monetary instrument, the interest rate, in reference to a neutral interest rate (Woodford, 2003).

Such lean 'against-the-wind' state-contingent policy rules provide benchmarks to evaluate the stance of monetary policy by comparing the interest rate to a

contingent hypothetical rate that is fully consistent with aiming at bringing inflation back to its official target level. Moreover, interest rules that depend on the output gap, and conceivably other indicators of tensions on prices, can indicate whether the current stance of monetary policy tends to counteract these tensions or, on the contrary, accommodate them (Woodford, 2003).

However, inflation targeting is better described by a framework for the conduct of policy than by a strict rule. For one, inflation targets are often defined as ranges and it is generally understood that point targets are indicative of the region where inflation should be, approximately. Inflation can be expected to be close to the target, on average on the business cycle and as close as possible to the target, but it can deviate temporarily because of specific supply shocks—for example, to energy or food prices. In addition, inflation targeting does not provide a strict operational rule (Goodfriend, 2007). The central bank can adjust the stance of monetary policy and the rate of growth in the money stock to accommodate either changes in the velocity of money or unexpected shocks that could harm the other objectives of monetary policy, such as stabilizing output and employment at their maximal non-inflationary levels. Finally, the central bank ought to acknowledge the uncertainty of the environment in which it operates. This uncertainty may entail temporary deviations of the monetary policy stance from the one required under a baseline scenario to prevent the risks of less likely outcome, typically a crisis in financial markets, which occurrence would imply prolonged economic unrest. This risk management approach to monetary policy, which has been formulated by Alan Greenspan (2004), has been particularly useful to describe the reaction of monetary policies to changes in the economic outlook that are outside the scope of standard macroeconomic models. In particular, the financial crises that we discussed in the previous section led central banks to alter the path of interest rates to restore confidence on financial markets. The risk management doctrine clarifies that such loosening of monetary policy stance is temporary and would not deter the central bank from its objective of price stability, defined over the longer run.

The success of the new monetary policy consensus seems remarkable in so far as muting inflation[7] did not entail higher variability of output and employment. However, one decade is a relatively short period in which to assess the robustness of this approach to monetary policy as a guard against future monetary and financial crises, a point we come back to, in light of the 2007–9 crisis, in the last section of this chapter.

Although this does not have a core role in the doctrinal debates about monetary policy, it should be stressed that, in the real world, a vast majority of central banks enforce convertibility into the currency that is central to a region or to the world. It is therefore essential to make a distinction between the central bank that dominates

[7] For instance, the Bank of England announced an inflation target of 2.5% in 1992. UK inflation was brought from 5% in 1991 to 2% in 1993 and close to its current (at time of writing) 2.5% since then.

the international monetary system and the vast majority of central banks on the periphery. Metallic standards also constitute monetary policy regimes where the value of the higher-powered currency is set externally, in the interplay of supply of and demand for the reference metal.

For central banks at the periphery, the value of the currency remains largely defined in terms or an external object over which the central bank has little control. The pound sterling, the US dollar, or the basket of currencies chosen to peg the domestic currency may change value similarly, although for entirely different reasons, as precious metals have at the time of the gold standard and earlier. Hence, the value of the domestic currency hinges on the country's ability to maintain enough reserves in the center's currencies to sustain the credibility of the peg.

The implementation of monetary policy has also evolved with major innovations in the technology of payments and of financial instruments more generally. This evolution starts centuries before the emergence of central banks as the institutions in charge of monetary policies.

Control of the unit of account in earlier times under metallic standards

For centuries Europe's monies worked under dualist systems. Units of accounts were separated from coins in use. Defined in old coins that no longer circulated, they became abstract units. Dualist systems were established in which the sovereign could change the value of the unit of account in terms of the galaxy of coins without having to alter the latter.[8] Whenever the king devalued the unit of account, he increased the purchasing power of the coins in use, because prices were slow to adjust. He did so to increase the money supply, but also to alleviate the burden of the public debt that was denominated in the unit of account. Monetary conditions were highly dependent on the availability of metal.[9]

In the seventeenth century, the nation states' eagerness to build large-scale factories required the immobilization of savings in long-run investments. However, when the unit of account depreciated, hoarders of species gained at the expense of creditors holding nominal claims. In England, silversmiths speculated on the recurrent devaluations of the pound sterling and on the debasement of species. They aggravated the monetary chaos by exporting the best coins. The Glorious

[8] In France, the 'livre tournois' goes back to Charlemagne around 800. In Great Britain, the 'pound sterling' originates in a Norman silver penny brought in with the conquest of 1066.

[9] For instance, the dualist system worked to mitigate the destructive forces of deflation in the terrible era that encompassed 150 years from the Black Plague of 1348 to the end of the fifteenth century. However, in the sixteenth century, the inflow of silver from the Potosi in Peru launched a long inflation exacerbated by the manipulation of the units of accounts that the sovereigns indulged in their rivalries for the supremacy in Europe.

Revolution of 1688 promoted a drastic change spurred by the need to finance the 1689 War of the Grand Alliance against France. In 1694, the merchants created the Bank of England and lent its entire capital of 1.2 million pounds to the king, William III (William of Orange). In return, the Bank was granted the right to discount bills and issue notes that later became legal tender.

However, in 1694–5, an inflationary spike was perceived as a bad omen for the acceptance of the Bank of England's notes. A hot debate raged about the means to restore trust in money. On one side, William Lowndes, Chancellor of the Exchequer, advocated another devaluation of sterling. On the other side, Locke pleaded for a complete re-coinage, which would entail getting rid of debased species entirely. This deflationary solution was finally accepted by the king. It entailed a terrible recession in 1697–8 and provoked a huge loss for the Crown. Nonetheless, the chosen ratio of gold to silver was 15.9 against 15 in Continental Europe. It attracted species from abroad and put the newly founded United Kingdom of Great Britain on a de facto gold standard. The dualist system was replaced by the convertibility rule that was only suspended during the wars against France, starting in 1797 and lasting beyond the Vienna treaty of 1815 to the complete monetary recovery in 1821.

The gold standard became an international monetary order much later. Meanwhile, gold and silver coexisted as long as Central Europe was on silver and France was the bi-metallic center of the system. It was not until 1871 that the new German Empire adopted the gold standard. Not long after, France decided to abandon silver coinage and the US to redeem the greenbacks issued in the Civil War in gold. In 1880, the world was on the gold standard.

The international monetary stability that prevailed until World War I is another example of the importance of hierarchy in payment systems, though this time at the international level, with a central role for the Bank of England. The gold standard was in essence a *key currency* system legitimated by gold convertibility. Sterling bills of exchange were the universal means of payment in international trade, while long-run capital exports from the City were negatively correlated with investment cycles in the UK. Since the rule of convertibility was everywhere considered as an intangible common good over national policy objectives, short-run capital flows were stabilizing. Banks all over the world held deposits in London because they discounted sterling bills and received sterling payments. Therefore sterling was *primum inter pares*. In handling its rate to keep the ratio of gold reserves to notes close to the required level, the Bank of England *ipso facto* regulated international liquidity, because all other countries kept their exchange rates against sterling within gold points.

The working of convertibility was so entrenched in the minds of people that trust in the nominal value of contracts was never shattered. In times of stress, such as the Baring Crisis of 1890 and the Bankers' Panic of 1907, ad hoc cooperation between the Banque de France and the Bank of England, in the form of gold loans by the former to the latter, helped to build up international rescues that retrieved confidence.

The City of London centralized the market for international bills of exchange. The bank rate had a paramount influence on other countries via the discount houses in London. It is why the financial conditions in London summed up the degree of tightness in world liquidity. Liquid balances of foreign banks in London were highly sensitive to the bank rate, which triggered stabilizing capital flows and synchronized the business cycle.

Monetary policy implementation in purely fiat monetary systems

The implementation of monetary policy since World War II is fundamentally different in the sense that money has effectively lost its physical, metallic reference. The high-powered money issued by the central bank has become purely fiat, with the suspension of convertibility of dollars into Gold in 1971. However, even before 1971, the convertibility into gold has had only a very marginal role in the conduct of monetary policies. The gold standard had been replaced by a gold exchange standard in the 1930s. Under the *exchange* standard, only monetary authorities could exchange gold for currencies among themselves. The system was put under pressure in the 1960s as financial markets became progressively convinced that the US treasury could not sustain an artificially low peg of dollars into gold. Bickering arose on both sides of the Atlantic while European central banks had begun accumulating excess dollar reserves. In 1965, President of France Charles de Gaulle accused the US of buying French assets cheaply and ordered the Banque de France to sell dollar reserves against Fort Knox gold at the official rate of $35 an ounce.

The growing abstraction of money takes the form of new means of payments and savings instruments that become nearly as liquid as the more traditional deposits that are used for payments. These evolutions led to the definition and the measurement of several monetary aggregates (e.g., M1, M2, and M3). This multiplication of operational definitions of money reflects the increasing difficulty of identifying the relevant set of financial instruments that best reflect the liquidity available in the economy.

Central banks use mainly *three instruments* to influence the pace of money creation in the economy. *Reserve requirements* are a first type of instrument. By law, central banks stipulate that all banks keep a fraction of their balance sheets (typically a specified money aggregate) in reserve at their account at the central bank. The reserve requirement ratio can be used to affect the cost of issuing deposits.[10] Given that the central banks often choose to pay no or low interest on

[10] A prominent example of such implementation of monetary policy is the ongoing increase of required reserve ratio in mainland China. The People's Bank of China increased its reserve requirement ratio nearly every month in 2007 in order to contain the expansion of credit and deposit, while at the same time limiting increases in interest rates.

the reserve, the cost of issuing any liability subject to reserve is directly affected by the level of reserve and their opportunity cost.

The second type of monetary policy instruments are *standing facilities* (also called 'discount windows') for banks to obtain liquidity in a bilateral transaction with the central bank. The interest rate of these transactions is typically called the *discount rate*. This rate can be superior to the money market interest rate so that these standing facilities are essentially insurance in case of unforeseen liquidity shortages. However, the discount window has been and still can be the foremost channel of central bank liquidity to the banking sector in countries where decentralized money markets are not mature.

The last major instrument of monetary policy is *open-market operations*. These operations take different forms. The European Central Bank organizes regular auctions where it provides reserve at a target interest rate. The Federal Reserve directly purchases and sells public sector securities against central bank money so as to achieve a certain overnight interest rate.

These three instruments are used to a various degree to control money supply. However, open-market operations are not usually described in terms of the interest rate they are meant to achieve, rather than in terms of the resulting level of monetary aggregates. One reason for this development is that the unpredictable rhythm of financial innovation alters the link between money growth and inflation. Another is that the control of monetary aggregates may require volatile short-term interest rates, as experienced in the US between 1979 and 1982. And, targeting larger monetary aggregates, while less conducive of financial instability, largely reflects the evolution of deposits, on which the central bank has a much looser grip.

Effectively, the level of the target interest rate has progressively become the dominant operational instrument of monetary policy. While the amount of liquidity exchanged by central banks is relatively small in comparison with the overall amounts of debt securities, the monopoly of the central bank over the supply of monetary base guarantees that the overnight interest rate rarely deviates from the target interest rate of the central bank for more than a day. Hence, at the frequency that is relevant for production, consumption, and financial planning—that is, over months or years—the central bank does control the level of short-term interest rates. Moreover, because this control of the cost of liquidity is ongoing, the central bank both controls the short end of the yield curve and influences the full maturity spectrum through the market expectations of future short-term rates.

Central banks therefore need to pick a level for this interest rate and explain the reason for this choice to market participants and the market at large. The conceptual framework used to decide on the level of short-term interest rate and the supply of central bank money is precisely the object of the monetary doctrines above discussed.

The last fifteen years have seen a large convergence in the conduct of monetary policy. Explicit inflation targeting was first introduced in New Zealand, Canada,

the UK, and Sweden in the early 1990s. More than twenty countries have adopted it since (see Crowe and Meade, 2007 for a list). The European Central Bank, that conducts monetary policy for, at the time this article is published, fifteen European countries, has a quantified inflation objective that has a level close to the inflation targets of inflation-targeting countries. The case of the US Federal Reserve Bank stands out because it has a dual mandate of price stability and full employment. However, some observers have argued that US monetary policy can be considered to aim at keeping inflation at a small positive level (Goodfriend, 2007).

Central banks announce and explain changes in the level of the short-term interest rate with reference to inflationary and deflationary pressures that follow from the degree of tensions demand puts on the productive capacity of the economy. Changes in the price of liquidity are hence typically associated with the risks of both inflation and economic activity, though the path of inflation is always sustainable at the low level of the explicit or implicit inflation target. This approach to monetary policy succeeds in providing a nominal anchor, although money has become a purely abstract concept that, potentially, can grow or shrink without limit. Central banks tie the price of liquidity, the nominal interest rate, through a state contingent rule, to the degree of tensions on the economy's productive capacity. An explicit nominal anchor is pursued, and, to a very large extent, achieved, through tailoring the money supply to the economy's changing need for liquidity.

The challenges of 2007–9: Financial crisis and beyond

This conclusion focuses on the current financial crisis. There are, however, other important challenges. First, means of payments have been evolving for a long time—from coins, to notes, checks, plastic cards, and electronic transfers. E-money poses several new forms of risk because it uses open networks to communicate instructions and transfer value in contrast to the closed nature of the interbank market. Open networks can remain efficient only if they maintain a critical mass of users, while they may collapse suddenly if would-be users anticipate that this threshold will not be reached. In that event, losing customers trigger a self-fulfilling flight from the network. Chaos could then spread through interconnection between networks. The control by the central bank of the degree of liquidity in the economy would be severely hampered if issuers of e-monies were freed from the reserves and supervision requirements that apply to banks. Second, while there is broad agreement that monetary policies have contributed to general stabilization

of inflation since 1995, this remarkable performance may also have resulted from globalization. The worldwide organization of production may indeed diminish the traditional bottlenecks that translated tensions on local factors of production into local inflation pressure. As a consequence, central banks may need to reconsider the geographic base of supply and demand in analyzing inflation tensions and envisage international coordination of monetary policies. As we have shown in this text, structural changes, technological developments, and the occurrence of new crises constantly raise new challenges for central banks. We therefore now turn to what we consider to be major threats for central banks at the current juncture.

While we are still, at the time of writing this text, in the midst of the largest financial crisis since the 1930s, it is essential to realize that the extension of the LOLR interventions taken by central banks since August 2007 have avoided the collapse of the banking systems. In our view, this shows that central bankers have learned from past financial crises that such a collapse should be avoided at all costs. In spite of the scale of the financial crisis, we can still reasonably expect that the de-leveraging phase now engaged would trigger 'only' the most severe recession since 1945, with unemployment rising 'only' by 2–4 percent, not so much higher than the levels visited in the early 1990s, and nowhere near the Great Depression levels (when unemployment reached 25 percent in the US).

However, the strength of the de-leveraging tide should not be underestimated. It encompasses bank and non-bank financial intermediaries, households, and non-financial firms. Monetary authorities should in particular seek to avoid that a self-fulfilling debt deflation spiral (Fisher, 1933) takes off. Given the willingness of the private sector to downsize its leverage, we risk an increase in precautionary savings that would put downward pressure on prices. This in turn can reduce income and increase real interest rates to a point where a larger proportion of income is used to reduce debts. At a point, bankruptcies may become the only way out of debt.

On the income side, demand can be stimulated either via income transfer to households with the highest propensity to spend (i.e., low-income households) or through public investment programs, preferably with a positive effect on long-term productivity (i.e., infrastructure and research). On the debt dynamic side, public policies should aim at lowering the effective interest rates on private debt—if possible, to below the growth rate of nominal income. Otherwise, the de-leveraging process can be prolonged indefinitely, or work through bankruptcies, with dire social costs. We believe that these principles have inspired the design of the current crisis management policies and we are confident that further steps in these policies will be taken if warranted by the evolution of the crisis.

Looking further into the future, the current financial crisis constitutes dramatic evidence that the economy remains vulnerable to the self-fulfilling dynamic of credit expansion and asset price appreciation. Against this background, the single focus of monetary authorities on the stability of goods and service prices, while asset prices have become widely volatile, may put them off the mark of their objectives of

nominal, financial, and real stability. More generally, economists and central banks have to identify whether taming asset price volatility is desirable, whether it can be tackled by traditional supervision policies only, or whether it should also become an objective of monetary policy. In the last case, the design of an operational framework that would take asset price volatility into account is warranted.

This aggiornamento of the central bank missions should improve the economy's resilience to the destabilizing impact of financial innovations. The last waves of financial crises (around 1990 and in 2007–9) have all been fueled by excessive credit growth, which itself relied on some form of financial innovations. In the case of the current crisis, securitization and credit risk transfers have seriously hampered the effectiveness of bank supervision, while it led to a dramatic increase in leverage of both financial intermediaries and households. This leverage increase has been located in part in a 'shadow banking system', which escaped capital requirements of bank regulations. For instance, the leverage of GSEs (Fannie Mae and Freddy Mac) have become higher than 60—that is, five times as high as the maximum leverage tolerated for commercial banks.

We can reasonably expect that future bank regulation will increase its perimeter to include shadow banks. However, the historical perspective reminds us that the next wave of financial innovation is likely to facilitate yet another way to increase leverage in the next optimistic phase of the asset price/economic cycle. This is why we may need to establish principles of macro-prudential policies that would set up contingent policies to counter the acceleration of credit growth away from its long-term trend. Such policies would lean against the wind of the credit cycle.

In case of a broad agreement in favor of assigning central banks the objective of fighting this source of financial instability, economists will have to design the appropriate policy instruments. Indeed, as exemplified by the current crisis, financial instability may arise in spite of an impressive record of price stability. Moreover, the interest rate can be powerless to counter speculative dynamics, while, at the same time, very costly for non-speculative investment.

A second instrument, such as, for instance, contingent capital adequacy ratios may be necessary to allow central banks to pursue both objectives. Possibly, central bankers will come up with a more desirable solution to limit financial instability. In any event, financial and monetary crises have repeatedly transformed the mandate of central banks and the 2007–9 crisis should be no exception.

References

Aglietta, M. (1992). 'Genèse des banques centrales et légitimité de la monnaie', *Annales Économies, Sociétés, Civilisations (ESC)*, 3: 675–98

Bagehot, W. (1873). *Lombard Street: A Description of the Money Market*, London: H.S. King.

BERNANKE, B. (2003). 'Constrained Discretion and Monetary Policy', Remarks given on February 3, 2003, www.federalreserve.gov.

—— and MISHKIN, F. (1997). 'Inflation Targeting: A New Framework for Monetary Policy?', *Journal of Economic Perspective*, 11: (Spring), 97–117.

CANTILLON, R. (1755). *Essai sur la nature du commerce en général*, ed. H. Higgs (repr. New York: A. M. Kelley 1964).

Committee of the Global Financial System (2008). 'Central Bank Operations in Response to the Financial Turmoil', GFS papers 31, Bank for International Settlements, July.

CROWE, C. H. and MEADE, E. (2007). 'The Evolution of Central bank Governance around the World'. *Journal of Economic Perspective*, 21: (Fall), 69–90.

European Central Bank (2004). *The Monetary Policy of the ECB*, www.ecb.int.

FISHER, I. (1933), The Debt Deflation, *Econometrica* 1: 1–33

FRIEDMAN, M. and SCHWARTZ, A. J. (1963). *A Monetary History of the United States*, Princeton.

GOODHART, C. (1988). *The Evolution of Central Banks*, Cambridge, MA, and London, England: MIT Press.

GOODFRIEND, M. (1988). 'Money, Credit Banking and Payment Systems Policy', in D. Humphrey (ed.), *The US Payments System: Efficiency, Risk and the Role of the Federal Reserve*, Kluwer Academic Publishers.

—— (2007). 'How the World Achieved Consensus on Monetary Policy', *Journal of Economic Perspective*, 21: (Fall), 47–68.

GREENSPAN, A. (1994). 'Risk and Uncertainty in Monetary Policy', *American Economic Review*, 94: 33–40.

HUME, D. (1752). 'Banks and Paper Money', reprinted in F. H. Capie (ed.), *History of Banking*, I, Cheltenham: Edward Elgar, 1993.

KYDLAND, F. and PRESCOTT, E. (1977). 'Rules Rather than Discretion: The Inconsistency of Optimal Plans', *Journal of Political Economy*, 85: 473–91.

SCHOLES, M. (2000). 'Crisis and Risk Management', *American Economic Review*, 90: 17–21.

TAYLOR, J. (1993). 'Discretion versus Policy Rules in Practice', *Carnegie-Rochester Conference Series on Public Policy*, 39: (December), 195–214.

THORNTON, H. (1802). *An Inquiry into the Nature and Effects of Paper Credit of Great Britain*, ed. F. A. Hayek, Fairfield: Augustus M. Keley.

WICKSELL, K. (1907). 'The Influence of the Rate of Interest on Prices', *The Economic Journal*, 17: (May), 213–20.

—— (1935). *Lectures on Political Economy*, London: George Routledge.

WOODFORD, M. (2003). *Interest and Prices, Foundation of a Theory of Monetary Policy*. Princeton University Press.

CHAPTER 10

THE ROLE OF BANKS IN THE TRANSMISSION OF MONETARY POLICY

JOE PEEK
ERIC S. ROSENGREN[1]

Introduction

While macroeconomists have traditionally focused on the role of inside money in the transmission of monetary policy, over the past two decades an increased emphasis has been placed on the other side of bank balance sheets. Although the traditional interest rate channel for the transmission of monetary policy remains intact, the importance of the credit channel in augmenting the impact of monetary policy on the economy has gained credibility. Still, while the broad credit channel version has gained widespread acceptance, the more narrow bank lending channel remains somewhat controversial. The main points of contention are the extent to

[1] The views I express (with my co-author) are my own, not necessarily those of my colleagues on the Board of Governors of the Federal Open Market Committee (FOMC).

which a shift in monetary policy affects bank loan supply, and, if it does, the extent to which a change in bank loan supply can affect economic activity. Here, we describe the mechanisms by which the banking system transmits changes in monetary policy and provide an overview of the evidence on the efficacy of the view that bank lending plays an important role in determining the magnitude of the effect of monetary policy on the economy.

Several factors may explain the renewed interest in the role played by bank lending in the transmission mechanism of monetary policy. First, financial innovation has resulted in monetary policy shifting from focusing on money aggregates to focusing on interest rates, with the Federal Reserve relying on the federal funds rate as its policy instrument. Second, in the early 1990s, the US experienced significant banking problems that resulted in banks limiting their lending as a result of capital constraints.[2] Similar concerns with a bank 'capital crunch' help explain some of the reaction of monetary policymakers to their own banking problems in countries as diverse as Japan, Sweden, and Argentina.[3] Third, episodes of liquidity problems in the US, such as those associated with the failure of Penn Central, the 1987 stock crash, the Long-Term Capital Management crisis, and the events of 9/11, have highlighted the important role of bank lending during liquidity and financial crises.[4] Fourth, the concerns about the bank 'capital crunch' of the early 1990s have once again been raised as numerous banks and non-bank lending institutions became capital constrained as a result of the credit problems that became apparent in August 2007. This crisis has already extended through 2008 and has caused countries from around the world to intervene in an effort to stabilize the banking system and credit markets more generally. In addition to the standard interest rate reductions, these interventions have included significant capital injections, such as the Troubled Asset Relief Program (TARP) capital injections in the US, and extensions of broad liability guarantees to ensure that banks would have the ability to rollover debt. Finally, interest rates around the world are approaching the zero bound on nominal interest rates. By the end of 2008, both the US and Japan had interest rates close to zero. With traditional interest rate policies now becoming limited by the zero lower bound, countries have increasingly looked to alternative monetary policy tools. In the US, the extensive use of lending facilities has been used in an effort to stimulate lending by financial institutions, and in some cases these facilities have provided lending directly to borrowers. While our recent experience has highlighted that banking problems can become a direct cause of systemic problems, banks also are a key component of crisis resolution, even when they are not directly involved in instigating the financial crisis, because

[2] See Chap. 27 in this volume for a discussion of earlier episodes and the role of banking distress in exacerbating adverse macroeconomic shocks.

[3] See Chap. 26 in this volume for a thorough discussion of banking crises.

[4] See Chap. 5 in this volume for a discussion of the important role of banks in providing 'funding' liquidity.

many financial markets and non-bank financial intermediaries are dependent on banks to provide liquidity and credit enhancements.

This chapter will provide an overview of recent research about the role played by bank lending in the transmission of monetary policy. The second section will begin with a description of the mechanisms, under both the money view and the credit view, by which monetary policy is transmitted to the economy through the banking sector. The third section will examine the empirical evidence on how bank lending responds to changes in monetary policy. This literature has focused both on changes in the borrowing by firms and on changes in the amounts and composition of bank assets following a change in the stance of monetary policy. In particular, we will discuss the evidence concerning which banks and which firms are likely to play the key roles in transmitting monetary policy to the macroeconomy through the bank lending channel. This section also includes a discussion of how the effects emanating from the bank lending channel can be derailed by bank capital constraints that limit the extent to which banks are able to expand their balance sheets in response to an easing of monetary policy. The fourth section briefly reviews some of the literature on the role of banks in other countries, many of which are far more dependent on bank lending than is the US. The fifth section discusses how the role of bank lending may be altered by recent financial innovations and provides observations on the implications of recent events for the effectiveness of the bank lending channel. The final section provides some conclusions.

How is monetary policy transmitted through the banking system?

Traditional interest rate, or money, view

The traditional interest rate, or money, view of the transmission of monetary policy focuses on the liability side of bank balance sheets. The important role played by banks in this transmission mechanism arises from the reserve requirement constraint faced by banks. Because banks rarely hold significant excess reserves, the reserve requirement constraint typically is considered to be binding at all times.[5] Thus, shifts in monetary policy that change the quantity of outside money result in changes in the quantity of inside money in the form of the reservable deposits that can be created by the banking system.

[5] While that accurately describes non-crisis times, two recent notable exceptions are the episodes of quantitative easing in response to the crises experienced by Japan in the 1990s and the United States currently.

The transmission mechanism functions as follows. When the monetary authority undertakes open-market operations in order to tighten monetary policy (by selling securities), the banking industry experiences a decline in reserves. The fractional reserve system then forces banks (as a whole) to reduce reservable deposits in order to continue to meet the reserve requirement. This exogenous (to the banking sector) shock thus constrains bank behavior. To induce households to hold less reservable deposits (transactions accounts), interest rates on other deposits and non-deposit alternatives must rise. That is, since the supply of transactions deposits has declined relative to those of alternative assets, interest rates on these alternative assets would have to rise to clear the market for transactions deposits. As the increase in the short-term interest rate is transmitted to longer-term interest rates, aggregate demand declines.

The broad credit channel

The broad credit channel, also referred to as the balance sheet effect or financial accelerator, does not require a distinction among the alternative sources of credit. Instead, it is predicated on credit market imperfections associated with asymmetric information and moral hazard problems. Research on the credit channel was motivated, in large part, by the puzzle that monetary policy shocks that had relatively small effects on long-term real interest rates appeared to have substantial effects on aggregate demand. This literature attributes the magnification, or propagation, of the monetary policy shocks to frictions in the credit markets (see, e.g., Gertler and Gilchrist, 1993; Cecchetti, 1995; Hubbard, 1995; Bernanke and Gertler, 1995; and Bernanke, Gertler, and Gilchrist 1996; Oliner and Rudebusch 1996a). Because of the information asymmetries between borrowers and lenders, external finance is an imperfect substitute for a firm's internal funds.

The broad credit channel posits that an increase in interest rates associated with a tightening of monetary policy causes a deterioration in firm health, in terms of both net income and net worth. A firm's net income is impaired both because its interest costs rise and because its revenues deteriorate as the tighter monetary policy slows the economy. A firm's net worth is adversely impacted as the lower cash flows emanating from the firm's assets are discounted using the higher interest rates associated with the tightening of monetary policy. The deterioration in the firm's net income and the reduction in the collateral value of the firm's assets, in turn, cause an increase in the external finance premium that must be paid by the firm for all sources of external finance. This increase in the cost of external funds for borrowers over and above the risk-free interest rate then results in a reduction in aggregate demand in addition to that due to the increase in the risk-free interest rate associated with the interest rate channel of the transmission of monetary policy.

The bank lending channel

With the bank lending, or credit, view, in contrast to the money view, the focus of the transmission mechanism operating through bank balance sheets shifts from bank liabilities to bank assets. Again, when monetary policy tightens, the reduction in available bank reserves forces banks to create less reservable deposits. Banks must then either replace the lost reservable deposits with non-reservable liabilities, or shrink their assets, such as loans and securities, in order to keep total assets in line with the reduced volume of liabilities. Typically, one would expect to observe some combination of these responses, although Romer and Romer (1990) question the extent to which banks, in an age of managed liabilities, are unable easily to replace reservable deposits. However, to the extent that banks are unable or unwilling fully to insulate their loan portfolio, the interest rate effect on aggregate demand is supplemented with an additional effect emanating from a reduction in the availability of bank loans that further slows aggregate demand.

In a simple world with three assets—money, government bonds, and bank loans—three conditions must be satisfied for the bank lending channel to be operational in the transmission of monetary policy (see, e.g., Bernanke and Blinder, 1988; and Kashyap and Stein, 1994). First, as with the interest rate view, prices must not adjust fully and instantaneously to a change in the money supply. That is, money is not neutral. Second, open-market operations must affect the supply of bank loans. Third, loans and bonds must not be perfect substitutes as a source of credit for at least some borrowers. Of course, the set of assets can be expanded to include private sector bonds and non-bank intermediated loans, in which case the more narrow bank lending channel is distinguished from the broad credit channel by requiring that private sector bonds and non-bank intermediated loans must not be perfect substitutes for bank loans as a source of credit for at least some borrowers. Because only the second and third conditions distinguish the bank lending view from the money view, and because substantial evidence exists that wages and prices are not perfectly flexible, that the first condition holds will be assumed for this discussion.

With respect to the second condition, when open-market operations reduce the quantity of bank reserves, the banking system has no choice but to reduce reservable deposits, given the reserve requirement. However, banks do have choices, and individual banks do differ with respect to how, and to what extent, they respond to this decline in reserves. Banks must either raise non-reservable liabilities to replace the lost reservable deposits, reduce assets such as securities and loans, or make some combination of these portfolio adjustments. To the extent that banks do not regard non-reservable sources of funds as perfect substitutes for reservable deposits, they will not fully replace the lost reservable deposits, and thus must shrink their assets in order to keep total assets in line with their reduced volume of liabilities.

Asymmetric information and credit market frictions will play an important role in determining how an individual bank will respond on the liability side of its

balance sheet. Banks primarily use uninsured non-reservable liabilities, such as large time deposits, as the marginal source of funds during a period of monetary policy-tightening. However, the ease of raising large time deposits varies by bank. For example, one would expect that more transparent (e.g., publicly traded compared to non-publicly traded), larger, and healthier banking organizations would have relatively better access to external (uninsured) funds, and thus would tend to replace a higher proportion of their lost reservable deposits, resulting in a relatively smaller shrinkage in their assets.

Given that some shrinkage in bank assets will occur, a bank must then decide on the distribution of that shrinkage across the various assets held in its portfolio. Because securities are relatively liquid and considered to be 'secondary reserves', one would certainly expect banks to shrink their holdings of securities. However, to the extent that banks do not consider securities and loans to be perfect substitutes in their asset portfolio, one would expect that at least part of the adjustment in assets would be composed of a shrinkage in the volume of their loan portfolio, although initially the loan portfolio might temporarily grow from distress borrowing as loan customers access credit from previously established loan commitments and lines of credit (Morgan, 1988).

Asymmetric information and credit market frictions also play an important role in determining the extent to which firms consider bonds, or, more generally, publicly issued credit market instruments, and non-bank intermediated loans as perfect substitutes for bank loans. That is, to distinguish the broad credit channel from the bank lending channel, one must address the degree to which borrowers consider non-bank sources of credit as perfect substitutes for bank loans. To the extent that non-bank sources of credit are perfect substitutes for bank loans from the viewpoint of borrowers, then when a tightening of monetary policy reduces the availability of bank loans, borrowers will merely substitute these alternative sources of credit for bank loans. In that case, one would observe no impact on aggregate demand emanating from the reduction in bank credit beyond that due to the increase in the external finance premium associated with the broad credit channel.

While non-bank financial intermediaries provide loans, open-market instruments are available for short-term credit, and trade credit is available to some firms, these alternative sources of credit are not perfect substitutes for bank credit for a variety of institutional reasons. With respect to the substitutability of intermediated loans and publicly issued credit market instruments, all firms do not have access to public credit markets. In particular, smaller firms are not able to issue such debt because the issue size would be too small to overcome the fixed costs of issuance at a reasonable interest rate. Similarly, firms that are sufficiently opaque or have a sufficiently low credit standing to require close monitoring by a financial intermediary would not have direct access to the credit markets. Still, even though large, highly rated firms can directly access public credit markets by issuing commercial paper, issuing unsecured commercial paper still may involve participation

by banks, in so far as the issuing firms obtain third-party guarantees from banks to enhance the credit rating of the commercial paper in order to lower the interest cost to the firm.

Similarly, alternative sources of intermediated loans are not perfect substitutes for bank loans for at least some borrowers. While bank loans share the attributes of many of the alternative forms of intermediated loans for firms, there are important differences. For example, insurance companies are very active in the commercial real estate market and are important providers of term financing that allows them to better match the maturities of their assets and liabilities. Similarly, finance companies provide asset-backed financing—for example, for loans collateralized by inventories and accounts receivable. Yet, for a small, opaque firm with few tangible assets, bank loans may be the only source of an unsecured line of credit or of a loan secured by an asset that might not be easily commoditized. As a result, the clientele effect in bank lending results in many firms being 'bank dependent', having few alternatives to banks should their bank credit be curtailed.

Empirical evidence on the role of bank lending for the transmission of monetary policy

There are several challenges faced by empirical researchers investigating the bank lending view. First, they need to determine if a change in monetary policy does affect bank lending. Then, if bank lending is affected, the issue becomes the extent to which shifts in bank loan supply do, in fact, affect aggregate demand. The difficulties in establishing the first point are twofold. First, to what extent are banks able to insulate their loan portfolios from monetary policy shocks by adjusting other components of their balance sheet? The second difficulty concerns identifying a bank-loan-supply shock, in so far as a decline in bank loans following a tightening of monetary policy may simply reflect a decline in loan demand rather than loan supply.

The effect of monetary policy on bank loan supply

While the theoretical conditions required for bank loan supply to be affected by changes in monetary policy are clear, the empirical disentangling of shifts in loan supply from shifts in loan demand is not straightforward. At an aggregate level, Bernanke and Blinder (1992), among others, show that bank lending does contract

when monetary policy becomes tighter. However, such an observed correlation may reflect a reduction in loan demand as the economy weakens in response to the tighter monetary policy, rather than being the result of a reduction in bank loan supply. Furthermore, even if one observed an initial increase in bank loans or a notable delay in the decline in bank loans following a tightening of monetary policy, such evidence would not necessarily conflict with an inward shift in bank loan supply in response to a tightening of monetary policy. For example, the initial response of firms to a tightening of monetary policy may be an increase in loan demand due to a need to finance the build up of inventories as aggregate demand initially declines faster than production. Even though banks may decrease loan supply immediately to borrowers without loan commitments, the total amount of bank loans may temporarily increase, in so far as banks are forced to honor existing loan commitments (Morgan 1998). Thus, the endogeneity issues associated with using aggregate data for total loans prevent obtaining a clear answer.

Kashyap, Stein, and Wilcox (1993) provide an alternative approach for identifying an effect of monetary policy on bank loan supply, although the analysis is still based on aggregated data. They investigate the change in the mix between bank loans and commercial paper in the composition of firm external finance, with the argument being that if the decline in loans is due to a general decline in credit demand associated with a slowing of the real economy, then demand for other types of credit should similarly decline. Finding that a tightening of monetary policy is associated with an increase in commercial paper issuance and a decline in bank loans, they conclude that a tightening of monetary policy does reduce bank loan supply rather than the decline in bank loans simply reflecting a reduction in loan demand as the economy slows. In the same vein, Ludvigson (1998) investigates the composition of automobile finance between bank and non-bank providers of credit. She finds that, in fact, a tightening of monetary policy reduces the relative supply of bank loans, consistent with the bank lending channel. In contrast, Oliner and Rudebusch (1996b) revisit the Kashyap, Stein, and Wilcox (1993) approach using a different measure of the mix of external finance and disaggregating the data into those for small firms and those for large firms. They argue that their evidence is consistent with the broad credit channel rather than the more narrowly defined bank lending channel. However, this only highlights the weaknesses associated with attempting to isolate bank-loan-supply shocks from shifts in loan demand using aggregate data. In fact, in their reply, Kashyap, Stein, and Wilcox (1996) close by suggesting that a more definitive answer will have to rely on an analysis using micro data at the individual bank and firm levels.

By advancing the analysis to focus on panel data, the literature has been able to obtain more definitive results about the impact of changes in monetary policy on bank loan supply. The key was relating cross-sectional differences in bank, or banking organization, characteristics to differences in the extent to which banks were able to insulate their loan portfolios from a tightening of monetary policy.

Two aspects of bank characteristics appear to have been the primary focus. First, the ability of banks to raise non-reservable liabilities to replace the lost reservable deposits is a key factor in determining the extent to which a bank must adjust its loan portfolio when a tightening of monetary policy occurs. Because these funds are, for the most part, uninsured liabilities, bank characteristics related to their access to external funds—for example, size, health, and having direct access to capital markets—play an important role in determining the ability of banks to insulate their loan portfolios from the effects of changes in monetary policy. Second, because banks face a capital requirement constraint in addition to the reserve requirement constraint on their activities, banks may differ in their response to a change in the stance of monetary policy depending on which constraint is most binding. If the capital ratio requirement is the binding constraint, easing the reserve requirement constraint through open-market operations should have little, if any, effect on bank lending. That is, because the binding constraint has not been eased, expansionary monetary policy, at least operating through the bank lending channel, would be like 'pushing on a string'.

Kashyap and Stein (1995) note that with a tightening of monetary policy, and the associated loss in reservable deposits, it is costly for banks to raise uninsured deposits. However, banks differ in the degree to which they have access to external funds. They hypothesize that bank size is a reasonable proxy for their degree of access to uninsured liabilities, with smaller banks having more limited access, and thus having their loan portfolio impacted more by a tightening of monetary policy. Indeed, they find empirical support for the proposition that small banks are more responsive (shrink their loan portfolios by more) to a monetary policy tightening than are large banks.

Subsequently, Kashyap, and Stein (2000) extend their analysis of the relative ease with which banks can raise uninsured deposits following a monetary policy tightening, noting that the bank loan response also will differ depending on the liquidity position of the bank. A bank that finds it relatively costly to raise uninsured deposits but that has large securities holdings has the option of adjusting to the shrinkage of reservable deposits by selling some of its securities, while a less liquid bank may be forced to shrink its loan portfolio by a greater degree. In a large cross-section of banks, they find evidence that the loan portfolios of smaller, more-illiquid banks are the most responsive to monetary policy shocks.

Campello (2002) distinguishes among these smaller banks based on whether the bank is affiliated with a large multibank holding company, finding that the lending of small banks that are affiliated with large multibank holding companies reacts less to a tightening of monetary policy than does the lending of similar small (stand-alone) banks that are not affiliated with multibank holding companies. Although this evidence indicates that small banks affiliated with multibank holding companies are better able to insulate their lending from a tightening of monetary policy, the extent to which this is due to the channeling of internal holding-company funds

to bank subsidiaries rather than being due to large multibank holding companies having easier access to external funds is not clear. Campello tries to address this issue by using capital-to-asset ratios to distinguish among bank holding companies. Similarly, Kishan and Opiela (2000) use the capital-to-asset ratio as the proxy for the bank's ability to raise uninsured deposits, finding that the loan portfolios of well-capitalized banks are less sensitive to monetary policy shocks than are those of poorly capitalized banks of the same size. However, for reasons discussed below, capital-constrained banks may behave differently for reasons other than their ability to raise uninsured deposits.

Holod and Peek (2007) utilize the distinction between publicly traded and non-publicly traded banks to classify banks by the ease with which they can access external funds. They find that after controlling for size, capitalization, and other factors, the loan portfolios of publicly traded banks shrink less than those of non-publicly traded banks when monetary policy tightens due to the banks' ability to raise external funds, including issuing large time deposits. Furthermore, as one would expect, when a distinction is made between tightening and easing monetary policy, the estimated effect can be attributed to the effects of monetary policy tightening (tightening a binding constraint) rather than monetary policy easing (possibly 'pushing on a string').

The second important characteristic of banks that can affect the extent to which the bank lending channel is operative is whether banks face a binding capital constraint. As a result of the 'headwinds in monetary policy' noted by Chairman Greenspan during the recovery from the 1990 recession, a variety of authors have examined the impact that significant bank health problems can have on the transmission of monetary policy. For example, Peek and Rosengren (1995a) examined the impact that being capital constrained had on a bank's ability to lend during the period of significant banking problems in the early 1990s in New England. Using a simple static model, they show that banks facing a binding capital constraint are limited in altering the size of their balance sheet, restricting the ability of capital-constrained banks to respond to monetary policy shocks. They document that experiencing an adverse capital shock that makes the capital constraint binding will cause banks to shrink both assets and liabilities. Peek and Rosengren (1995a) also show that the behavior of capital-constrained banks in New England differed from that of unconstrained banks, with the loan portfolios of unconstrained banks responding more to monetary policy shocks than those of the capital-constrained banks.

In a subsequent study, Peek and Rosengren (1995b) focus on the direct impact of the enforcement of capital regulations by bank supervisors on the ability of capital-constrained banks to lend, and thus to be able to increase loans in response to an easing of monetary policy. They examine the impact on bank lending of formal regulatory actions (cease and desist orders and written agreements) imposed on banks that experienced asset quality problems. They find that the enforcement

actions by bank regulators included explicit capital targets that needed to be achieved over a short time frame. The result was an immediate and significant reduction in bank loan portfolios associated with the imposition of the enforcement action that persisted for some time thereafter while the bank continued to operate under the enforcement action.

Alternatively, banks can become capital constrained as a consequence of changes in capital regulations. A number of authors have examined whether such changes can cause banks to be particularly responsive to their capital constraint, and, by implication, less responsive to changes in monetary policy. For example, Hall (1993) finds that the introduction of the Basel I Accord had a significant impact on bank portfolios. Hancock and Wilcox (1994) also found that the implementation of the Basel I Accord affected banks' willingness to lend. However, Berger and Udell (1994) do not find evidence that the Basel I Accord created a bank 'capital crunch'. More recently, a concern raised with the proposed Basel II Accord has been that the new capital regulations would magnify potential capital constraints during recessions (e.g., Kashyap and Stein 2004), making banks less responsive to an easing of monetary policy. Thus, a very real concern with the effectiveness of the bank lending channel, and thus the overall effectiveness of monetary policy, is whether banks are capital constrained at the time of an easing of monetary policy.

Real effects of shifts in bank loan supply

Given that the empirical evidence generally supports the proposition that banks, particularly those that may find it relatively expensive to raise uninsured liabilities, respond to a monetary policy tightening by reducing loans, we turn to the second link in the bank lending channel mechanism. For the reduction in bank loans to have an impact on economic activity, firms must not be able easily to substitute other sources of external finance when bank loan supply is cut back. At a somewhat aggregated level, Gertler and Gilchrist (1994) find that the investment of an aggregate of small firms is more responsive to changes in monetary policy than is the investment of an aggregate of large firms, a set of firms that presumably is less bank dependent. Similarly, Ludvigson (1998), comparing bank and non-bank sources of automobile loans, finds that the composition of automobile credit impacts automobile sales, even after controlling for the standard factors that probably impact automobile demand.

Additional evidence at an aggregate level is provided by Driscoll (2004) who uses a panel of state-level data to investigate the extent to which bank loan-supply shocks affect output. Using state-specific shocks to money demand as an instrumental variable to address the endogeneity problem, he does not find a meaningful effect of loan-supply shocks on economic activity at the state level. Ashcraft (2006), similarly basing his analysis on state-level data, attempts to exploit differences

between stand-alone banks and banks affiliated with multibank holding companies in their degree of access to external funds in order to identify loan-supply shocks related to changes in monetary policy. While he does find a difference between the two types of banks in their lending response to changes in monetary policy, he does not find a significant effect of these bank-loan-supply shocks on state income growth. In contrast, Ashcraft (2005), using the cross-guarantees of two failed Texas bank holding companies as his identification mechanism to address the endogeneity problems, finds that the failures of healthy banks forced by the cross-guarantee provisions were associated with reduced local economic activity. Thus, this suggests that bank lending is special, in so far as it appears that other lenders (even other banks) did not fill the gap created by the sharp reduction in lending by the failed banks, and is consistent with an operative lending channel.

Another approach that provides direct evidence that a reduction in bank loan supply adversely affects macroeconomic activity is provided by Peek and Rosengren (2000). Using the banking problems in Japan as the source of an exogenous loan-supply shock in the US, they are able to avoid the common endogeneity problem faced by studies that rely on domestic shocks to bank loan supply. Furthermore, by focusing on commercial real estate loans that tend to have local or regional markets, they are able to exploit cross-sectional differences across geographic regions to show that the decline in loans had real effects. That is, the pull-back by Japanese banks in local US markets was not fully offset by other lenders stepping in to fill the void.

Taking still a different tack, Peek, Rosengren, and Tootell (2003) obtain evidence of a macroeconomic effect of shifts in bank loan supply. They find that adverse shocks to bank health weaken economic activity in the major GDP components that one would expect to be most affected by bank-loan-supply shocks—for example, the change in business inventory investment, while not impacting other major components of GDP whose fluctuations would be correlated with demand shocks.

While such aggregate evidence is more than simply suggestive of an operative bank lending channel, to obtain even more convincing evidence about the efficacy of the bank lending channel, one must turn to disaggregated data, preferably at the firm level. One way to test whether bank lending is special is to determine whether a monetary policy tightening disproportionately impacts borrowers that are more reliant on bank lending as a source of external finance. A variety of authors have examined individual firm-level data to determine whether financial constraints cause non-financial firms to react more to monetary policy shocks, for example, by reducing investment more in response to a monetary policy tightening (e.g., Fazzari, Hubbard, and Peterson, 1988). The proxies for liquidity constraints have included dividend payouts (Fazzari, Hubbard, and Peterson, 1988), size (Gertler and Gilchrist, 1994), and bond ratings (Kashyap, Lamont, and Stein, 1994). The evidence tends to support the proposition that external funds are more costly to raise

relative to internal funds, so that firms that depend more on external finance are likely to be more adversely impacted by a reduction in bank loan supply.

This, of course, presumes that bank loans are special to firms, so that such loans cannot be easily replaced with non-bank loans or by issuing credit market instruments. A large literature speaks precisely to this point. For example, James (1987) notes that the stock price of a firm rises in response to an announcement of a new loan agreement. Slovin, Sushka, and Poloncheck (1993) observe that when Continental Illinois Bank failed, it adversely impacted borrowers that had a close banking relationship with that bank. However, this relationship did not hold if the Continental Illinois loan was part of a loan participation, but did hold if Continental Illinois was the lead underwriter of the loan. In terms of the strength of the banking relationship, Peterson and Rajan (1995) note that a firm's banking relationship often involves both a deposit and a lending relationship. They find that the strength of lending relationships, as indicated by a firm holding deposits at the bank, is indicative of how extensively the firm relies on bank lending. Finally, Fields, et al. (2006) argue that the value of lending relationships has diminished substantially over time, due in part to the further development of financial markets and the increased availability of information about borrowers. However, their sample includes only publicly traded firms, precisely those firms that are the least likely to be bank dependent.

Thus, the evidence from studies based on individual non-financial firms supports the proposition that many firms are, in fact, bank dependent, and that their economic activity is adversely affected by reductions in bank loan supply. While other financial intermediaries provide external finance to firms, this credit tends to be directed to specific types of loans. Finance companies tend to focus on asset-backed lending, such as receivables, while insurance companies tend to make longer duration loans more closely to match the duration of their liabilities. Thus, banks remain the primary source of funding for smaller firms that do not have ready access to external finance from other sources.

Non-US evidence on bank lending and the transmission of monetary policy

The role of banks in the transmission of monetary policy is potentially more relevant in many other countries than in the US because of their relatively greater reliance on bank finance. A good example is Japan, where banks continue to have a significant role in financing large as well as small companies, although due to the deregulation of bond markets the largest companies are increasingly able to tap

directly into financial markets (e.g., Hoshi and Kashyap, 2001). The Japanese economy also is particularly interesting because of a variety of characteristics that make bank–firm ties particularly close, such as widespread cross-shareholding, bank representatives placed on firms' board of directors, and bank-centered keiretsu groups (see, e.g., Kaplan and Minton, 1994; Kang and Shivdasani, 1995; and Morck and Nakamura, 1999).

Even before the Japanese banking problems that began in the early 1990s were able to have their full impact, Hoshi, Scharfstein, and Singleton (1993), using the mix-effect technique with aggregate data, found that when monetary policy tightened, the share of bank loans compared to insurance company loans declined. Furthermore, for Japanese firms not affiliated with bank-centered keiretsu groups, and thus less closely connected to banks, firm liquidity was a more important determinant of their investment when monetary policy tightened and bank credit became less available.

After the bubble burst and bank health began to deteriorate, Japanese banks faced with potential capital constraints sought ways to continue to lend to domestic borrowers while still shrinking their balance sheets. Peek and Rosengren (1997; 2000) find that global Japanese banks initially shrank their assets abroad in order to insulate their domestic lending. As the Japanese banking problems continued, domestic borrowers and the lending channel were affected. For example, Ito and Sasaki (2002) find evidence of a 'credit crunch' in Japan as binding capital constraints became an important factor in the ability of Japanese banks to continue to lend. Similarly, Kang and Stulz (2000) find that the banking problems in Japan had the greatest adverse impact, both in terms of stock prices and their investment expenditures, on the firms that were the most bank dependent. In addition, several studies argue that problems were compounded by Japanese banks applying international bank capital standards that resulted in the need for Japanese banks to increase their capital ratios (e.g., Hall, 1993; and Montgomery, 2005).

Given the relative importance of bank lending as a source of credit in Japan, it is likely that the severe banking problems weakened the bank lending channel and contributed to the prolonged malaise in the Japanese economy throughout the 1990s and early 2000s, even though monetary policy reduced interest rates to near zero. This was magnified by distortions that had a broader adverse impact on the economy emanating from the combination of lending relationships and the perverse incentives faced by banks that led to a misallocation of much of the credit that was provided by banks (Peek and Rosengren, 2005). In fact, Caballero, Hoshi, and Kashyap (2008) link the misallocation of credit to broader economic problems in Japan, finding that investment by firms was seriously distorted by the desire of banks to support 'zombie' (insolvent) firms.

Similarly, the evidence suggests that the deterioration in bank health created similar problems when other Asian countries experienced financial difficulties in the late 1990s. For example, Ferri and Kang (1999) found that South Korean

problems caused a significant 'credit crunch' when bank capital became constrained. Given that the chaebols (business groups) in Korea have some similarities to the keiretsus in Japan, one might expect that some of the same results found in Japan also would hold in emerging market economies such as Korea.

Turning to the evidence from Europe, Europe is somewhere between Japan and the US in terms of bank dependence. While European countries do not have to the same degree the formalized banking relationships of the Japanese keiretsus, European firms have not had the same direct access to financial markets as do many US firms, although the conversion to the euro and the integration of European financial markets has improved their access.

Although bank lending in Europe is relatively more important than in the US, the evidence on the bank lending channel is mixed. Ehrmann, et al. (2001) apply a number of the empirical tests conducted on US data to European data in order to investigate the importance of the bank lending channel in Europe. Consistent with the results in the US, they find that monetary policy does alter bank loan supply, particularly for those banks that are liquidity constrained. However, they do not find that the size of the bank influences the bank's reaction to a monetary policy shock. Similarly, Gambacorta (2005), using Italian data, finds that a tightening of monetary policy reduces bank lending, with the effect being mitigated for banks that are well capitalized, are relatively liquid, or can benefit from the operation of internal capital markets by being affiliated with a bank holding company. Again, bank size is not important. In contrast, Angeloni, et al. (2003) find that most of the European evidence is consistent with a classic interest rate channel for the transmission of monetary policy, although they do find some country-specific evidence supporting a bank lending channel, with bank loan supply reacting to changes in monetary policy in a number of countries.

Some observations on recent events

The evolution of financial markets, as well as other changes in the economy, raises questions about the potency of the bank lending channel in the future. First, banks have been consolidating in the US as well as elsewhere, resulting in a much greater concentration of banking assets in the largest banking institutions that are likely to have a variety of sources of liquidity. Second, the largest banks that are least likely to be liquidity constrained are finding ways to compete for many loans that used to be the focus primarily of small banks. For example, many large banks are commoditizing small business loans in cases where the owners of the small business have significant net worth and/or a high credit score. Third, non-bank financial

intermediaries increasingly are competing in banking markets, often through finance companies, mortgage companies, or industrial loan companies that are not under the traditional bank regulatory umbrella. Also, the ability to securitize assets has increasingly placed loans in asset-backed financial instruments that are widely held, rather than being retained in bank portfolios. These trends could weaken the bank lending channel, as both banks and businesses continue to gain better access to capital markets.

Similarly, one would expect the lending channel to weaken in Europe as the conversion to the euro has led to capital markets becoming more integrated and financial regulations becoming more standardized. As in the US, bank consolidation should contribute to the ability of banks to insulate their loan portfolios from the impact of a tightening of monetary policy. At the same time, the better access to the corporate bond market as it becomes integrated across countries will provide many firms with improved alternatives to bank loans should bank credit be adversely impacted by a tightening of monetary policy.

However, concerns with the relevance of financial intermediaries have undergone significant revision as a result of the credit crisis of 2007–8. Most securitization markets did not fare well. Concerns about the accuracy of ratings and uncertainty about the financial strength and transparency of many securitization vehicles caused a dramatic decline in securitization activity in 2008. In addition, non-bank financial intermediaries have increasingly found it difficult to secure financing in illiquid markets. Major investment banks have either been acquired by commercial banks or have changed their charter to become a bank holding company. Similarly, many non-bank financial intermediaries, such as American Express and GMAC, have converted to bank holding companies. Bank holding company status enables the organization to gather bank deposits that provide a less volatile source of financing, as well as providing access to the discount window, including the many related lending facilities that were created in response to the 2007–8 financial crisis.

Other recent changes suggest an enhanced role for banks going forward. The commercial paper market came under severe stress in 2008, causing the Federal Reserve to create a new lending facility for issuers of commercial paper. However, the facility was available only for high quality commercial paper (A1/P1), making it increasingly difficult for A2/P2 issuers to finance their lending. As the backup lines of credit for commercial paper are renewed, this market is likely to remain more bank dependent and more skewed to high-quality borrowers than it was prior to the crisis.

Similarly, the role of money market mutual funds is in the process of evolving. In September 2008, the Reserve Fund 'broke the buck', causing a significant run on prime funds that invested in short-term credit instruments such as commercial paper and bank certificates of deposit. This run caused policymakers to extend insurance to money market mutual funds (MMMFs) and to establish a new

lending facility, the asset-backed commercial paper money market-fund liquidity facility (AMLF). In addition, short-term Treasury rates have been close to zero, making it increasingly difficult for MMMF's to provide a positive return without taking on more risk. These developments make more uncertain the role money market mutual funds will play in the future, possibly resulting in the need for more short-term financing to occur through the banking system.

The roles of banks and bank regulation also are likely to be re-examined as a result of the credit crisis. Starting in July 2007 with the substantial downgrades of financial instruments tied to the subprime mortgage market, many commercial banks found it difficult to obtain short-term funding. Leveraged buyout loans, jumbo mortgage loans, and asset-backed commercial paper were all difficult to place in securitized products. As a result, many of these assets flowed back onto bank balance sheets, the interest rates for assets normally placed in asset-backed securities rose significantly, and banks became the primary source of continued financing.

In many respects, the liquidity problems that began in July 2007, and were still continuing through the end of 2008, have shown that liquidity problems in the banking system can be a significant constraint that requires the attention of policymakers. Particularly notable has been the elevated spread between LIBOR rates and overnight index swap (OIS) rates (which provide a measure of the expected federal funds rate). Since LIBOR is used as the reset rate on most subprime mortgages, is often used as the base rate on credit cards, and is frequently used as the base rate for business loans, the widening of this spread has served to increase credit costs to a wide variety of borrowers. Because the increase in the LIBOR/OIS spread during the Fall of 2007 frequently exceeded seventy-five basis points, the liquidity problems merely offset much of the initial easing by the Federal Reserve for many borrowers. The spread peaked in the fall of 2008 in the wake of the failure of Lehman Brothers and AIG, and at the end of 2008 still remained well above levels experienced prior to the credit crisis.

The dramatic intervention in banking markets through the substantial equity infusions into the banking system and the broad guarantees on bank liabilities highlights policy concerns that banking problems could disrupt the transmission of monetary policy and cause a broader downturn in the real economy. As the financial problems have mounted, the US economy has become more bank dependent and moved to simpler financial contracts with higher credit standards. Furthermore, the deleveraging of financial institutions has exacerbated other problems in the economy, causing both aggressive fiscal and monetary policy responses in the US.

As monetary policy has reached the zero lower bound, monetary policies have moved from a focus on the federal funds rate to increased purchases of other assets, either indirectly through lending facilities or through direct purchases. This dramatic expansion of the Federal Reserve's balance sheet is designed to offset some of the lost credit availability as financial institutions and securitization markets are no longer able to sustain the same volume of lending. While it is too

soon to estimate the full impact of these actions, the link between financial institutions and the transmission of monetary policy is likely to be much stronger than before the onset of the crisis.

At the time of the writing of this article, the US economy has already been in a recession for a year, and many interest rate spreads remain elevated by historical standards. These events highlight that many of the financial market innovations used to remove assets from bank balance sheets still depend on liquidity provision by banks in the event of liquidity problems, and that in the future, an increased emphasis on returning assets to bank balance sheets may occur. For example, instruments such as structured investment vehicles (SIVs) that relied on market triggers for liquidation, rather than more traditional backup liquidity from banks, proved not to be able to weather the liquidity problems, resulting in the liquidation of many SIVs and the need to find alternative financing mechanisms.

These events make clear that banks remain a critical component in financing assets. In fact, the combination of the large losses generated by the subprime problems and the swelling of bank balance sheets as a result of both banks moving off-balance sheet assets back onto their balance sheets and their inability to move leveraged buyout loans off their balance sheets have given rise to complaints about a bank-induced 'credit crunch'. While it will take the passage of time to evaluate how significantly capital constraints affect the US or European economies, the potential threat from capital-constrained banks highlights the important role still played by banks. The implications of these issues for the transmission of monetary policy remain an area where future research would be beneficial.

Conclusions

Theoretical and empirical studies produced over the past two decades have emphasized the important role for banks in the transmission of monetary policy. Much of this work has highlighted a role for changes in bank assets in response to a monetary policy shock, above and beyond the familiar interest rate channel operating on the liability side of bank balance sheets. The empirical evidence provides substantial support for the view that liquidity-constrained banks and bank-dependent borrowers can be adversely impacted by a tightening of monetary policy. The evidence also indicates that a bank lending channel can be important in an international context, especially in countries where banks and firms have less direct access to financial markets.

In addition, a significant body of research highlights that during 'credit crunches' the bank lending channel can be short-circuited, requiring larger changes

in monetary policy instruments to obtain the same desired change in aggregate demand. Again, the international evidence indicates that capital-constrained banks make it difficult for monetary policy to have as large an impact as would occur if banks were not capital constrained.

Looking forward, the bank lending channel may be enhanced, as off-balance sheet and non-bank financing become more difficult to obtain as a result of the credit and liquidity shocks recently experienced. Given the extensive need for government support of banks and credit provision more generally, the current financial crisis should be a particularly fruitful period for investigation by researchers. The extent of the recent problems indicates that banks are likely to continue to play an important, although changing, role as financial markets continue to evolve.

References

Angeloni, I., Kashyap, A. K., Mojon, B., and Terlizzese, D. (2003). 'Monetary Transmission in the Euro Area: Does the Interest Rate Channel Explain It All?', National Bureau of Economic Research Working Paper 9984.

Ashcraft, A. (2005). 'Are Banks Really Special? New Evidence from the FDIC-Induced Failure of Healthy Banks', *American Economic Review*, 95: 1712–30.

—— (2006). 'New Evidence on the Lending Channel', *Journal of Money, Credit and Banking*, 38: 751–75.

Berger, A. N. and Udell, G. F. (1994). 'Did Risk-Based Capital Allocate Bank Credit and Cause a "Credit Crunch" in the U.S.?', *Journal of Money, Credit and Banking*, 26: 585–628.

Bernanke, B. S. and Blinder, A. S. (1988). 'Credit, Money, and Aggregate Demand', *American Economic Review*, 78: 435–9.

—— —— (1992). 'The Federal Funds Rate and the Channels of Monetary Transmission', *American Economic Review*, 82: 901–21.

—— and Gertler, M. (1995). 'Inside the Black Box: The Credit Channel of Monetary Policy Transmission', *Journal of Economic Perspectives*, 9: 27–48.

—— —— and Gilchrist, S. (1996). 'The Financial Accelerator and the Flight to Quality', *The Review of Economics and Statistics*, 78: 1–15.

Caballero, R. J., Hoshi, T., and Kashyap, A. K. (2008). Zombie Lending and Depressed Restructuring in Japan, *American Economic Review*, 98: 1943–77.

Calomiris, C. W. (2009). 'The Great Depression and Other Contagious Events', in Berger, A., Molyneux, P., and Wilson J. O. S. (eds.), *The Oxford Handbook of Banking*. Oxford: Oxford University Press.

Campello, M. (2002). 'Internal Capital Markets in Financial Conglomerates: Evidence from Small Bank Responses to Monetary Policy', *Journal of Finance*, 57: 2773–805.

Caprio, G., Jr. and Honohan, P. (2009). 'Banking Crises', in Berger, A., Molyneux, P., and Wilson J. O. S. (eds.), *The Oxford Handbook of Banking*. Oxford: Oxford University Press.

Cecchetti, S. G. (1995). 'Distinguishing Theories of the Monetary Transmission Mechanism', *Federal Reserve Bank of St. Louis Review*, May/June, 83–97.

Driscoll, J. C. (2004). 'Does Bank Lending Affect Output? Evidence from the U.S. States', *Journal of Monetary Economics*, 51: 451–71.

Ehrmann, M., Gambacorta, L., Martìnez-Pagès, J., Sevestre, P., and Worms, A. (2001). 'Financial Systems and the Role of Banks in Monetary Policy Transmission in the Euro Area', European Central Bank Working Paper 105.

Fazzari, S. M., Hubbard, R. G., and Petersen, B. C. (1988). 'Financing Constraints and Corporate Investment', *Brookings Papers on Economic Activity*, 141–95.

Ferri, G. and Kang, T. S. (1999). 'The Credit Channel at Work: Lessons from the Republic of Korea's Financial Crisis', World Bank Policy Research Working Paper 2190.

Fields, L. P., Fraser, D. R., Berry, T. L., and Byers, S. (2006). 'Do Bank Loans Relationships Still Matter?', *Journal of Money, Credit and Banking*, 38: 1195–209.

Gambacorta, L. (2005). 'Inside the Bank Lending Channel', *European Economic Review*, 49: 1737–59.

Gertler, M. and Gilchrist, S. (1993). 'The Role of Credit Market Imperfections in the Monetary Transmission Mechanism: Arguments and Evidence', *The Scandinavian Journal of Economics*, 95: 43–64.

——— (1994). 'Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms', *The Quarterly Journal of Economics*, 109: 309–40.

Hall, B. J. (1993). 'How Has the Basel Accord Affected Bank Portfolios?', *Journal of the Japanese and International Economies*, 7: 408–40.

Hancock, D. and Wilcox, J. A. (1994). 'Bank Capital and the Credit Crunch: The Roles of Risk-Weighted and Unweighted Capital Regulation', *Journal of the American Real Estate and Urban Economics Association*, 22: 59–94.

Holod, D. and Peek, J. (2007). 'Asymmetric Information and Liquidity Constraints: A New Test', *Journal of Banking & Finance*, 31: 2425–51.

Hoshi, T. and Kashyap, A. N. (2001). *Corporate Financing and Governance in Japan*. Cambridge, MA: MIT Press.

——— Scharfstein, D., and Singleton, K. J. (1993). 'Japanese Corporate Investment and Bank of Japan Guidance of Commercial Bank Lending', in Singleton, K. J. (ed.), *Japanese Monetary Policy*, Chicago: University of Chicago Press, 63–94.

Hubbard, R. G. (1995). 'Is There a "Credit Channel" for Monetary Policy?', *Federal Reserve Bank of St. Louis Review*, May/June, 63–77.

Ito, T. and Sasaki, Y. N. (2002). 'Impacts of the Basel Capital Standard on Japanese Banks' Behavior', *Journal of the Japanese and International Economies*, 16: 372–97.

James, C. (1987). 'Some Evidence on the Uniqueness of Bank Loans', *Journal of Financial Economics*, 19: 217–35.

Kang, J. K. and Shivdasani, A. (1995). 'Firm Performance, Corporate Governance, and Top Executive Turnover in Japan', *Journal of Financial Economics*, 38: 29–58.

——— and Stulz, R. M. (2000). 'Do Banking Shocks Affect Borrowing Firm Performance? An Analysis of the Japanese Experience', *Journal of Business*, 73: 1–23.

Kaplan, S. N. and Minton, B. A. (1994). 'Appointments of Outsiders to Japanese boards: Determinants and Implications for Managers', *Journal of Financial Economics*, 36: 225–58.

Kashyap, A. K., Lamont, O. A., and Stein, J. C. (1994). 'Credit Conditions and the Cyclical Behavior of Inventories', *The Quarterly Journal of Economics*, 109: 565–92.

——— and Stein, J. C. (1994). 'Monetary Policy in Bank Lending', in Mankiw, N. G. (ed.), *Monetary Policy*, Chicago: University of Chicago Press, 221–56.

——— ———, (1995). 'The Impact of Monetary Policy on Bank Balance Sheets', *Carnegie-Rochester Conference Series on Public Policy*, 42: 151–95.
——— ——— (2000). 'What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?', *American Economic Review*, 90: 407–28.
——— ——— (2004). Cyclical Implications of the Basel II Capital Standard, Federal Reserve Bank of Chicago *Economic Perspectives*, First Quarter, 18–31.
——— ——— and Wilcox, D. W. (1993). Monetary Policy and Credit Conditions: Evidence from the Composition of External Finance, *American Economic Review*, 83: 78–98.
——— ——— ——— (1996). Monetary Policy and Credit Conditions: Evidence from the Composition of External Finance: Reply, *American Economic Review*, 86: 310–14.
Kishan, R. and Opiela, T. (2000). Bank size, bank capital, and the bank lending channel, *Journal of Money, Credit and Banking*, 32: 121–141.
Ludvigson, S. (1998). The channel of monetary transmission to demand: Evidence from the market for automobile credit, *Journal of Money, Credit and Banking*, 30(3): 365–83.
Montgomery, H. (2005). The effect of the Basel accord on bank portfolios in Japan, *Journal of the Japanese and International Economies*, 19: 24–36.
Morck, R. and Nakamura, M. (1999). Banks and corporate control in Japan, *Journal of Finance*, 54(1): 319–39.
Morgan, D. P. (1988). 'The Credit Effect of Monetary Policy: Evidence Using Loan Commitments', *Journal of Money, Credit and Banking*, 20: 102–18.
Oliner, S. D. and Rudebusch, G. D. (1996a). 'Is There a Broad Credit Channel for Monetary Policy?', *Federal Reserve Bank of San Francisco Economic Review*, 1: 3–13.
——— ——— (1996b). 'Monetary Policy and Credit Conditions: Evidence from the Composition of External Finance: Comment', *American Economic Review*, 86: 300–9.
Peek, J. and Rosengren, E. S. (1995a). 'The Capital Crunch: Neither a Borrower Nor a Leader Be', *Journal of Money, Credit and Banking*, 27: 625–38.
——— ——— (1995b). 'Bank Regulation and the Credit Crunch', *Journal of Banking & Finance*, 19: 679–92.
——— ——— (1997). 'The International Transmission of Financial Shocks: The Case of Japan', *American Economic Review*, 87: 495–505.
——— ——— (2000). 'Collateral Damage: Effects of the Japanese Bank Crisis on Real Activity in the United States', *American Economic Review*, 90: 30–45.
——— ——— (2005). 'Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan', *American Economic Review*, 95: 1144–66.
——— ——— and Tootell, G. M. B. (2003). 'Identifying the Macroeconomic Effect of Loan Supply Shocks', *Journal of Money, Credit and Banking*, 35: 931–46.
Petersen, M. A. and Rajan, R. G. (1995). 'The Effect of Credit Market Competition on Lending Relationships', *The Quarterly Journal of Economics*, 110: 407–43.
Romer, C. D. and Romer, D. H. (1990). 'New Evidence on the Monetary Transmission Mechanism', *Brookings Papers on Economic Activity*, 1: 149–213.
Slovin, M. B., Sushka, M. E., and Polonchek, J. A. (1993). 'The Value of Bank Durability: Borrowers as Bank Stakeholders', *Journal of Finance*, 48: 247–66.
Strahan, P. E. (2009). 'Liquidity Production in Twenty-first-Century Banking', in Berger, A., Molyneux, P., and Wilson, J. O. S. (eds.), *Oxford Handbook of Banking*. Oxford: Oxford University Press.

CHAPTER 11

LENDER OF LAST RESORT AND BANK CLOSURE POLICY

XAVIER FREIXAS
BRUNO M. PARIGI[1]

Introduction

The financial crisis of 2007–9 redefines the functions of the lender of last resort (LOLR) of the twenty-first century—first, by placing it at the intersection of monetary policy, supervision, and regulation of the banking industry, and the organization of the interbank market, and, second, by extending its role to cover the possible bailout of non-bank institutions.

Since the creation of the first central banks (CBs) in the nineteenth century, the existence of a LOLR has been a key issue for the structure of the banking industry. The banking system has to provide mechanisms to manage banks' liquidity risk because one of the major functions of banks is to offer access to the payment system and facilitate property rights transfer, and because it is efficient to combine these functions with opaque long-term investments on the asset side (delegated

[1] We would like to thank Falko Fecht, several conference audiences, and the editors for their comments. We gratefully acknowledge funding from the Italian Ministry of University and Research, and jointly from the Italian and Spanish Ministries of University and Research through a grant Azioni Integrate/Accion Integradas 2006–8.

monitoring) and with demand deposits on the liability side (as justified by Diamond, 1984; Diamond and Dybvig, 1983; Calomiris and Kahn, 1991, or Diamond and Rajan, 2001). Although in any developed economy the principal mechanism to cope with both excesses and shortages of liquidity will be the interbank market, the well functioning of the banking system might still require an additional mechanism to avoid that both aggregate and bank-specific liquidity risk mismanagement results in a bank defaulting on its contractual obligations. The terminology 'of last resort' itself emphasizes that this institution is not intended to replace existing regular market mechanisms, but should make up for its possible, although infrequent, failures. This justifies the existence of a discount window in the US and the marginal lending facility in the Eurozone.

The basic objective of LOLR lending were first formulated by Thornton (1802) and Bagehot (1873) who argued that it was necessary in order to support the whole financial system and to provide stable money growth (Humphrey, 1989). Since then, the role of the LOLR has become a more controversial issue. The debate is inherent to the fact that, by providing insolvent banks with liquidity we are both allowing them to escape market discipline and promoting forbearance. Clearly, there is a consensus among academics and central bankers that a mechanism should exist to allow solvent banks to obtain liquidity if the interbank market fails to operate correctly. Also, everyone agrees that insolvent banks should not access the standard liquidity facilities and that, if necessary, their insolvency should be dealt with on a case-by-case basis. The problem arises because liquidity shocks affecting banks might be indistinguishable from solvency shocks. So, the debate about the role of the LOLR is connected with the efficient bank closure policy and, more generally, with the costs of bank failures and of the safety net.

This connection between the LOLR and bank bailout policy is not yet fully accepted. This may be due to the fact that access to liquidity, as well as the role of the LOLR, have evolved through history. Those accepting Bagehot's view of the LOLR may argue that it relates to a world where solvent banks were to be protected against sudden deposit withdrawals without the recourse to a well-developed repo market and without the CB privilege of issuing fiat money. With the emergence of a well-functioning repo market, today's conception of the role of the LOLR is completely different. The LOLR may step in exceptionally to prevent a collapse of the payment system that could be triggered by the lack of liquidity, but this should normally be dealt with by means of the appropriate monetary policy. So, *if* the money markets are well functioning, the LOLR should manage aggregate liquidity only and leave the issue of solvency to the market that will eliminate the lame ducks.

The critical step in this argument is the assumption of perfect money markets. Once we consider imperfect money markets we are forced to consider cases where it is impossible to distinguish whether a bank is solvent or insolvent. So, we have to acknowledge that in solvency cases the LOLR is sometimes acting to channel liquidity and therefore is improving the efficiency of the monetary policy framework,

while in the second case it is part of the safety net and directly related to the overall regulatory framework. Therefore, the design of an optimal LOLR mechanism has to take into account both the monetary framework and the banking regulation context.

The panic of 2008, which originated with the subprime crisis of 2007 in the US, illustrates vividly the new role of LOLR. Years of accommodating monetary policy, regulatory arbitrage to save capital, and waves of financial innovations—which by definition tend to escape traditional prudential regulation—have created the conditions for slack credit standards without the rating agencies calling for adequate risk premiums. The opacity of the assets of the banks and of the off-balance sheet finance vehicles created to hold mortgages have resulted at some point in a dramatic and sudden reappraisal of risk premiums. As with a thin market typical of the Akerlof lemon problem (Freixas and Jorge, 2008) financial intermediaries have become reluctant to lend to each other if not for very short maturities. The fear that the interbank market might not work well and might fail to recycle the emergency liquidity provided by the CBs around the world in various and coordinated ways has induced banks to choose the rational equilibrium strategy to hoard some of the extra liquidity instead of recycling it to the banks in deficit. Thus, channelling emergency liquidity assistance through the interbank market has not worked precisely because the interbank market is not functioning properly. To limit the systemic feedbacks of the sudden deleveraging of financial institutions in 2008 the Fed has taken the unprecedented steps of increasing the list of collateral eligible for CB discount lending, of extending emergency liquidity assistance to investment banks, government sponsored entities, money market mutual funds, a large insurance company, of entering swap agreements with other central banks to provide dollar liquidity to banks outside the US, and of acquiring bank capital. Preventing a complete meltdown of the financial system has required that the CB guarantees (and accepts potential losses) that most, if not all claims, on financial institutions will be fulfilled—which is more the resort of the bank regulatory authority than of the CB alone. As a consequence, the balance sheet of the FED has grown from about $900 billion in September 2008 to more than $2 trillion in December 2008, largely financed by the creation of high-powered money and by loans from the US Treasury. Discount window lending has grown from a few hundred million dollars under normal circumstances to more than $500 billion at the height of the crisis. In the four months from September to December 2008 the Fed has put more than $600 billion of reserves into the private sector (against September's 2007 total outstanding level of reserves of about $50 billion) in what Lucas (2008) has described as the boldest exercise of the LOLR function in the history of the Fed.

Consequently, we argue that the panic of 2008 has showed that it would be erroneous to adopt a narrow definition of the LOLR, stating that its role should be limited to funding illiquid but solvent depository institutions, while capital injections should be the Treasury responsibility. This would lead to a very simplistic

analysis of the LOLR functions, as the complex decisions would be either ignored or handed over to the Treasury. In our view, such a narrow view of the LOLR would create an artificial separation between lending by the LOLR at no risk and the closure or bailout decision by the Treasury that could lead to incorrect policy assessments.

On the contrary, the rigorous view of the LOLR that we take here has to be a broad one, encompassing the closure or bailout decision defining the LOLR as 'an agency that has the faculty to extend credit to a financial institution unable to secure funds through the regular circuit'. This definition omits any mention to the fact that the institution is illiquid or insolvent. Obviously, this does not preclude that a separation between LOLR and Treasury decisions might prove efficient. Yet, what it implies is that it has to be proven so. This broad definition has the additional benefit also to encompass the management of overall banking crises, which would be difficult to consider from the narrow perspective of pure liquidity provision.

Once we agree that the LOLR policy has to be part of the overall banking safety net, the interdependence of its different components becomes clear. First, the existence and the extent of the coverage of a deposit insurance system, as is well documented in Santos (2006), limits the social cost of a bank's bankruptcy, and, therefore, reduces the instances where a LOLR intervention will be required. Second, capital regulation reduces the probability of a bank in default being effectively insolvent, and so has a similar role in limiting the costly intervention of the LOLR. Third, the procedures to bail out or liquidate a bank, determined by the legal and enforcement framework, will determine the cost–benefit analysis of a LOLR intervention.

Clearly, the LOLR policy and its efficiency will depend upon the overall financial environment. When a liquid market for certificates of deposit (CDs), T-Bill, and securitized loans—or even simply for the loans themselves—exists, banks will only exceptionally encounter difficulties in coping with their liquidity shocks. Adopting a perspective of an all-embracing safety net does not mean that the safety net has to be the responsibility of a unique agent. Often, several regulatory agencies interact, because different functions related to the well functioning of the safety net are allocated to different agents. It is quite reasonable to separate monetary policy from banking regulation, and the separation of the deposit insurance company from the CB makes the cost of deposit insurance more transparent. Also, the national jurisdiction of regulation makes cross-border banking a joint responsibility for the home and host regulatory agencies.[2] This implies that regulation will be the outcome of a game among different agents that may cooperate or may be facing conflicts.

[2] In the Fall of 2008, the Belgium–Dutch banking and insurance conglomerate Fortis was rescued by a joint financial effort of the monetary and fiscal authorities of the Benelux countries, and the good offices of the President of the European Central Bank. This rescue was subsequently complicated by the Belgian courts rejecting the legal structure of the deal on the ground that shareholders rights had not been respected.

Finally, as part of the financial environment, the regulatory structure will be crucial. In particular, LOLR functions are usually attributed to the CB, while another institution, often the deposit insurance company, is in charge of closure. So, how the two decisions are coordinated is clearly an issue to be considered.

The rest of the chapter is structured as follows: in the second section we will examine the justification of LOLR lending in a simplified framework where only liquidity shocks arise. Then, the third section will consider contagion in the interbank market. The fourth section will be devoted to the case where liquidity shocks cannot be disentangled from solvency ones. The fifth section discusses the issues raised by the implementation and decentralization of the LOLR policy within the safety net. The sixth section concludes.

Pure liquidity shocks

As already mentioned, one of the major features of banks, and a justification of their existence, is that they combine assets with a long maturity with short-lived liabilities. As a consequence, an institution providing liquidity to the banking system has a key role in the well-functioning of the whole credit, deposit, and payment system. We will study here what types of liquidity shocks might affect banks and how emergency liquidity assistance (ELA) may help them cope with those shocks. Still, setting a framework that explains why banks may face liquidity risk does not mean that a LOLR should exist. First, it could be argued that monetary policy, jointly with peer monitoring, could solve the problem. Second, even if a specific institution is required, a private LOLR without any privileged access to CB liquidity could provide liquidity to the banks that need it.

We will first examine the different models of pure liquidity shocks, then turn to the analysis of a pure liquidity shock event, the disruption of the market as a result of the events of 9/11 2001, and close this section by discussing the pros and cons of a private LOLR based on the historical evidence.

Maturities transformation

The main motivation for LOLR in a modern economy is the need to prevent the threat of systemic risk whereby the crisis of one financial institution may affect others. In turn, the fragility of individual financial institutions stems from the very notion of the 'fractional reserve system' where short-term deposits finance illiquid long-term investments.

The classical models of Bryant (1980) and Diamond and Dybvig (1983) show that a bank that offers demand deposits, invests the proceeds in illiquid assets, and keeps an amount of liquidity equal to the expected value of the liquidity needs of its depositors, can offer a valuable insurance function to consumers who are uncertain about the time of their consumption needs. However, the transformation of maturities exposes the bank to the threat of bank runs if a large number of depositors decide to withdraw their money for reasons other than liquidity.

In this approach there are two possible equilibria. In the efficient one, depositors withdraw only to satisfy their interim consumption needs thus allowing the illiquid investment to mature. But, since the value of bank assets does not cover the contractual obligations of the bank with its depositors at the interim stage there is also an inefficient equilibrium, where it is optimal for all depositors to withdraw early (a run), even for those that have no immediate consumption needs. This may cause the 'fire sale' of long-term or illiquid assets, which, if generalized, may further depress asset values and cause a vicious circle as we will see in the sequel. Although deposit insurance and prudential regulation for many decades have essentially confined bank runs to textbook phenomena they have not completely disappeared even in sophisticated banking systems—witness the run on the deposits of the UK mortgage lender Northern Rock in 2007,[3] and in 2008 the runs on the deposits of, among others, IndyMack, a Californian bank, and of Washington Mutual, the largest US thrift.

The traditional way to address equilibrium selection is to imagine that depositors behave in one way or in another depending on an exogenous event (a 'sunspot', in the jargon of this literature). Since in one equilibrium, banks increase welfare, and in the other, they decrease welfare, the impossibility to establish which equilibrium will prevail makes it impossible to determine whether it is *ex ante* desirable that banks arise as providers of intertemporal consumption insurance. In other words, it is not clear why consumers would find it optimal to deposit their money in a bank in the first place. As a consequence, absent regulatory safeguards, policy recommendations are based on the assumption that a particular equilibrium will prevail, an issue that more recent modeling approaches using global games is not faced with, as we will see in the sequel. Despite this shortcoming, the Bryant–Diamond–Dybvig approach has been the modern draught-horse for the study of financial instability and systemic risk.

In a modern economy liquidity transformation takes on a different form from that envisioned in the classical Bryant–Diamond–Dybvig set-up. Two major changes that have occurred are relevant here. First, banks have dramatically lowered the fraction of their liquid assets over total assets. Goodhart (2008) observes that in the 1950s liquid assets were around 30 percent of British clearing banks' total assets

[3] This was the first bank run in Britain since 1886. Northern Rock was later nationalized by the UK Government.

while currently they are about 1 percent of total liabilities. Second, since long-term funding is more expensive than short-term funding, banks have funded an increasing fraction of long-term illiquid assets with short-term borrowing on the wholesale market. As a result, banks have replaced a relatively stable source of short-term funding like demand deposits with short-term interest-sensitive wholesale funding and rolling over debt. Brunnemeir (2008) observes that in 2006 and 2007 the short-term overnight repos were around 25 percent of the assets of brokers/dealers thus implying that the whole balance sheet must be refinanced every four days. These related changes have put a tremendous pressure on any financial institution in case of funding problem, as we will analyze in detail in the sequel.

One of the major features of the subprime crisis of 2007–9—the fact that with the widespread adoption of the so-called 'originate-to-distribute' model of banking, maturity transformation takes place in part off-balance sheet, and therefore escapes banking regulation and the traditional regulatory mechanisms to prevent runs—is to be considered also from this perspective: a liquidity crisis in a conduit or special purpose vehicle (SPV), that is funded through a rollover of short-term debt, is akin from the point of view of liquidity to a holding bank company with an unregulated subsidiary where bank runs can occur.

Systemic risk

Financially fragile intermediaries are exposed to the threat of systemic risk. Systemic risk may arise from the existence of a network of financial contracts from several types of operations: the payment system, the interbank market, and the market for derivatives. The tremendous growth experienced by these operations in the last decades increases the degree of interconnections among operators and among countries and thus the potential for contagion.

A number of papers have modeled contagion among banks and the ways to prevent it. The discussion will focus here on the two we consider most relevant. Allen and Gale (2000) show that financial contagion can emerge in a banking system of a multi-region economy. The interbank deposit market offers insurance against regional liquidity shocks but provides also a channel through which the shocks to the agents' preferences in one region can spread over other regions. Allen and Gale (2000) consider a version of the Diamond–Dybvig model with several regions in which the number of early consumers (the ones demanding liquidity at an interim stage) and late consumers fluctuate. An interbank market in deposits allows insurance as regions with liquidity surpluses provide it to regions with shortages. This constitutes an efficient mechanism provided there is enough aggregate liquidity. But, if there is shortage of aggregate liquidity the interbank deposit market can turn into the channel through which a crisis spreads. Facing a liquidity crisis, before liquidating long-term investments, banks liquidate their deposits in other banks, a

strategy that in the aggregate just cancels out. In case of shortage of aggregate liquidity the only way to increase consumption early is eventually to liquidate long-term investments. A financial crisis in one region can thus spread via contagion. Note that the nature of the crisis, and of the solution, is different with respect to the market for retail deposits as, for example, in the Diamond–Dybvig model. In the retail market runs occur because banks liquidate when they have insufficient liquidity to meet the fixed payment of the deposit contracts. Hence, by making the contracts contingent or discretionary, incentive to run can be eliminated. In the interbank markets instead, the reciprocal nature of the deposit agreements makes these solutions impossible. Moreover, the likelihood that contagion happens depends on the architecture of the interbank deposits. If each region is connected with all the others the initial impact of the crisis can be attenuated and contagion avoided. On the contrary, if each region is connected with few others the impact of the initial crisis may be felt strongly on the neighbouring regions.

Using the notion that participants in financial markets may have different beliefs Castiglionesi (2007) extends the basic framework of Allen and Gale (2000) to ask if there is any instrument to avoid contagion. He argues that contagion is due to the impossibility to sign contingent contracts on unforeseen contingencies. This happens because the CB and market participants (banks and depositors) disagree on the prior of certain events; in particular they don't think that aggregate liquidity shortage is possible, hence they don't write contracts contingent on this event; the CB instead believes that with positive although small probability aggregate liquidity shortage is possible and can thus improve matters by imposing reserve requirements.

In Freixas, Parigi, and Rochet (2000) a system of interbank credit lines arises because depositors face uncertainty about where they need to consume. Financial connections reduce the cost of holding liquidity but make the banking system prone to experience speculative gridlocks even if all banks are solvent. The mechanism of the gridlock is the following: if the depositors in one location, wishing to consume in another location, believe that there will be not enough resources for their consumption at the location of destination, their best response is to withdraw their deposits at their home location. This triggers the early liquidation of the investment in the home location, which, by backward induction, makes it optimal for the depositors in other locations to do the same. The CB can play a role of crisis manager—when all banks are solvent the CB's role is simply to act as a coordinating device by guaranteeing credit lines of all banks. Since the guarantees are not used in equilibrium this action entails no cost. When instead one bank is insolvent because of poor returns on its investment the CB has a role in the closure of this bank, which has to be conducted in an orderly fashion to maintain the well-functioning of the payment network despite the closure of one bank.

Both Allen and Gale (2000) and Freixas, Parigi, and Rochet (2000) emphasize the key role the interbank market plays in propagating a crisis through the intertwining of their balance sheets, the default of one bank generating an immediate

loss to all its unsecured creditors. Both emphasize that the structure of payments, with more or less diversification or more or less relationship lending will be a key characteristic of the resilience of the banking system. Yet, from a policy point of view, the two models have a crucial difference. In Allen and Gale any CB emergency liquidity injection allows solving the crisis no matter where the liquidity is injected, as it is profitable for one liquidity long institution to lend to a liquidity short one. In Freixas, Parigi, and Rochet (2000), since the crisis does not originate in an unpredicted liquidity shortage but in a rational alternative equilibrium strategy for depositors, injecting additional cash in the aggregate will not help. Even in the case where every bank has access to sufficient liquidity, the inefficient gridlock equilibrium exists where banks resources are used in an inefficient way. Solving the crisis in the Freixas, Parigi, and Rochet (2000) model is more the resort of the bank regulatory authority than of the CB, as it requires guaranteeing that all claims on banks will be fulfilled. So, despite apparent similarities, the LOLR has a role of liquidity provider in the Allen–Gale model, while it has a role of crisis manager in the Freixas–Parigi–Rochet one.

The panic of 2008 offers a clear example of the distinction between systemic risk in Allen and Gale (2000) and Freixas, Parigi, and Rochet (2000). The fear that the interbank market might not work well and might fail to recycle the emergency liquidity provided by the CBs around the world in various and coordinated ways has induced banks to choose the rational equilibrium strategy to hoard some of the extra liquidity instead of recycling it to the banks in deficit. The resulting equilibrium resembles closely the gridlock described by Freixas, Parigi, and Rochet (2000) where the fear that a debtor bank will not honour its obligations induces the depositors of the creditor bank to withdraw deposits thus triggering the liquidation of assets in a chain reaction style. This is the modern form of a 'bank run' where financial intermediaries refuse to renew credit lines to other intermediaries thus threatening the very survival of the system.

In a logic similar to the previous models, another way in which the CB can prevent a crisis is to reallocate toward the correct risk the liquidity insurance of agents particularly sensitive to extreme events. Caballero and Krishnamurthy (2007) construct a model of the benefits of CB intervention during flight to quality episodes induced by such preferences. Agents deposit their wealth in financial intermediaries that insure them against shocks. Agents' preferences, however, do not exhibit the traditional risk aversion characteristics. Rather, agents are uncertainty-averse in the sense that they are uncertain about the functioning of the economy. This is modelled assuming that agents maximize the minimum expected utility that they receive. In other words they are extremely sensitive to extreme events. Extreme events are modelled as repeated waves of shocks that induce the agents to demand liquidity. Each agent is concerned about the scenario in which he is the last one to receive a shock and there is little liquidity left. Since they all have max–min preferences, this induces an upward bias in the probability

of the worst-case scenario. When aggregate liquidity is insufficient agents waste valuable liquidity by self-insuring against worst-case scenarios which are impossible.

A CB with different preferences can improve matters. The CB has no more information than the private agents but just sees the world through a different lens—namely, does not incorporate in its objective function the worst-case probability assessments of the private agents. Its objective function allows seeing the world from the aggregate which makes it apparent that the worst-case scenario that all agents are guarding against is impossible. Thus CB policy works by reducing the agents' 'anxiety' that they will receive a shock when liquidity is depleted.

Building on their work, Caballero and Krishnamurthy (2008) offer insights on the development and the propagation of the panic of 2008 and what the CB policy should be. At the start of 2007, banks were well capitalized and liquid and the price of risk was quite low by historical standards. The subprime segment of the mortgage industry was widely known to be risky and defaults were expected. They argue that the root of the crisis is not in subprime losses, that in the worst case estimates were around $250 billion, large for any individual financial institution but quite small with respect to any measure of global financial wealth. Rather the root cause of the crisis has been the increase in uncertainty related to the *complexity* (see also Gorton, 2008) of the financial instruments and derivatives structures traded in financial markets, combined with *lack of historical records* that would offer a guide on how these instruments would behave under stress. These two factors—complexity and lack of history—argue Caballero and Krishnamurthy (2008), have induced investors to start questioning the valuation of a myriad other credit products structured in much the same way of subprime mortgages. This increase in uncertainty has led investors to make decisions based on imaginary worst-case scenarios and hoard liquidity to cover worst-case shocks that may never occur, as in the Freixas, Parigi, and Rochet (2000) model. Thus, this view allows us to understand why, although the subprime shock was a small one, the actions of panicked investors made it large. Caballero and Krishnamurthy (2008) argue that in this case the standard recipe for the CB is credibly to promise the injection of large amounts of liquidity in the event of a meltdown of the financial system. However, the credibility of these interventions introduces a new layer of uncertainty if the markets perceive that the CB runs out of policy ammunitions or is not ahead of the game, either for political or other reasons.

The LOLR and liquidity shocks: the 9/11 case study

An important criticism on the classical view of the LOLR in today's financial market has been raised by Goodfriend and King (1988). They argue that the existence of a fully collateralized repo market allows CBs to provide the adequate amount of liquidity which is then allocated by the interbank market. Since individual interventions would

no longer be necessary, the discount window is made obsolete. Well-informed participants to the interbank market are capable of distinguishing between illiquid and insolvent banks. These arguments have been so influential that the Bagehot view of the LOLR is often considered obsolete in well-developed financial markets. Yet, Goodfriend and King's argument contradicts the asymmetric information assumption that is regarded as the main justification for financial intermediation. Goodfriend and King's argument is even less attractive if we consider Goodhart's (1987) criticism that liquidity and solvency shocks cannot be disentangled.

Here we will begin with the analysis of the conduct of the LOLR in a particular instance that offers a clear example of a system-wide liquidity shock. The liquidity effects of the events of 9/11 2001 illustrate well the systemic threats posed by the interdependencies in payment flows even in the absence of solvency shocks. McAndrews and Potter (2002) make the point that on 9/11 banks experienced severe difficulties in making payments because of the widespread damage to the payments' infrastructure. The nettable nature of payment flows allows banks to operate in the Fedwire system—a real-time gross payment system—with an amount of reserves which is about 1 percent of their total daily payments, with the rest coming from the inflows of payments from other banks. This high velocity of circulation exposes the system to great risk if the normal coordination and synchronization of payments collapses as it happened on 9/11. The events of that day resulted in an uneven distribution of liquidity in the banking system: while some banks unable to send payments accumulated higher-than-desired balances, other banks' increased uncertainty about the flow of incoming payments led them to raise their precautionary demand for liquidity. McAndrews and Potter (2002) observe that the incident that triggered the liquidity shortfall was well known to all market participants and was generally perceived as a pure liquidity shock, unrelated to the fundamental solvency of any major financial institution. However, the fear of a systemic threat due to the breakdown of the coordination mechanism that banks use in their normal handling of payment flows induced the Fed to act. McAndrews and Potter (2002) and Coleman (2002) document that, on 9/11 and in the following days the Fed took a number of steps to make sure that market participants would know that the Fed was ready to provide the liquidity that the market demanded. The Fed released a statement encouraging the banks to borrow from the discount window with the result that discount loans grew from $200 million to $45 billion on 12 September; it waived daylight overdraft fees and overnight overdraft penalties so that overnight overdraft increased from an average of $9 million in August 2001 to more than $4 billion on 12 September; later on, with markets beginning to function better, the Fed increased liquidity in the interbank market via open-market operations (OMO) from $25 billion to $100 billion. The Fed did not simply inject liquidity; it also invited the banks to benefit from the discount window by lifting the stigma that is usually attached to this type of borrowing. Still, it could be argued that, had all the operations been channelled

through the open market, the effect would have been the same. If so, it would be impossible to see if the liquidity crisis was of the Allen and Gale type or of the Freixas–Parigi–Rochet type.

Nevertheless, McAndrews and Potter (2002) point out another important lesson from these events that help discriminating between the two models. Banks that are reluctant to pay one another are also reluctant to lend to one another. Thus, in these circumstances, injecting liquidity through OMO, as advocated, for example, by Goodfriend and King (1998), may be ineffective at redistributing balances because the additional funds may not be circulated where needed, contrary to discount window interventions. Only once coordination among banks has been re-established OMO may be preferred as they leave to the market the task to allocate liquidity.

That lending to the market via OMO may not be completely effective if banks are reluctant to lend to each other as they fear hidden losses in their counterparties balance sheets has been quite evident during the crisis of 2007–9, as we see later.

To reinforce the previous point notice that the response of the Fed to the events of 9/11 2001 was facilitated by the fortuitous circumstance that the US banking system was in a relatively healthy condition at the onset of the crisis so that the Fed did not have to be selective about the account holders through which to channel reserves. Rather, the Fed could lend freely to solvent banks according to the LOLR principles articulated by Bagehot, consistent with one of the missions of the Federal Reserve Act: 'to furnish an elastic currency' (Lacker, 2004).

Private LOLR

Although central banks and their LOLR functions are relatively new institutions, the financial history of the US before the creation of the Federal Reserve System in 1913 offers good examples of private arrangements to solve bank crises—namely, the commercial bank clearing houses (CBCs) (See Gorton, 1985 and Gorton and Mullineaux, 1987 for a detailed analysis of CBCs). Originally developed to facilitate check clearance, the CBCs became organizations that performed a variety of tasks. During bank panics, the CBC ceased to behave as an authority regulating competing banks and instead effectively combined the member banks into a single organization, with the group accepting corporate liability for the debts of each individual member. Among the most significant actions of the CBC during a bank panic were the suspension of the publication of individual banks' balance sheets and the publication instead of aggregate balance sheet information for the clearing house as a whole, the suspension of convertibility of deposits into currency, and the issuance of loan certificates. Loan certificates were liabilities of the clearing house that member banks could use in the clearing process and could circulate as currency. These loan certificates, issued up to a fraction of the market value of the assets of the member bank seeking them, were in effect fiat money of the clearing house.

In the US, cooperation among banks produced stable interbank relationships, which in the case of the Suffolk system—an important example of a self-regulating bank clearing system operating throughout New England from the 1820s to the 1850s—were even more resilient than anticipated by their proponents (Calomiris, Khan, and Kroszner, 1996). Many observers pointed out that the Federal Reserve System was a development of the existing CBCs (White, 1983; Gorton, 1985; Calomiris, Kahn, and Kroszner, 1996; Timberlake, 1978; and Timberlake, 1993). However, one criticism of the functioning of the CBCs was that their membership criteria were too stringent and designed to reflect only the interest of the member banks, not the public interest. For example, the New York Clearing House Association demanded a very high level of reserves to qualify for membership so that many banks preferred to opt out of the clearing system (Sprague, 1910). In the panic of 1907, a solvent Trust Company, the Knickerbocker Trust—which did not belong to any CBC—was forced to suspend as a result of liquidity problems. As argued by Friedman and Schwartz (1963: 159), 'Had the Knickerbocker been a member of the Clearing House, it likely would have been helped, and further crisis developments might thereby have been prevented'. The consequence was one of the severest contractions in US economic history, and the impetus for the founding of the Federal Reserve System.

The example of the US CBCs shows that the financial linkages that expose banks to contagion threats may allow agents to obtain *ex post* mutual insurance in the form of private bailouts even though formal *ex ante* commitments are impossible. A recent study by Leitner (2005) provides a model that shows that linkages that create the threat of contagion may be optimal. Assume that the project of an agent can succeed only if he and the other agents whom he is linked to make a minimum level of investment. Since an agent's endowment is random he may not have enough resources to make the necessary investment. His inability to commit to repay may prevent him from borrowing against future cash flow or from entering an insurance contract *ex ante*. In the absence of contagion through financial linkages, agents with high endowments have no incentives *ex post* to lend to agents with low endowments; hence, some positive net present value projects are not realized. But, if agents are linked to one another, then those with high endowments have the incentive to bail out those with low endowments, because, if they don't, all projects fail by contagion. Financial linkages, thus, can motivate banks to help one another even in cases in which they could not commit to do so *ex ante*. The consolidation of the balance sheets of the member banks of the CBC in crisis time, and the issuance of loan certificates drawn on the CBC, by providing *ex post* mutual insurance, seems to have performed privately modern LOLR functions.

So, the evidence seems to indicate that, as expected, CBCs are more concerned about their own narrow interest than about the risk of contagion that may result from the bankruptcy of a bank outside their network. This is why a LOLR should have a mandate of preserving financial stability and should therefore encompass all

banks—not just the ones affiliated with its network. The question of semi-public institutions has not been the object of much research. Yet, during the subprime crisis of 2007 it has been argued that the Federal Home Loan Banks (FHLBs) have massively lent to the banking industry and, by so doing, avoided an exacerbation of the mortgage crisis.[4] In retrospect, it is clear that the impact of FHLBs Fannie Mae and Freddie Mac in fueling the mortgage market by purchasing the securitized loans has been to postpone the exacerbation of the mortgage crisis rather than to avoid it.

Pure solvency contagion: The domino effect

Although it is clear that systemic risk increases in situations of financial fragility, the issue of contagion is characteristic of the banking industry. The 'domino effect' is the evocative term used to illustrate it, and with good reason. Banks are related to one another through a network of assets and liabilities, and a joint reputation. When assessing the cost–benefit of a LOLR operation, contagion and systemic risk will be the first factors to be considered. Central banks have been clear in asserting that they will bail out banks that are systemic, thus comforting the market prevailing view that banks are not equally treated, as some banks are 'too big to fail'. The important issue is here to assess whether contagion is a myth or a reality.

From that perspective, it is important to start by reviewing the literature on contagion, and then point out some issues that affect the measure of contagion. Since the prevention of systemic risk is one of the main rationales behind the LOLR, it is important to assess and quantify it. Unfortunately, lack of data availability has limited so far the analysis. Researchers have thus resorted to study particular market segments or made particular assumptions about bilateral banks' exposures, and have concentrated on specific countries. Typically, they take as given the failure of a bank and track its effects through the banking system. Although no general conclusions arise, most studies show that the failure of one bank may have significant knock-on effects on others, depending on the architecture of the interbank links, the concentration of the banking industry, the extent of cross-border banking, and the presence of de facto safety nets.

[4] 'FHLB advances rather than borrowing from the discount window reflect several factors: 1. the lower cost of FHLB advances; 2. the ability to borrow at longer terms from the FHLB; and 3. the lack of stigma in using FHLB advances as a source of funding.' William Dudley, Executive Vice-President of the New York Fed, 17 October 2007, 'May You Live in Interesting Times'.

Computing cross-banks contagion

Humphrey (1986) was the first to investigate the extent of contagion by using data from the Clearing House Interbank Payments System. His simulation of the failure of a major participant in the payment system showed that this could lead to a cascading effect. More recently, estimates of contagion in the US federal funds market were obtained by Furfine (2003) that studied the particular segment of settlement data to compute bilateral exposures in the US federal funds market. He found that contagion is quite limited: even in the worst-case scenario of the failure of the largest bank with a 40 percent loss given default, only between two and six banks fail, with 0.8 percent of total bank assets involved. Since his data incorporate only federal funds transactions, which account for only 10–20 percent of total interbank exposures, the estimates of contagion are potentially conservative. However, illiquidity presents a greater threat: if a large federal funds debtor becomes unable to borrow, illiquidity could spread to banks representing almost 9 percent of the US banking system by assets.

One characteristic of modern real-time payment systems is the high velocity of circulation of payments which allows economizing on the stock of reserves but exposes the systems to risk of disruptions. Besides the disruptions to the physical infrastructure identified by McAndrews and Potter (2002) and discussed above, there is a second endogenous response of participating banks that makes high-value payment systems vulnerable. Afonso and Shin (2008), using lattice-theoretic simulation methods, show that precisely as the synchronization and coordination of payments of real-time systems creates a virtuous circle of high-payment volumes, the decline in a bank's willingness to pay and the decision to postpone payments to the end of the day to conserve liquidity can cause an increase in the demand of intraday liquidity and disruption of payments. Afonso and Shin (2008) show that when a bank is identified as vulnerable to failure other banks may choose to stop payments to that bank with systemic repercussions. This chain of events probably played a role in the disruption of interbank wholesale funding and in the near meltdown of the major US investment banks in the middle of September of 2008, leading to the collapse of Lehman Brothers: nobody wanted to do business with a company that could fail.

Studies conducted for smaller economies or with a bigger size of cross-border transactions show a more pronounced risk of systemic repercussions. In particular, Blavarg and Nimander (2002) study the impact that the failure of one of the four largest Swedish banks may have on the rest of the system. They find that such a failure could push Tier 1 capital of the three other banks below 4 percent and hence force their closure. An even more dramatic effect could result from the failure of the largest foreign counterparty.

Wells (2004) and Upper and Worms (2004) use a similar methodology to study contagion, starting with estimates of the matrix of bilateral exposures of banks in

the interbank market. However, since one can only observe each bank's total interbank claims and liabilities, it is not possible to estimate such a matrix without imposing further restrictions. Their approach is to choose a distribution that maximizes the uncertainty (the 'entropy') of these exposures. The result is that these exposures reflect the relative importance of each institution in the interbank market via the size of its total borrowing and lending. With this methodology, Wells (2004) studies the effect of the sudden and unexpected insolvency of a single bank in the UK. He shows that this can lead to a substantial weakening in the capital holdings of other banks, but in most cases does not result in additional bank failures. By assuming complete loss given default, his model shows that, in the extreme cases, single-bank insolvency could trigger knock-on effects leading in the worst case to the failure of up to one-quarter of the UK banking system. At the same time, a further quarter of the banking system would suffer losses amounting to more than 10 percent of their Tier 1 capital. For loss given default levels of less than 50 percent, contagion affects, at worst, less than 1 percent of total banking system assets. However, even with low losses given default, a narrow shock can considerably reduce the capital reserves of many banks. Furthermore, if the initial shocks hit during a period where the banking system is already weakened (e.g., during a recession), the effect of contagion can be more pronounced.

Upper and Worms (2004) use German banks' balance sheet information to estimate a matrix of bilateral credit exposures. They find that interbank lending is relatively concentrated and, as the theoretical literature suggests (Allen and Gale, 2000 and Freixas, Parigi, and Rochet, 2000) this makes contagion a real possibility. Despite the presence of institutional guarantees that prevent the failure of the savings and cooperative banks sector, there is considerable scope for contagion as the failure of a single bank could lead to a break down of up to 15 percent of the banking system in terms of assets.

Changing patterns in cross-bank contagion

Lack of data availability has forced previous researchers to conduct simulations based on bilateral exposures on a short period of time, while contagion risk has evolved over time as function of banking consolidation and the increased importance of cross-border banking. The availability of detailed confidential bank balance sheet data about Belgium allows Degryse and Nguyen (2006) to go beyond the existing literature on several respects. First, using time series data on interbank exposures they examine the evolution over time of the contagion risk associated with the failure of a Belgian bank. Second, they conduct a regression analysis that identifies the major determinants of contagion. They find that a move from a 'complete' structure—one where each bank lends to each other—toward a 'multiple money centers' bank structure and the increase in concentration in the

lending market decreases domestic contagion. They also find that an increase in the proportion of cross-border assets decreases the risk and the impact of domestic contagion. These results contrast with those of Mistrulli (2005) for the Italian interbank market. He finds that the importance of cross-border exposures has decreased and that the transition from a 'complete' toward a 'multiple money centers' structure has increased contagion risk. Part of the contrasting results can be attributed to the different methodologies, as the Mistrulli's study is based on simulations while Degryse and Nguyen use regression analysis which allows sorting out the different sources of contagion.

An alternative approach

The exposure of the LOLR to systemic risk can be quantified using standard risk management techniques that take into account the correlations between banks' assets portfolios. While most studies take the probability of default as given and trace the impact of a bank default on the rest of the system, a new methodology proposed by Lehar (2005) allows estimating the probability of default and pricing the liabilities of the regulators with respect to the banks as contingent claims in the classic framework of Merton (1977). Assuming that banks' assets values follows a geometric Brownian motion one can link equity price, asset values, and bank debt in a standard equation that sees equity (whose price is observable from the market) as a call option on banks' assets with a strike price equal to the notional value of bank debt (observable from the banks' balance sheet). Using a maximum-likelihood estimation procedure the value of the banks' assets is then obtained and the exposure of the regulator computed. Lehar (2005) uses a sample of 149 international banks from 1988 to 2002 to identify the banks with the highest contributions to systemic risk and the countries which threaten the stability of the global financial system. Correlations of North American banks' asset portfolios have increased but the systemic risk of the North American banking system has decreased over time as banks have increased their capitalization. Instead, the capitalization of the Japanese banks has declined dramatically causing that system to become very unstable. Not surprisingly, the estimated regulator's liabilities increased sharply at the time of the Asian crisis in 1997–8.

Is contagion a myth?

On the basis of the previous estimates, one might have been tempted to conclude right before the crisis of 2008 that contagion was a myth. The exposures of banks one to another are limited and should not be a major concern for the regulator. But, taking the estimates at their face value without considering the whole contributions

that the theoretical models allow us to make would be an oversimplified view of contagion. From this point of view, several important criticisms should be formulated. They concern the indirect contagion through the behavior of depositors, the business cycle, the price of bank assets during a crisis, and the impact of liquidity.

First, the empirical evidence is based on the network of banks' assets and liabilities. Still, it may well be the case that the failure of one bank implies contagion through demand depositors. Their rational updating of the chances of another bank of similar characteristics may lead them to withdraw their deposits in a fly to quality. Many banking crises illustrate this phenomenon, as during the Great Depression in the US, or in the ethnic bank crisis in the aftermath of BCCI in the UK. The default of the Madhavpura Mercantile Cooperative Bank in India in 2001 was used as a case study by Iyer and Peydró (2006) to examine the contagion taking place through demand deposits. Still, the overall analysis of the joint impact of a bank failure through the network of reciprocal liabilities and through depositor's reaction remains to be done.

The second remark is that the measure of contagion is different in good times and in bad times. The impact of an individual bank when the banking system is healthy is the object of the above analysis. Yet, from the policy analysis point of view, it is not clear that this is the best measure of contagion. An individual bank is more likely to go bankrupt when all banks are in trouble. This, of course, makes the analysis much complex, because, in such a case, contagion-induced and macroeconomic-induced systemic risk are simply undistinguishable. So, the new challenge in the measurement of contagion would be to try to compute the impact of a bank bankruptcy conditionally on the banking sector health, using some measure of the loan impairment in the banks' loan portfolio.

A third remark is the impact of a number of bank failures or large reduction of the size of their balance sheet on the value of assets. The main impact, first identified by Irving Fisher (1933), concerns the price of assets that are used as collateral. In a debt deflation situation, the value of assets decreases and this lowers the amount of collateralized loans, and therefore the amount of available credit, which, in turn, reduces output. This output decline will again impact into the price of assets, thus leading to a further reduction in asset prices until outside investors buy the assets (see also Kyotaki and Moore, 1997 and, more recently, Gorton and Huang, 2004 and Acharya and Yorulmazer, 2007).

Fourth, a number of studies show how the impact of a relatively small shock, like defaults on subprime mortgages in 2007, may be amplified in a full-blown financial crisis in a way that cannot be captured by domino models of financial contagion—for example, like Freixas, Parigi, and Rochet (2000). As pointed out by Adrian and Shin (2008), an implication of domino models is that if bank A has borrowed from bank B, and bank B has borrowed from bank C, and so on, should bank A default then bank B would suffer a loss. If the loss is large enough, B's capital will be wiped out and also B defaults. But, if the loss is small relative to the

size of the capital of the banks—as it was in the case of defaults on subprime mortgages—it can easily be absorbed by the capital of the financial sector. Furthermore the widespread use of securitization has further spread risk among financial institutions making it easier to absorb potential losses.

An alternative to domino models of financial contagion is offered by Brunnemeir and Pedersen (2008) and Brunnemeir (2009), who argue that liquidity spirals may cause aggregate liquidity to dry up as a result of minor shocks. If leveraged investors suffer even minor losses on their assets, in order to maintain the same leverage they have to sell assets, hence contributing to depress asset prices even further if *market liquidity* for the asset is low. In addition to this loss spiral, Brunnemeir and Pedersen (2008) identify a margin spiral arising from the fact that, typically, financial assets are purchased on credit (*funding liquidity*) using the purchased assets as collateral (margin) for the loan, often a short-term one. This exposes the investors to the risk that the fall in value of the assets triggers a reduction of the margin and a risk that the rollover of short-term borrowing becomes more costly or impossible. The margin spiral reinforces the loss spiral as investors suffering losses have to sell assets to meet higher margin demands—that is, to lower the leverage ratio. Adrian and Shin (2008) confirm this spiral empirically for the five major US investment banks in the period 1997 to 2007. They identify a strong positive relationship between the value-weighted change in leverage and the change in assets, hence showing that leverage is highly procyclical. The amplification of the financial cycle that arises from the balance sheet of financial intermediaries helps explain, according to Adrian and Shin (2008), how modest losses on US subprime mortgages triggered the most severe financial crisis since the Great Depression.

Finally, and more tentatively, the cross-banks link could be underestimated as well if we restrict the analysis to solvency. In fact, a bank lending overnight to a peer financial institution that happens to be in default may not be fully satisfied with the knowledge that it will recover 95 percent of its claims in five years, after the liquidation of the failing institution is complete. This may trigger the lending bank to liquidate some of its assets later at 'fire sale', possibly increasing the impact on the price of assets. Recently, the possibility of contagion from the asset side of interlinked balance sheets has received explicit attention in the literature. Acharya and Yorulmazer (2007) examine the efficient closure/bailout policies and find that if the number of banks facing distress is large enough, it may be *ex post* optimal for the regulator to bail out some failed banks or alternatively to provide liquidity assistance to surviving banks in the purchase of failed banks. In both cases, the LOLR has to step in, as in Allen and Gale (2000) or Freixas, Parigi, and Rochet (2000). Schnabel and Shin (2004) and Cifuentes, Shin, and Ferrucci (2005) show that changes in asset prices may interact with solvency requirements or with internal risk control and amplify the initial shock in a vicious circle in which the reduction of the value of a bank's balance sheet may force the sale of assets or the disposal of a trading position, further depressing asset prices as illustrated above.

This point appears particularly relevant in the current crisis. While contagion was expected to occur through the interlinkages between the different banks assets and liabilities, it occurs through the financial institutions lack of liquidity. The lack of liquidity led banks to sell some of their assets; this, in turn led to a decrease in the value of those assets. Banks were, therefore, confronted with losses and an increase in their risk, thus leading to a reduced solvency. Thus, it seems that during the current crisis assets liquidity has been the channeling vehicle for solvency shocks to be transmitted from one bank to another.

The awareness of this risk is linked to a number of steps taken by the regulators to soften liquidity requirements in the face of crisis. Thus, for instance, the Financial Services Authority (FSA) responded to the decline in stock prices in the Summer 2002 by diluting the solvency test for insurance companies and in 1998 the Fed orchestrated the rescue of the hedge fund LTCM to prevent the negative impact of asset values that would have resulted from the unwinding of its positions.

Regarding the extent of contagion, it is clear that whatever the form it takes it will depend on macroeconomic conditions. The term 'macroeconomic fragility' has been coined precisely to express the vulnerability of financial institutions at some point in time, when a systemic crisis is possible. The extent of macroeconomic fragility depends upon a number of macroeconomic factors that are present in the current crisis, such as asset bubbles, linked to exuberant expectations, expected stable interest rates, and expected high growth. Still, it also depends on the level of procyclicality of the financial system. On this point it should be mentioned that the combination of the accounting standards that impose marking assets to market and Basel II has made the supply of credit more procyclical. Also, banks have neither taken into account the risks they were generating, nor have the regulatory authorities (possibly with the exception of the Bank of Spain) increased the required capital level—a prerogative they have under the pillar two of Basel II, to take into account the increased level of risk. The Bank of Spain, instead, in full defiance of International Financial Reporting Standards (IFRS), made Spanish banks abide by a dynamic provisioning rule that forced them to make provision at the time when the credit was granted, and not at the time when the loan becomes distressed.

Distinguishing between insolvent and illiquid banks

The difficulty to distinguish between an illiquid and an insolvent bank has been acknowledged at least since Bagehot's Lombard Street, when he argued 'Every banker knows that if he has to prove that he is worthy of credit, however good may be his

arguments, in fact his credit is gone'. Modeling such a framework has been done only recently. Two different approaches are possible: one based on unobservable liquidity and solvency shocks and the other based on the coordination of interbank market lenders' strategic responses to fundamental, public and private solvency signals.

Unidentifiable shocks

The difficulty of sorting out liquidity and solvency shocks stems also from the unique position that banks have in creating aggregate liquidity. Diamond and Rajan (2005)—building on their previous work (Diamond and Rajan, 2001)—argue that banks perform two complementary functions: they have loan-collection skills without which borrowers could not credibly commit to repay their loans, and they issue demand deposits to commit not to extract rents from investors. If a sufficiently large fraction of a bank's portfolio needs refinancing (a solvency problem) the bank will be unable to borrow against its future value. But, in that case, there will be a shortage of liquidity in the economy to fund current consumption (a liquidity problem). A solvency problem or a liquidity problem alone can lead to a run on a bank if depositors anticipate losses. A run, in turn, destroys a bank's ability to extract money from borrowers and thus the ability to channel funds from surplus agents to those in need to consume. Thus, after a run, aggregate liquidity is destroyed (an effect not present in bank runs of the type described in Diamond and Dybvig, 1983) and liquidity is also trapped in the wrong place, very similarly to what happened during the panic of 2008; hence, the difficulty of distinguishing between illiquid and insolvent banks. The appropriate policy response depends on the cause of the problem. When the source of the problem is a liquidity shortage, Diamond and Rajan (2005) advocate to lend freely to prevent a drop in the money stock. When solvency is the problem, their advice is to recapitalize banks. Recapitalization, however, can be harmful if the problem is lack of liquidity since capital infusion will simply push interest rates up, potentially causing more bank failures. Liquidity infusion instead has the least downside and thus it fits the test of doing no harm.

The approach followed by Freixas, Parigi, and Rochet (2004) is also based on the impossibility of distinguishing illiquidity from insolvency. They consider that banks are confronted with shocks that may come from uncertain withdrawals by impatient consumers (liquidity shocks) or from losses on the long-term investments that they have financed (solvency shocks), and that the two types of shocks cannot be disentangled. In acting as a LOLR the CB faces the possibility that an insolvent bank may pose as an illiquid one and borrow either from the interbank market or from the CB itself. Then the bank may 'gamble for resurrection'—that is, it may invest the loan in the continuation of a project with a negative expected net present value. This assumption is in line with the criticism of the LOLR during the

S&L crisis in the US during the eighties, and justifies why CBs are reluctant to be more liberal in their use of ELA. This setting allows the authors to focus both on the incentive issues of ELA and under which macroeconomic conditions the CB should provide ELA, at the cost of abstracting from modelling contagion. In periods of crisis, when banks' assets are very risky, borrowing in the interbank market may impose a high penalty because of the high spread demanded on loans. Freixas, Parigi, and Rochet (2004) show that ELA should be made at a penalty rate so as to discourage insolvent banks from borrowing as if they were illiquid, but it should happen at a rate lower than the interbank market. The reason the CB can lend at a better rate than the market is that the CB can lend collateralized and thus override the priority of existing claims. By penalizing insolvent banks that demand ELA, the CB provides banks with the appropriate incentives to exert effort to limit the probability that a bank becomes insolvent in the first place.[5] The implications of this approach can be clearly seen in the assessment of the 2007–8 crisis. The classical view of the interbank market, according to which the interbank market works perfectly, was that the spreads on interbank loans were understating risk, and that the observed turmoil was a correction in pricing on all assets and contracts that depended on the price of risk: real estate, mortgages, and unsecured loans to banks. Instead, the Freixas, Parigi, and Rochet (2004) approach views the crisis as a joint one of liquidity and solvency, so that, absent CB intervention, the interbank market may exacerbate the adverse selection problems. Taking the argument to the extreme, as modeled, for instance, in Freixas and Holthausen (2005) or Freixas and Jorge (2008), this may lead to a thin-market equilibrium as in the classical market for lemons. The policy implications are vital, since if the differential diagnostic is a correction back to the long-term price of risk, the optimal policy may be for the CB not to intervene except in so far as to reduce the cost of bank failure. If, instead, adverse selection in the interbank market leads to a stand still, then the LOLR liquidity provision to individual institutions is capital.

The global games approach

Rochet and Vives (2004) provides a theoretical foundation of Bagehot's doctrine in a modern context. They shift the emphasis from maturity transformation and liquidity insurance of small depositors to the modern form of bank runs where large well-informed depositors refuse to renew their credits in the interbank market, as in the market freezing of 2008. Building on the theory of global

[5] However, when the Fed made an emergency senior loan to the largest US insurance company AIG in the middle of September 2008 it caused a 40% loss for AIG's junior creditors. This may well be one of the reasons why unsecured interbank markets fail as the CB's interventions have diluted the value of pre-existing loans dramatically.

games, they investigate the optimal behavior of bankers that observe noisy signals about banks' fundamentals. This allows obtaining a unique equilibrium, in contrast with the Diamond–Dybvig classical result. The global games approach overcomes the problem of equilibrium selection (Carlsson and Van Damme, 1993 and Morris and Shin, 1998) by linking the probability of occurrence of a crisis to both the fundamentals and the information of depositors. The proof of uniqueness of the equilibrium hinges on the assumption of global strategic complementarities—namely, an agent's incentive to take an action increases monotonically with the number of agents who take the same action. Rochet and Vives (2004), Dasgupta (2004), Goldstein and Pauzner (2004), and Goldstein and Pauzner (2005) show the theoretical possibility of a solvent bank failing because agents withdraw their investments for fearing that others will. All these papers show that the introduction of noisy signals to multiple-equilibria games may lead to a unique equilibrium. Since the signals are noisy, even a small asymmetry of information can lead to strategic uncertainty about the other agents' actions. This prevents the agents from coordinating their actions and reduces the set of possible equilibria. These models are panic based, that is, driven by bad expectations: depositors want to withdraw early because on the basis of their signal they fear that a sufficiently large number of other agents will withdraw. The beliefs of the investors are uniquely determined by the realization of the fundamentals in the sense that the fundamentals serve as a device coordinating agents' beliefs on a particular equilibrium. This approach allows reconciling two seemingly different views of banks runs: runs originated by negative real shocks, and runs originated by coordination failures.

Rochet and Vives (2004) apply this approach to the interbank market where, as a result of the signal, a fraction of bankers decide to withdraw from other banks. A banker withdraws if and only if the probability of failure of the bank, conditional on the signal and the behavior of the other bankers, is large enough. This leads to the following equilibrium if the signal about returns is poor bank failures caused by insolvency; but if the signal about returns is good failures caused by illiquidity, this may still occur if many other bankers withdraw. The uniqueness of equilibrium and the fact that it is based on bank fundamentals allow Rochet and Vives (2004) to develop policy recommendations. Even if liquidity and solvency regulation can solve the coordination problem, Rochet and Vives (2004) show that the cost in terms of foregone investment is too large. Thus prudential measures must be complemented with LOLR interventions. CB interventions can be in the form of OMOs that lower the need to 'fire sale' bank assets, or discount window lending. If the 'fire sale' premium is high because of temporary liquidity crises then OMO may be preferred. Discount windows interventions may instead be preferred if the 'fire sale' premium is thought to originate from an adverse selection problem. In either case, since the penalty rate exacerbates strategic complementarities among depositors, the LOLR lending should be at below market rate in contrast with Bagehot's doctrine of lending at a penalty rate.

A common feature of Freixas, Parigi, and Rochet (2000), Rochet and Vives (2004), and Freixas and Jorge (2008) is of interest in the analysis of the current crisis, and this is the existence of hoarding—that is, liquidity that is not intended to be used in the market. From that perspective we could argue that there are two types of liquidity: *effective liquidity* that is invested in the market as the market conditions guarantee a sufficient expected return and *potential liquidity* that cannot be invested in the market at the market conditions regarding information and/or uncertainty of other agent's strategies. This is relevant as, in this case, the Central Bank, acting as a LOLR will try to convert potential liquidity into effective liquidity and by so doing acts as a financial intermediary of last resort. This is possible through the issue of Government debt and the risky investment of its proceeds, by supporting commercial paper and banks newly issued debt, with future generation supporting the risk.

Efficiency, regulators objective function, and decentralization

Although the characterization of the optimal LOLR policy, as we have derived it in the previous section, brings a number of important insights, its implementation might be confronted with serious difficulties. First, the objective function of the regulator may be biased, not reflecting the correct values for the social costs and benefits of the continuation vs. liquidation decision. Second, the regulatory structure may be decentralized and therefore will combine decisions of two or more institutions with different objective functions.[6] This decentralization may take different forms, depending on whether it is within a country, where a CB and a deposit insurance company have to coordinate their policies, between monetary authorities and institutions in charge of prudential regulation (as the FSA in the UK), or between several countries as it is the case for multinational banks.

A unique regulator

In order to analyze the bias in the regulator's objective function, consider first, as a benchmark, the decision of the regulator in the absence of a liquidity shortage. A distinction parallel to the one that is drawn regarding the autonomy of Central

[6] Note that while the issue of the interplay between different regulatory instruments, as, for instance, deposit insurance and banking supervision, has been extensively examined, the interplay between different regulators has only recently been considered.

Banks may here be useful (Lybeck and Morris, 2004). A regulator is entrusted with *goal autonomy* if it has the power to determine its primary objective from several objectives; it is entrusted with *target autonomy* only if it has autonomy over one clearly defined primary objective, usually stipulated in the law. Absent moral hazard on behalf of the regulator, goal autonomy is preferred to target autonomy, because unrestricted welfare maximization clearly dominates the maximization of other objective functions. This could be the result of a tendency for the regulator to overestimate the cost of a liquidation or the benefits of the continuation, or, more likely, could be the consequence of an incorrect objective function for the regulator in terms of its career, compensation, and reputation, which, as stated in the classical Stigler (1971) and Peltzman (1976) 'capture theory' might lead to collusion between regulator and the regulated bank.

Consequently, the effective comparison is not between restricted and unrestricted unbiased regulatory policies, but, rather between two second-best policies: cost minimization, as a form of *target autonomy*, and welfare maximization, within a *goal autonomy* which is open to the regulator's discretionary interpretation and is, therefore, more sensitive to the biases in the regulator's objective function. In the US, the first approach has been selected: the regulatory choice has been to tie the regulator's hands by giving him a precise mandate of cost minimization.

Restricting the analysis to cost minimization implies that the regulator's objective function is biased. This bias arises from the very fact that prudential regulation is concerned with downside risk only and disregards the upside potential for profits that the continuation policy may involve. Because of this, the cost-minimizing LOLR will always be biased toward liquidation (Kahn and Santos, 2005). On the other hand, the regulator's own objective function may be biased toward overestimating the cost of liquidation and underestimating the cost of continuation and this will go in the opposite direction. In what follows we will survey the main results in the literature that makes the assumption of cost minimization. In the case of decentralization, this will imply that each regulator will consider only the costs that it has to bear, and not the total social cost.

A remark is here in order regarding the legal context. Although theoretically the regulator's mandate gives him the power to remove management and close down the bank, it may provide these options at very high costs (e.g., years of litigation) that the regulator may want to avoid. This issue has become prominent with the rescue of Northern Rock, where the bank had to be nationalized in spite of the interest shown by some group of investors, or, in a different vein, in the case of Fortis, where the selling was deemed illegal by the Belgian court. Two cases have to be considered: either these costs are limited and the regulator has the power to close down the bank independently of the liquidation shortage, or else it is only when the bank is forced to resort to the LOLR that the regulator is able to decide whether the bank should be granted access to the LOLR facilities (and therefore will be able to continue its activity) or whether it should be denied it, in which case it would be closed down.

The case where the closure decision is to be taken in connection with a LOLR operation can be viewed, from a theoretical perspective, as the case where the regulator observes a non-verifiable signal in a context where, for legal reasons, the decision to close down the bank cannot be based upon a non-verifiable signal.[7] A bank faced with a liquidity shortage that the interbank market is not ready to cover is forced to leave the continuation vs. closure decision in the hands of the CB. The lack of liquidity support from its peers, and the absence of liquid securities to pledge or sell, also provide additional information that will allow the LOLR to update its information.

Multiple domestic regulators

Decentralization between deposit insurance and LOLR has been initially analyzed by Repullo (2000) and then by Kahn and Santos (2005). With the trends toward the externalization of supervision and banking regulation outside the CB—for instance, in the hands of a FSA—this issue is highly topical. From a theoretical point of view, the coordination of the two institutions should not be a problem as the two regulators compensation package could be optimally determined. Yet, in practice this is a clear issue, as both institutions may have different views regarding the costs and benefits of a loan to the defaulting institution. Although the total cost of a bank failure, whether cost of liquidation or reputation cost to the regulators, could be shared, the positions regarding the effect of a loan are quite different and affect each institution's incentives to intervene to rescue or to liquidate a defaulting institution. For the CB, the cost of granting a loan is the amount of the loan, because, in the event of failure, part of the cost will be borne by the deposit insurance. On the other hand, from the point of view of the deposit insurance company, the issue is continuation or liquidation rather than the cost of extending credit. This is so because the deposit insurance company will take into account the cost of reimbursing all the insured deposits, not just the cost of the loan loss. As a consequence, both institutions may have opposite biases: the deposit insurance may be excessively prone to liquidation, because, as explained in the unique regulator case, the regulator considers only downside risks, while the CB, not bearing the full cost of paying back insured depositors, may be bent toward continuation.

The allocation of power between the two regulators as well as their incentives will be capital in determining the type of LOLR policy. This issue might have played a key role in the Northern Rock crisis and will therefore be at the forefront of the regulatory analysis in the forthcoming years. Two cases are to be considered depending on whether the agreement of the two institutions is necessary, or

[7] Casual observation seems to indicate that regulatory authorities are quite reluctant to enter into a legal battle with a commercial bank that might end in a Pyrrhic victory. The case of BCCI, where the Bank of England waited for almost a year before accumulating sufficient evidence of fraud prior to the coordinated closure of the institution across the world in July 1991, illustrates this point.

whether the allocation of the authority to close down a bank depends upon the extent of the liquidity shortage.

An interesting new issue has emerged with the 2007–9 crisis. As financial innovations have allowed banks to sell risks to other agents, it has turned out that the crisis has affected these agents and this, in turn, has threatened financial stability. This has occurred with AIG, that, as an insurance company, was out of reach of US banking supervision. Yet the AIG bankruptcy, because of the huge amount of CDSs it held in its portfolio, would have been systemic. It also happened with money market mutual funds after the losses generated by Lehman Brothers led to negative returns that, in turn, created a run. Again, the Fed intervention was necessary, although these institutions were beyond the scope of the Fed's regulatory mandate.

So, the paradox of the 2007–9 crisis is that risk sharing with non-bank agents has been welcome, as it was thought that this could not only provide additional diversification across different types of financial risks, like banking and insurance, but also that it would provide an additional cushion to support losses on banking instruments. Yet, when these losses become effective, they crash into the limited liability of non-bank institutions and the LOLR is forced to bail them out, as otherwise it would generate additional systemic risks. Thus, it has to be acknowledged that the commercial banking industry safety net once we face a systemic crisis becomes much more extended.

Functional specialization

When the mandates of the two regulatory agencies are clearly differentiated, the CB specializes in last resort lending while the deposit insurance company is in charge of the continuation vs. closure decision. In this case, the two regulatory institutions have to agree on continuation, as otherwise the bank is closed down. In other words, a bank may be closed down either by the deposit insurance company because of the signal on the bank future profitability, or else by the CB, because of the excessive cost of the LOLR operation. So, because each regulator has a veto power, the joint decision reduces the forbearance problem (Kahn and Santos, 2005). Note that, again, the deposit insurance decision to close down the bank is biased toward excessive liquidation because it does not internalize the upper tail of the bank's profit, while the CB decision is also biased because the cost of a loan depends on its size; so for small loans the CB is prone to forbearance while for large ones it has a bias toward excessive liquidation that is even bigger than the deposit insurance company's one. So, at high levels of illiquidity, the CB may refuse to lend despite the good signal the deposit insurance company receives on the bank's future profitability.[8]

[8] The issue of how the deposit insurance scheme is funded is also relevant for the analysis of the bailout decision: a deposit insurance that is funded *ex ante* may lead the regulatory authorities to understate the cost of a bailout, as they do not have to raise additional funds; on the other hand, if it is

Regulatory powers allocated depending on the size of the liquidity shortage

Repullo (2000) considers an incomplete contract approach where both the CB and the deposit insurance company are able to observe the same unverifiable signal on the bank future profitability, but where the right to act as the LOLR and therefore to take the decision as to whether the liquidity short bank should continue or should be closed down depends upon a verifiable variable, the amount of the required loan.

The main result of Repullo is that, if the CB is a junior creditor with respect to the deposit insurance company, and if it lends at a zero interest rate, then it is optimal to allocate control to the CB when withdrawals are small and to the Deposit Insurance Company when they are large. As before, the intuition is based on the biases of the two regulatory institutions. For small loans, the CB's decisions are closer to the optimal ones, while for larger amounts, the deposit insurance company's decisions are preferred. Contingent allocation on the basis of the realized liquidity needs leads, therefore, to an improvement with respect to the case where a unique regulator is in charge.

Extending the framework

Both Repullo (2000) and Kahn and Santos (2005) assume that the LOLR is a junior creditor. The reason for making this assumption is not that this is the most common structure, as frequently the CB is senior. The reason is simply that, if the CB has a senior claim and is certain always to recover the full amount of its loan, it has an incentive systematically to forbear. Thus, from a theoretical perspective, the issue of whether the loan should be collateralized or not, or, the closely related one of whether the CB should be senior, is trivial: it reduces the cost of LOLR operations and therefore increases the tendency to forbearance. From that perspective, the current crisis illustrates the behavior of CBs confronting a systemic crisis. Although it is clear that any form of support by the Central Bank will be senior, and, as such will increase the risk of banks' market funding, as market funds will be junior with respect to Central Bank loan, CBs have no choice but to intervene, and by so doing substitute market funding by Government funding.

The specific form the CBs loan takes depends upon the financial environment. In the absence of a systemic crisis, CBs are quite reluctant to lend to financial institutions except against good collateral, thus following Bagehot's principles. Both the US Federal Reserve Discount Window and the ECB marginal lending facility operate on the basis of collateralized lending. Still, there is a crucial difference between them: the ECB marginal lending facility collateral is based on securities traded in financial markets. So, it appears to be a substitute for a repo

funded *ex post*, this may lead both to additional costs, as the deposit insurance has to raise funds when the crisis is developing, and to additional contagion. We are grateful to Falko Fetch for pointing this out to us.

market with insufficient liquidity. Nevertheless, under section 10B of the Federal Reserve Act, Reserve Banks can accept any assets satisfactory to them as collateral for discount window advances, a facility that has been used more and more extensively during the crisis of 2007–9.[9] So, the discount window mechanism cannot be considered a substitute for the repo market and it allows extending a larger amount of loans, possibly depriving the deposit insurance company of valuable assets in case of the bank bankruptcy. The same is true of the European marginal lending facility, which is open to a wider range of collateral than the weekly main refinancing operations auctions, even if lending at a penalty.

If there is a unique regulator, the regulator internalizes the cost that this generates on the deposit insurance company. This will then have an impact on uninsured claim holders that will see their probability of recovering the whole amount of their claim increase, while, at the same time, they will be confronted with a larger loss given default in case of liquidation. On the other hand, if the regulator is not unique, its seniority rights, or its use of collateralized loans, will generate an externality on the deposit insurance fund. The LOLR operations will not be decided based on the real cost, but on the fraction of the cost that the CB incurs and this will lead to even a larger bias toward forbearance.

The functional specialization can be pushed one step forward if the Treasury is guaranteeing the loan of the CB to the illiquid (and may be insolvent) bank. In this case, the cost to the CB is only reputational. Such a scheme would parallel the one existing in the UK with the Bank of England lending with the guarantee of the Treasury. The mechanism has been used to guarantee all the deposits at Northern Rock. The Repullo–Kahn–Santos approach predicts that the CB would be prone to excessive leniency.

Information gathering and information sharing

Up to now we have assumed that both regulators were endowed with an informative signal regarding the future profitability of the banks projects. In fact, one of the two regulatory agencies may have the informative signal and the question is whether it has incentives to share it with the other one. Kahn and Santos (2006) consider this issue. They show that when only the CB has access to information about the bank future profitability it has no incentive to transmit this information. The implication is that when the costs for the deposit insurance are prohibitively high, the optimal institutional mechanism is to allocate the liquidation vs. closure power to the CB for small liquidity shocks, and to always support the bank's continuation for large liquidity shocks.

[9] The Federal Reserve currently accepts a wide variety of instruments as collateral under sect. 10B, including customer notes, mortgages on one- to four-family homes, commercial real estate loans, credit card receivables, collateralized mortgage obligations, asset-backed securities, and a host of other common debt obligations (Board of Governors of the Federal Reserve System, 2002).

Although this statement may seem a bit extreme, it might nevertheless prove useful in the understanding of the Northern Rock crisis in 2007. The Bank of England considers that the Northern Rock crisis is not systemic. Yet, the market for liquidity dries. The FSA is in charge of solvency and issues a favorable report, confirming the 'pure liquidity' assumption. The Bank of England cannot institute a mechanism similar to the ECB or to the discount window that allows for a much larger class of admissible collateral, and is therefore forced to resort to a special LOLR operation that guarantees all deposits. Without entering on the structure of deposit insurance in the UK or the fact that the actual terms and conditions of the credit facility are not public, the FSA had considered Northern Rock as following a safe banking strategy.[10] The Bank of England had to rely on this information when choosing to extend a credit line. The difficulty of the Bank of England to avoid been perceived as encouraging moral hazard by not resisting to the lobbies of the uninsured debt holders is clear in this case. Despite previous official statement to the contrary, the images of depositors lined up to withdraw from Northern Rock branches in the end forced the UK authorities to guarantee all depositors. Since, at the same time, the credit line is guaranteed by Her Majesty's Treasury, the incentives to find information contradicting the FSA are quite narrow. The Northern Rock case illustrates the notion that the commitment not to save lenders from their excesses is not realistic even in a sophisticated financial system once the CBs are confronted with the fragility of the banking system and the possibility of a systemic crisis.

Monetary policy and the LOLR

One of the major achievements of theoretical and empirical research in the last two decades has been to establish that the independence of Central Banks in setting monetary policy is one of the conditions for economic stability. Still, when it comes to the independence of monetary policy and LOLR policy, the issue is more involved. To begin with, LOLR policy is part of the safety net. So, a preliminary question would be to consider the pros and cons of having monetary policy and prudential regulation responsibilities delegated to two distinct institutions. The theoretical argument in favor of separation is the existence of possible conflicts of interest: as LOLR, the CB may feel compelled to bail out banks if this is necessary to prevent a systemic crisis. The conflict of interest is all the more serious in that monetary policy is countercyclical, while prudential policy is procyclical, as bank bankruptcies occur in slowdowns (Goodhart and Schoenmaker, 1993). The cross-country empirical analysis of Goodhart and Schoenmaker (1993; 1995) establishes that central banks that have supervisory responsibilities experience higher inflation rates. On the other hand, Peek, Rosengren, and Tootell's (1999) empirical analysis

[10] On this it was not alone: Moody's in April 2007 and Standard & Poor's in August 2006 had raised Northern Rock rating by one notch.

shows that information obtained from bank supervision helps the CB to conduct monetary policy more effectively. More recently, Ioannidou (2005), examining the behavior of the three primary US federal regulators—the Federal Deposit Insurance Company (FDIC), the Office of the Comptroller of the Currency (OCC), and the Federal Reserve Board—shows that indicators of monetary policy do affect actions of the Federal Reserve Board, while it does not affect those of the FDIC or the OCC. Now, when we consider LOLR operations, the impact on monetary policy could, theoretically, be sterilized.[11] So, when the bailout operation concerns an individual bank crisis, monetary policy should not be affected. Of course, when facing a generalized crisis, as in the case of 2007–8, the CB has to consider the impact of the banking crisis on expected growth and inflation patterns, and therefore is expected to intervene. In summary, the empirical evidence indicates that whether the responsibilities of monetary policy and prudential regulation are joint or separate does affect the way they are implemented. Still, this does not tell us which of the two models, is more efficient.

An extreme version of the idea that the assets market is not perfectly elastic arises when not all assets can be used to purchase other assets. Gorton and Huang (2004) show that when there are such 'liquidity-in-advance' constraints it is privately efficient for agents to hoard liquidity but it is not socially efficient given the opportunity cost of foregone investment opportunities. When the amount of the assets to be sold is so large that it would have been inefficient for private agents to have hoarded liquidity, the government can improve welfare by creating liquidity to bail out banks by taxing solvent projects. The drawback is that if the government tax capacity is too small, the government cannot bail out all banks, and forbearance arises. This is indeed the case for small countries (witness the crisis in Iceland in October 2008) and poses a serious risk for midsize countries with large banks like Switzerland, where the assets of the banking system are several times larger than the country's GDP. The link with monetary policy and the conflict of interest it implies is clear. Banking crises will materialize in a downturn, under a tight monetary policy. This puts pressure on the prices of assets, thus setting the stage for a debt deflation. If, simultaneously, the LOLR has to bail out banks in distress, there is a clear case for coordination of policies and weighting of the cost of a higher inflation versus the cost of banking crises.

Multiple international regulators

The previous analysis of multiple regulators can be extended to international regulatory bodies. If regulators internalize correctly the costs of bankruptcy, the

[11] Although, as noted by Goodhart and Huang (1999a), the residual uncertainty on the precise amount of the bailout operation may have a macroeconomic impact.

Repullo–Kahn–Santos approach could lead to interesting insights. Regulators in different countries may have different signals, and, so, may have different views on the continuation vs. liquidation decision. Still, there are additional complexities, because of the possibility to free-ride on the subsidies provided by the other country. This issue constitutes one of the major challenges of banking regulation and is vital for the future of European Financial integration, where regulation of an institution could be in the hands of one institution while monetary policy is in the hands of another, and this asymmetry affects also the private and social costs of a bankruptcy. The classical textbook market imperfections of externalities and the provision of public goods are present here. When the major banks in some Eastern European countries take the form of a branch and are regulated by other countries' regulatory bodies (the Bank of Italy or the Oesterreichische Nationalbank), it is a source of concern whether the maximization of the regulator's home country objectives takes into account all important externalities on the host country. The public good provision problem arises because a pan-European bank rescue appears as a public good to be financed through coordination of a number of Treasuries across Europe. This issue is considered in Freixas (2003) and in Goodhart and Schoenmaker (2006). Its implications regarding LOLR policy are that the bailout of an institution may not be possible, simply because of the free-riding problem. The equilibrium outcome will be more biased toward liquidation than in the case of a unique multi-country regulator, as illustrated by the Fortis and the Icelandic banks' case.

The implementation through several regulatory bodies of LOLR decisions implies an additional number of restrictions. It means, therefore, that additional restrictions are imposed on the LOLR policy. This implies that type-one and type-two errors will occur, with the rescue of banks that should have been closed down and the liquidation of banks that should have been bailed out.

Conclusions

To conclude, it is worthwhile to compare the classic view of the LOLR with the complexities of the above analysis, and to try to summarize it by drawing the major lines of the recent advances in the contemporaneous approach of LOLR, as compared with the 'wisdom of our ancestors'. What is left today of the simple clear-cut guidelines suggested by Thornton and Bagehot who recommend lending to solvent illiquid institutions against good collateral and at a penalty rate?

First, lending to the market through OMO is the standard way for a CB to prevent an aggregate liquidity shock. This is the contemporaneous version of

'lending against good collateral', characteristic of developed financial markets. Yet, recent models of interbank lending teach us that market imperfections may lead to other inefficiencies that require the LOLR support to extend beyond the pure CB responsibility of aggregate liquidity management and lend to individual banks, either unsecured or against collateral of lower quality, or guaranteeing their future liquidity (Freixas, Parigi, and Rochet, 2000).

The second classical recommendation was to lend at a penalty. This point is now clearly controversial. In the presence of *ex ante* moral hazard, as in Freixas, Parigi, and Rochet (2004), a penalty provides managers with the right incentives to be diligent in their lending. Still, in Rochet and Vives (2004) the recommendation is the opposite—to lend at a rate inferior to the market rate. When, in addition, we consider decentralization between several regulatory agencies, a penalty on interest rates decreases the expected cost of the LOLR loan and imposes a better discipline in banks' liquidity management. This will, therefore, make the LOLR more prone to forbearance, which, as mentioned, could either increase or decrease the efficiency of LOLR. In the case of an unbiased regulator, this will be efficient, because, in case of success, the LOLR will obtain a share of the bank's profits. So, on the penalty issue there is no clear consensus yet and, hopefully, future work will help regulators to implement the efficient policy depending on the economic and financial environment.

Regarding the use of good collateral, Freixas, Parigi, and Rochet (2004) argue that this is a key feature in the intervention of the CB. In their approach, the CB intervenes through fully collateralized loans at a rate above the T-Bills rate, so that it encompasses a penalty, but below the interbank market rate.

The recent financial markets' turmoil offers several lessons regarding the role of the LOLR in a *systemic* crisis. First, we have witnessed how repeated and coordinated aggregate liquidity injections are not sufficient to solve the crisis: the illiquidity of financial institutions around the world is, in fact, directly linked not only to their solvency but also to asset prices. Second, it is important to notice that CBs around the world have been much more flexible in providing support to the banking industry than what was initially expected, or, in other words, that CBs cannot credibly commit to a bailout policy. Indeed, the arguments in support of the bailout of banks only if their closure could have a systemic impact ('too big to fail'), that were intended for an individual bank facing financial distress, were soon discarded in favor of a more realistic approach. The case of Northern Rock, certainly not a systemic bank, illustrates this point. Its liquidation in such a fragile banking environment would have triggered a domino effect with contagion from one institution to another. From that perspective, the lesson is that when facing a systemic crisis the LOLR has to take into account also the 'too-many-to-fail' perspective, and consider how it will treat all banks that are in a similar position. A third point is that in a systemic crisis the perimeter of the safety net is extended to non-bank institutions. This may be the result of the waves of financial innovation.

Yet, because AIG had been issuing CDSs its bankruptcy would have affected the fragility of the banking industry by leading to losses and a lower capital.

In the end, the above discussion highlights the important notion that the LOLR of the twenty-first century lies at the intersection of monetary policy, supervision and regulation of the banking industry, and the organization of the interbank market—a long way from the Bagehot doctrine.

References

Acharya, V. and Yorulmazer, T. (2007). 'Cash-in-the-Market Pricing and Optimal Resolution of Bank Failures', *Review of Financial Studies*, 21: 2705–42.

Adrian, T. and Shin, H. S. (2008). 'Liquidity and Financial Contagion', *Banque de France, Financial Stability Review, Special Issue on Liquidity*, 11: 1–7.

Afonso, G. M. and Shin, H. S. (2008). 'Systemic Risk and Liquidity in Payment Systems', Federal Reserve Bank of New York Staff Report 352, October.

Allen, F. and Gale, D. (2000). 'Financial Contagion', *Journal of Political Economy*, 108: 1–33.

Bagehot, W. (1873). *Lombard Street: A Description of the Money Market*. London: H.S. King

Blavarg, U. and Nimander, P. (2002). 'Inter-bank Exposures and Systemic Risk', *Sveriges Riksbank Economic Review* 2: 19–45.

Board of Governors of the Federal Reserve System (2002). 'Alternative Instruments for Open Market and Discount Window Operations', *Federal Reserve System Study Group on Alternative Instruments for System Operations*. Washington, DC, December.

Brunnemeir, M. (2008). Thoughts on a New Financial Architecture. http://www.princeton.edu/~markus/research/papers/new_financial_architecture.pdf.

—— (2009). 'Deciphering the Liquidity and Credit Crunch 2007–08', forthcoming *Journal of Economic Perspectives*, 23: 77–100.

—— and Pedersen, L. (2008). 'Market Liquidity and Funding Liquidity', *Review of Financial Studies*. <http://rfs.oxfordjournals.org./cgi/content/abstract/hhn098v2>.

Bryant, J. (1980). 'A Model of Reserves, Bank Runs and Deposit Insurance', *Journal of Banking & Finance* 4: 335–44.

Caballero, R. J. and Krishnamurthy, A. (2007). 'Collective Risk Management in a Flight to Quality Episode', National Bureau of Economic Research Working Paper 12896, *Journal of Finance*, 63: 2195–230.

—— —— (2008). 'Musical Chairs: A Comment on the Credit Crisis', *Banque de France, Financial Stability Review, Special Issue on Liquidity*, 11: 9–11.

Calomiris, C. W. and Kahn, C. M. (1991). 'The Role of Demandable Debt in Structuring Optimal Banking Arrangements', *American Economic Review*, 81: 497–513.

—— —— and Kroszner, R. S. (1996). 'The Efficiency of Self-Regulated Payments Systems: Learning from the Suffolk System', *Journal of Money Credit and Banking*, 28: 766–97.

Carlsson, H. and Van Damme, E. (1993). 'Global Games and Equilibrium Selection', *Econometrica* 61: 989–1018.

Castiglionesi, F. (2007). 'Financial Contagion and the Role of the Central Bank', *Journal of Banking & Finance*, 31: 81–101.

Cifuentes, R., Shin, H. S., and Ferrucci, G. (2005). 'Liquidity Risk and Contagion', *Journal of the European Economic Association*, 3: 556–66.

Coleman, S. P. (2002). 'The Evolution of the Federal Reserve's Intraday Credit Policies', *Federal reserve Bulletin*, 88: 67–84.

Dasgupta, A. (2004). 'Financial Contagion through Capital Connections: A Model of the origin and Spread of Bank Panics, *Journal of the European Economic Association*, 2: 1049–84.

Degryse, H. and Nguyen, G. (2006). 'Interbank Exposures: An Empirical Examination of Contagion Risk in the Belgian Banking System', Tilburg University Working Paper 2006–016.

Diamond, D. (1984). 'Financial Intermediation and Delegated Monitoring', *Review of Economics Studies*, 51: 393–414.

—— and Dybvig, P. (1983). 'Bank Runs, Deposit Insurance, and Liquidity', *Journal of Political Economy*, 91: 401–19.

—— and Rajan, R. (2005). 'Liquidity Shortages and Banking Crises', *The Journal of Finance*, 60: 615–47.

—— and Rajan, R. (2001). 'Liquidity risk, Liquidity Creation and Financial Fragility: A Theory of Banking', *Journal of Political Economy*, 109: 287–327.

Fisher, I. (1933). 'Debt Deflation Theory of the Great Depression', *Econometrica*, 4: 337–57.

Freixas, X. (2003). 'Crisis Management in Europe', in J. Kremers, D. Schoenmaker, and P. Wierts (eds.), *Financial Supervision in Europe*, Cheltenham: Edward Elgar.

—— and Holthausen, C. (2005). 'Interbank Market Integration under Asymmetric Information', *Review of Financial Studies*, 18: 459–90.

—— and Jorge, J. (2008). 'The role of Interbank Markets in Monetary Policy: A Model with Rationing', *Journal of Money, Credit and Banking*, 40: 1151–76.

—— Parigi, B. M., and Rochet, J.-C. (2000). 'Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank', *Journal of Money, Credit and Banking*, 32: 611–38.

———— (2004). 'The Lender of Last Resort: A 21st-Century Approach', *Journal of the European Economic Association*, 2: 1085–115.

Friedman, M. and Schwartz, A. J. (1963). *A Monetary History of the United States, 1867–1960*, Princeton, NJ: Princeton University.

Furfine C. H. (2003). 'Interbank Exposures: Quantifying the Risk of Contagion', *Journal of Money Credit and Banking*, 35: 111–28.

Goldstein, I. and Pauzner, A. (2004). 'Contagion of Self-Fulfilling Crises Due to Diversification of Investment Portfolios', *Journal of Economic Theory*, 119: 151–83.

———— (2005). 'Demand Deposit Contracts and the Probability of Bank Runs', *Journal of Finance*, 60: 1293–327.

Goodhart, C. A. E. (1987). 'Why Do Banks Need a Central Bank?', *Oxford Economic Papers*, 39: 75–89.

—— (1999). 'Myths about the Lender of Last Resort', *International Finance*, 2: 339–60.

—— (2008). 'Liquidity Risk Management', *Banque de France, Financial Stability Review, Special Issue on Liquidity*, 11: 39–44.

—— and Huang, H. (1999). 'A Model of Lender of Last Resort', IMF Working Paper 99/39.

—— and Schoenmaker, D. (1993). 'Institutional Separation Between Supervisory and Monetary Agencies', LSE Financial Market Group Special Paper 52.

———— (1995). 'Should the Functions of Monetary Policy and Bank Supervision be Separated?', *Oxford Economic Papers*, 39: 75–89.

——— ——— (2006). 'Burden Sharing in a Banking Crisis in Europe', *Swedish Riksbank Economic Review*, 2: 34–57.

Goodfriend, M. and King, R. (1988). 'Financial Deregulation Monetary Policy and Central Banking', in W. Haraf and R. M. Kushmeider (eds.), *Restructuring Banking and Financial Services in America*, American Enterprise Institute Studies 481, Lanham, Md.: University Press of America.

Gorton, G. (1985). 'Clearinghouses and the Origin of Central Banking in the United States', *Journal of Economic History*, 2: 277–83.

——— (2008). 'The Panic of 2007', Yale School of Management, mimeo.

——— and Huang, L. (2004). 'Liquidity, Efficiency and Bank Bailouts', *American Economic Review*, 94: 455–83.

——— and Mullineaux, D. (1987). 'The Joint Production of Confidence: Endogenous Regulation and Nineteenth-Century Commercial Bank Clearinghouses', *Journal of Money Credit and Banking*, 4: 457–68.

Humphrey, D. B., (1986). 'Payments Finality and Risk of Settlement Failure', in A. Saunders and L. White (eds.), *Technology and the Regulation of Financial Markets: Securities, Futures, and Banking*, Lexington, MA: Lexington Books.

Humphrey, T. (1989). 'The Lender of Last Resort: The Concept in History', *Federal Reserve Bank of Richmond Economic Review*, 75: 8–16.

Ioannidou, V. P. (2005). 'Does Monetary Policy Affect the Central Bank's Role in Bank Supervision?, *Journal of Financial Intermediation*, 14: 58–85.

Iyer, R. and Peydro, J. L. (2006). 'Interbank Contagion: Evidence from Real Transactions', University of Amsterdam, mimeo.

Kahn, C. M. and Santos, J. A. C. (2005). 'Allocating Bank Regulatory Powers: Lender of Last Resort, Deposit Insurance and Supervision', *European Economic Review*, 49: 2107–36.

——— ——— (2006). 'Who Should Act as Lender of Last Resort? An Incomplete Contracts Model: A Comment', *Journal of Money Credit and Banking*, 38, 1111–18.

Kyotaki, N. and Moore, J. (1997). 'Credit Cycles', *Journal of Political Economy*, 105: 211–48.

Lacker, J. M. (2004). 'Payment System Disruptions and the Federal Reserve Following September 11, 2001', *Journal of Monetary Economics*, 51: 935–65.

Lehar, A. (2005). 'Measuring Systemic Risk: A Risk Management Approach, *Journal of Banking & Finance*, 29: 2577–2603.

Leitner, Y. (2005). 'Financial Network: Contagion, Commitment, and Private Sector Bailouts', *Journal of Finance*, 60: 2925–53.

Lucas, R. E. (2008). 'Bernanke is the Best Stimulus Right Now', *The Wall Street Journal*, December 23.

Lybeck, T. and Morris, J. A. (2004). 'Central Bank Governance: A survey of Boards and Management', IMF Working Paper 04/226.

Merton, R. C. (1977). 'An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees: An Application of Modern Option Pricing Theory, *Journal of Banking & Finance*, 1: 3–11.

McAndrews, J. and Potter, S. (2002). 'Liquidity Effects of the Events of September 11, 2001', *Federal Reserve Banks of New York Economic Policy Review*, November, 59–79.

Mistrulli, P. (2005). 'Interbank Lending Patterns and Financial Contagion', Bank of Italy, mimeo.

Morris, S. and Shin H. S. (1998). 'Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks', *American Economic Review*, 88: 587–97.

Peek, J., Rosengren, E., and Tootell, G. (1999). 'Is Bank Supervision Central to Central Banking?', *Quarterly Journal of Economics*, 114: 629–53.

Peltzman, S. (1976). 'Toward a More General Theory or Regulation', *Journal of Law & Economics*, 19: 211–40.

Repullo, R. (2000). 'Who Should Act as a Lender of Last Resort? An Incomplete Contract Model', *Journal of Money, Credit and Banking*, 32: 580–605.

Rochet, J.-C. and Vives, X. (2004). 'Coordination Failures and the Lender of Last Resort: Was Bagehot Right After All?', *Journal of the European Economics Association*, 2: 1116–47.

Santos, J. A. C. (2006). 'Insuring Banks against Liquidity Shocks: The Role of Deposit Insurance and Lending of Last Resort', *Journal of Economic Surveys*, 20: 459–82.

Schnabel, I. and Shin, H. S. (2004). 'Liquidity and Contagion: The Crisis of 1763', *Journal of the European Economic Association*, 2: 929–68.

Shin, H. S. (2006). 'Risk and Liquidity in a System Context', Bank of International Settlements Working Paper 212.

Sprague, O. M. W. (1910). *History of Crises Under the National Banking System*, Philadelphia: Government Printing Office.

Stigler, G. J. (1971). 'The Theory of Economic Regulation', *The Bell Journal of Economics and Management Science*, 2: 3–21.

Timberlake, R. H. (1978). *The Origins of Central Banking in the United States*, Cambridge, MA: Harvard University Press.

—— (1993). *Monetary Policy in the United States*, Chicago: University of Chicago Press.

Thornton, H. (1802). *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain*, London: Hatchard.

Upper, C. and Worms, A. (2004). 'Estimating Bilateral Exposures in the German Interbank Market: Is There a Danger of Contagion?', *European Economic Review*, 48: 827–49.

Wells, S. (2004). 'Financial Interlinkages in the United Kingdom's Interbank Market and the Risk of Contagion', Bank of England Working Paper 230.

White, E. N. (1983). *The Regulation and Reform of the American Banking System, 1900–1929*, Princeton, NJ: Princeton University Press.

CHAPTER 12

REGULATION AND SUPERVISION

AN ETHICAL PERSPECTIVE

EDWARD J. KANE[1]

Introduction

Although the yoke between 'love and marriage' has become markedly more flexible than it was portrayed in Sammy Cahn and Jimmy Van Heusen's popular song of 1955, 'regulation and supervision' remain tightly conjoined. Regulation focuses on rules; super*vision* looks to their enforcement. Etymologically, regulation is a grammatical extension of *regula*: the Latin word for 'rules'. In every country, governments make rules that define formally what a bank is, what different kinds of things banking organizations may and may not do, how and where bankers may and may not do permissible things, and what reciprocal rights and duties bankers and regulators owe to one another. Moreover, for compelling historical, cultural, economic, and political reasons, these definitions vary across countries—often greatly (Barth, Caprio, and Levine, 2006).

[1] This chapter extends and refocuses analysis first presented in Kane (1998). For helpful comments, the author wishes to thank Gerard Caprio and John Wilson.

Differences in rules and enforcement support what has become a worldwide market for regulatory services. Just as bank managers might explore the market for any other support service that they plan to outsource, they sort through alternative regulatory schemes to ascertain the particular jurisdiction that offers them the best mix of costs and benefits for the various pieces of their product lines. In the absence of switching costs, each bank would design a series of substitute asset, liability, and hedging instruments and negotiate with (i.e., 'lobby') would-be suppliers so that each deal they write could be booked in the most favorable jurisdiction.

To sort out cross-country and cross-product differences in the quality and offering prices of different regulatory entities, this chapter develops the concept of an incentive-conflicted *regulatory culture*. In each country, the broad outline of this culture is determined by inherited ethical norms for individual, industry, and government behavior. However, the institutional details that constitute a particular *regulatory scheme* and go on to shape its particular policy instruments and their operative costs and benefits are developed cooperatively in response to the push and pull of lobbying pressures. The conflict between a top regulator's duties and outside political forces incorporates into every real-world system of bank regulation contradictory controls and subsidies that, when left unchallenged through time, tempt client banks to expose themselves to a growing chance of economic insolvency.

Many countries have experienced a banking crisis in recent years (Caprio and Klingebiel, 1996; and Honohan and Klingebiel, 2003). Of these crises, a high proportion was triggered by losses generated by government efforts to allocate bank credit to well-connected firms in politically influential sectors. The upside of these crises is that, in exposing inefficient, contradictory, and anti-egalitarian elements of regulatory policies in particular countries, they can generate pressure for effective reform.

Ethics of supervision

Economic theory presumes that, subject to external constraints, individuals choose a series of behaviors that maximize through time a personal objective function. Rules come into existence in situations where people fear that gaps in other individuals' ethical standards might allow them to behave in ways that would jeopardize the goals of a rule-making community to which they belong.

To constrain the choices that targeted parties make, rules must be backed up by supervision. Supervision entails surveillance and enforcement. Regulated parties ('regulatees') must be supervised when—and to the extent that—their objective

functions tempt them to make themselves better off by disobeying either the spirit or the letter of particular rules.

A bank's incentive to circumvent or violate a given rule increases with the weight of the burdens that full compliance threatens to impose on its efforts to create value and manage risk. Dutiful enforcement revises bank incentives by rewarding compliance, punishing evasion, and searching out and closing loopholes that regulatees might use to skirt the rules.

Rulemakers spell out in capital letters the behaviors that they wish either to avoid or to promote and usually in language almost anyone can understand. However, most rules contain a set of loopholes that is communicated either in very small print or in coded language that only the lobbyists that sponsored them can immediately see or understand. To quantify the economic burden of any rule, one must study not only the costs and benefits of compliance, but the opportunity costs of circumvention as well.

Loopholes sustain gaps in supervisory enforcement that generate a second set of rules. These secondary rules are designed to discourage appeals to higher authority and are at least partially conjectural. For example, although the formal speed limit on a given highway might be posted at (say) 55 miles per hour, drivers confidently expect the limit that police actually enforce to be higher than the posted one and to adapt predictably to exceptional circumstances (such as personal emergencies) as these unfold.

Common law and the 'commonsense school' of ethical theory maintain that, across any contract in which one party delegates authority to one or more others, agents, and principals owe one another duties of loyalty, competence, and care. On this hypothesis, supervisors owe four key duties to the community that employs them:

1. *A duty of vision*: they should continually adapt their surveillance systems to counter regulatee efforts to disguise their rulebreaking;
2. *A duty of prompt corrective action*: they should stand ready to discipline rulebreakers whenever a violation is observed;
3. *A duty of efficient operation*: they should produce their services at minimum cost;
4. *A duty of conscientious representation*: they should be prepared to put the interest of the community they serve ahead of their own.

In principle, supervisors committed to the fourth duty would bond themselves to disclose enough information about their decision-making to allow the community to make them accountable for neglecting or abusing these responsibilities. In practice, institutional arrangements do not hold supervisors strongly accountable for the distributional effects of how they resolve incentive conflicts. To the contrary and in country after country, politicians require bank lending to favor designated sectors of the economy. To obtain a quid pro quo, bank stakeholders expect these

loans to be supervised with a lighter hand, especially in times of banking turmoil (Kane, 1989).

Traditionally, supervisory duties have been exercised locally and—in a narrow and formal sense—schemes for regulating and supervising banks are still shaped and administered on a nation-by-nation basis. Changes in rules and duties respond to the interplay of economic events with changing governmental goals and with the waxing and waning of industry pressure to relax burdensome rules or to control disruptive behaviors. Kane (1977; 1981; and 1988) describes a dialectical process in which regulation-induced innovation engenders regulatory adjustments and regulatory adjustments (termed 're-regulation') engender new sources of regulatee avoidance.

Today, national schemes and resulting regulatee burdens are increasingly being influenced by competition from foreign regulatory systems. In world markets, movements of financial capital and changing asset values overlay onto the domestic policy scene a series of unfamiliar political, economic, and reputational pressures that individual country regulatory decision-makers must take into account. Arguably, these pressures have persuaded authorities in financial center countries to acquiesce in loophole-ridden agreements for coordinating cross-country supervision (Basel I and II).

This chapter introduces the concept of a regulation-induced financial crisis and uses it to explain how offshore regulatory competition can either reinforce or attenuate inefficient or anti-egalitarian elements of incentive-conflicted banking regulation in individual countries. Regulatory competition does this mainly by inducing increases and decreases in the banking business a country's banks can capture. With technological change intensifying the influence of offshore regulators, mis-steps promise to come to a boil sooner, but may still have severe and long-lasting effects on ordinary taxpayers. This chapter exemplifies the process by analyzing how regulatory competition simultaneously encouraged incentive-conflicted supervisors to outsource much of their due discipline to credit rating firms and encouraged banks to securitize their loans in ways that pushed credit risks into corners of the universe where supervisors and credit ratings firms could not see them.

The market for regulatory services and its imperfections

Although a large literature treats banking regulation as if it were simply a *tax* on bank income, bankers understand that banking regulation is better conceived as a

back-office financial *service* that, for participants in banking markets, generates benefits as well as costs. Its benefits lie in three realms: improving customer confidence, improving customer convenience, and supporting or resisting bank efforts to accumulate and exercise market power. Because banking regulation requires resources to produce, authorities can both produce it more or less efficiently and finance it more or less fairly. Whether or not the costs of producing regulation are minimized, political activity determines its level and allocates its production costs across society. Any firm or individual implicitly pays an endogenously determined price for regulatory services. This price corresponds to the difference between the benefits that firm or household receives from bank regulation and the costs that banking regulation imposes on it. We conceive of this variable as a given taxpayer's 'net regulatory benefit (or burden) from banking regulation', or NRB.

Parties that feel a stake in banking regulation routinely join together into political coalitions and lobby collectively for improvements in their NRBs. In principle, each sector's lobbyists compete self-interestedly with lobbyists from other sectors to generate regulatory benefits for their members and to shift the costs of financing their production toward parties located in other sectors.

In a world in which banking markets are globalized, services that provide regulatory benefits are available from foreign as well as domestic suppliers. Hence, the struggle by citizens and firms of any one country to maximize net benefits spills across its borders into what has become a worldwide market for financial regulation.

The market for regulatory services comprises a body of persons that carry on extensive transactions in the specific activity of promulgating, enforcing, and accepting regulatory restrictions. Regulation is supplied competitively and accepted voluntarily to the extent that entry and exit opportunities exist for banks willing to incur the transaction costs of switching all or part of their regulatory business to another supplier. Hence, although a regulator's clientele is fixed in the very short run, the jurisdictions in which a regulatee operates are voluntary over longer periods. Geographic overlaps in the global market for financial regulatory services have expanded as entry and exit costs for foreign financial institutions have declined around the world. Ongoing downward trends in costs of entering and exiting offshore financial markets render the margin of regulatory competition—even in developing countries—increasingly global.

Rules and enforcement systems are continually tested and reshaped by changes in the net regulatory benefits that other jurisdictions offer. Nevertheless, jurisdictional competition for most financial products is inherently imperfect. An incumbent regulator may be said to have market power in any line in which it can lower the NRB it offers clients without completely surrendering its clientele to another regulator. Alternatively, we might say that the *leaders* of a regulatory agency have market power whenever the various labor, capital, and political markets from

which they draw economic resources cannot hold them (and the elected politicians that appoint and sustain them) accountable for policy decisions that simultaneously lower net regulatory burdens for their clientele of lenders and borrowers and increase them for other important economic sectors.

The vigor of regulatory competition is enhanced by technological change and diminished by information asymmetries, leadership turnover, and various sources of principal–agent conflict that are inherent in governmental decision-making. The essence of a government's social contract is that taxpayers—as principals—award financial resources and coercive powers to governmental agents. Taxpayers hope that government officials will exercise the assigned powers to promote the 'common good'. However, the common good cannot be observationally defined. Moreover, especially in the short run, an agency's conception of the common good may be distorted by sectoral pressures.

Regulators routinely adopt reporting systems that make it difficult for citizens to gather information either about subsidiary goals that policymakers might be pursuing or about sectoral, bureaucratic, or personal benefits that regulatory activity might generate. Even when evidence of discriminatory or inefficient performance surfaces, it is difficult for outside observers to sort out its root causes or to correct the incentive defects responsible for it.

The value of regulatory competition lies in supplying indirect economic checks on the even-handedness and efficiency of net regulatory burdens. On the demand side, competition encourages parties that feel overburdened by their government's system of regulation to reconfigure their business to slide it into the jurisdiction of a more-advantageous supplier of regulatory services. It does not matter whether the new supplier is a domestic agency or a foreign one. What matters is that the regulatees gain some relief, the new regulator gains budgetary resources, and the old regulator loses them. The lower the transition costs of moving to a less-burdensome regulatory supplier, the more complete the demand-side check becomes.

On the supply side, entry and exit costs confer competitive advantages on incumbent regulators. In competing with would-be private regulatory enterprises, government entities are advantaged by the financial strength imparted to them by the presumption that they can assign catastrophic losses to taxpayers and by their ready access to the coercive power of the state. To a non-traditional supplier, the costs of actively gearing up to oversee even a narrow category of banking deals can be substantial. The existence of these costs means that the number of potential new entrants that can economically supply regulatory services to banks in a given country is relatively limited in the short run.

Successful entry requires more than a capacity for exercising disciplinary power. To displace a seasoned regulator, would-be entrants need specific skills, a source of moral authority, and substantial financial and reputational capital. Entrants must be able to promise credibly that they can fairly and efficiently produce regulatory

services and are committed and able to sustain this promise for a long while. They must be able to manipulate a system of rewards and punishments that is strong enough to change the behavior of potential regulatees. The entry of newly chartered *private* regulators into regulatory arenas is discouraged by the costs of accumulating sufficient public standing and moral authority to be trusted with this kind of coercive authority.

In brief, the inherited market structure for regulatory services is distorted by market power that the law freely gives to government enterprises and by reputational advantages enjoyed by incumbent private regulators. On the one hand, representative democracy confers renewable monopoly power on elected politicians and the regulatory leaders they appoint. Because policymaking authority may be canceled by voters or limited *ex post* by the courts, this authority becomes all the stronger, the more confidently incumbent politicians may count on holding power and the more that top bureaucrats may count on holding onto their offices and avoiding vigorous prosecution or public censure for questionable acts.

Even in the private sector, market power is conferred in lasting fashion on a successful regulatory enterprise. It is interesting that such traditionally hard-to-dislodge incumbent regulators as a country's major stock and commodities exchanges are being subjected today to pressures from cross-country partners that hope to take over their franchise. It is unfortunate that, for key regulatory bureaus, central banks, and ministries of finance, takeover discipline cannot be so direct.

The role of incentive conflicts and regulatory subsidies in banking fragility

Banking environments and patterns of banking regulation vary greatly from country to country. Financial institution supervision combines a capacity to observe fluctuations in balance sheet values ('vision') with a capacity to influence managerial actions ('control') and an *incentive system* that governs the pursuit and exercise of these capacities. Even when portfolios and attendant risks are concentrated within a single country, it is difficult to establish a combination of adequate oversight of institutional balance sheets, adequate authority to intervene in timely fashion, and bureaucratic incentives to detect and resolve insolvent institutions in ways that adequately protect taxpayer interests. As a result, individual countries solve this contracting problem in different ways. Although many

commonalities of interest exist, systems for setting and enforcing financial rules are infested with incentive conflict. Even within a country, conflicts exist between and among:

- regulators and the firms they regulate;
- particular regulators and other societal watchdogs;
- regulators and the politicians to whom they must report;
- taxpayers and the politicians and regulators they put in office.

How a country approaches and resolves these conflicts is in part hard-wired into its political and institutional structure. For example, while many European Union countries supervise banks separately from other financial institutions, some do not. A few European countries (Austria, Denmark, Germany, Sweden, the Netherlands, and the UK, in particular) have established agencies that supervise bank and non-bank financial institutions in an integrated way; others have to some degree integrated the oversight of at least their bank and securities sectors (Schüler, 2003).

Every country relies on its ethical norms, government regulators, and other professional watchdogs to bridge gaps in the bonding, deterrent rights (deterrency), and transparency inherent in its private contracting environment. Over time, the interaction of private and government watchdogs generates a *regulatory culture*. A culture may be defined as customs, ideas, and attitudes that members of a group share and transmit from generation to generation by systems of subtle and unsubtle rewards and punishments. A regulatory culture constrains the ways in which an uncooperative or even unscrupulous individual bank can be monitored and disciplined. It comprises a matrix of attitudes and beliefs about how regulators should act. These slowly changing attitudes and beliefs often express a distrust of government power that traces back to abuses observed in a possibly distant past when the country was occupied, colonized, or run by a one-party government. The culture's taboos and traditions define standards for the fair use of government power. Behind these standards are higher-order social norms that underlie a nation's political and legal environments.

The character of a country's regulatory culture is spanned by six specific components:

- legal authority and reporting obligations;
- formulation and promulgation of specific rules;
- technology of monitoring for violations and compliance;
- allowable penalties for material violations;
- duties of consultation (to guarantee fairness, regulated parties enjoy a right to procedural due process that specifies burdens of proof that regulators must meet before they can penalize violators);
- regulatee rights to judicial review (to bond the fairness guarantee, penalized parties have access to inside and outside appeals procedures).

In large part, the details of each component are shaped by:

- recognition and response lags generated by the interaction of weakness in the transparency of the nation's accounting system with bureaucratic incentives and statutory and bureaucratic checks and balances;
- regulatory competition brought about by the entry of foreign or differently regulated institutions;
- regulatory personnel's exposure to influence activity from a discipline-resistant firm's political clout, consultation rights, and appeal privileges;
- social norms that protect fraudsters and bumblers against prompt regulatory discipline.

Lobbying activity seeks to reshape the particular norms that officials stress and to constrain the trade-offs they make. Within limits set by a country's regulatory culture, how particular policy strategies officials adopt actually work is determined by regulatees' ability to delay or stymie decisive intervention and to find and exploit circumventive loopholes. Some of these loopholes involve the ability to relocate loss exposures that are more closely supervised either by the home country (or by a particular host) to venues that monitor or discipline risk taking less effectively.

The regulatory cultures of almost every country in the world today embrace in one form or another three strategic elements:

1. *politically directed subsidies to selected bank borrowers* (the policy framework either explicitly requires—or implicitly rewards—banks for making credit available to selected classes of borrowers at a subsidized interest rate);
2. *subsidies to bank risk taking* (the policy framework commits government officials to providing on subsidized terms explicit or implicit conjectural guarantees of repayment to depositors and other bank creditors);
3. *defective monitoring and control of the subsidies* (the contracting and accounting frameworks used by banks and government officials fail to make anyone directly accountable for reporting or controlling the size of either subsidy in a conscientious or timely fashion).

Taken together, the first two elements in the strategy tempt banks to extract wealth surreptitiously from taxpayers and constrain loan officers to pass some of the benefits to politically favored borrowers (such as builders and would-be homeowners—especially low-income households—in the US). Favored borrowers tend to be blocs of voters regularly courted by candidates for political office and financial supporters or cronies of influential government officials.

The third piece of the framework minimizes regulators' exposure to blame when things go wrong. It makes it impossible for outsiders to hold supervisors culpable for violating their ethical duties. It prevents outsiders from readily monitoring the true costs and risks generated by the first two strategies and interferes with efforts to subject the intersectoral flow of net regulatory benefits to informed debate. This gap exists

because accounting systems do not report the value of regulatory benefits as a separate item for banks that receive them. In modern accounting systems, the capitalized value of regulatory subsidies is treated instead as an *intangible* source of value that, if booked at all, is not differentiated from other elements of a bank's so-called 'franchise value'. Of course, some of the subsidy is offset by *tangible* losses that politically influenced loans eventually force onto bank balance sheets and income statements. In principle, a tangible reserve for expected losses ought to be set up as part of the process of making a poorly underwritten or deliberately underpriced loan. Not reserving for losses imbedded in a loan's preferential terms may be conceived as planting a time bomb in the asset and net-worth values shown on conventional bank balance sheets. Over time, the cumulative damage from politically favored loans becomes harder and harder to hide. Between one crisis and the next, the amount of government-favored loans grows larger and larger in bank portfolios. Eventually, a shortfall of contractual cash flows makes it harder to gain financing for pools of mispriced and poorly structured loans. This is how poorly documented mortgage-backed securitizations came a cropper in the US and Europe during the summer of 2007. Although officials resist the idea, creating an enforceable obligation for regulators to estimate in transparent and reproducible ways the ebb and flow of the dual subsidies would empower external watchdog organizations in the private sector to force authorities to explain whether and how these subsidies benefit taxpayers.

Sooner or later, savvy large-denomination creditors come to appreciate the unreported hole that overvalued loans imbed in the opportunity–cost value of their bank's enterprise-contributed net worth (NW_E). By NW_E, we mean the value that an informed buyer would pay for the bank if safety net guarantees did not exist. If a bank's NW_E declines through zero, it becomes a 'zombie' institution. A 'zombie' is an insolvent institution that stays active only because the black magic of government guarantees leaves its creditors with no reason to force it into a corporate grave. A zombie's ability to renew its deposit funding and other debt depends entirely on the continuing credibility of the explicit and implicit government guarantees that safety net managers attach to its obligations.

Accounting loopholes allow a zombie institution to show positive accounting net worth long after its NW_E has turned negative. For example, although we now know that in June 2007 the British mortgage lender Northern Rock was well on its way to becoming a zombie, management was able to post an accounting net worth equal to roughly 2 percent of its assets.

Systemwide fragility F increases with the number of zombies or near-zombies (Z) and with the aggregate size of the losses thought to be imbedded in their economic balance sheets:

$$F = F\left[Z, \sum_{j=1}^{Z} NW_E\ (j)\right] \tag{1}$$

Funding problems begin *not* when a bank becomes a zombie, but when suppliers of large-denomination funds begin to doubt whether officials can or will continue to support its existence. Funding problems for a region's or country's banking system are intensified when doubts arise about arrangements for making taxpayers absorb the cost of guaranteeing the area's potential zombie institutions. The triggering condition is that the upper bound on the uncertain value of implicit and explicit government guarantees G rises so high that taxpayer resistance threatens to make it hard for authorities to raise the funds needed to pay the bill promptly or in full. Massive withdrawals or requests for collateral by sophisticated creditors are sometimes described as 'silent runs', because servicing the demands that a troubled bank receives from large creditors generates far less publicity than the queue of panicked small depositors that impatiently mills about in a conventional run.

However, silent runs greatly weaken bank balance sheets. The deposit outflows that troubled banks experience must be financed by selling liquid assets and issuing costly debt. A troubled bank's first line of defense against a silent run is to arrange loans from government institutions or from relatively well-informed banks with which it has correspondent relationships. Private rescuers usually insist on receiving appropriately high interest rates and demand collateralization and an upside potential for their claims. In deciding to help a correspondent bank to weather a silent run, foreign banks are apt first to lobby the IMF, the host government, and even their own government for assurances that they will not be stuck with the bill for whatever losses the rescue effort might incur.

Until officials increase the transparency and credibility of their credit support, silent runs on weak institutions tend to escalate. Troubled banks' sales of good assets and increasing funding costs reduce future income and make the fragility of their condition apparent to more and more outside observers. When a troubled bank collateralizes its good assets at or above their market value, its unbooked losses on poorly performing loans become a larger proportion of the assets that remain unpledged. The more funding a troubled bank obtains at high credit spreads, the more severely its future accounting and economic profits are squeezed and the more likely it is to engage in go-for-broke lending and funding activities that severely pressure the profit margins of healthy competitors.

A silent run puts pressure on regulators because it progressively undermines the willingness of taxpayers and stronger banks to tolerate the regulatory status quo. As a silent run unfolds, reduced profit margins spread zombieness and disturbing information is revealed about the size of taxpayers' potential involvement. At the same time, net regulatory benefits for weak and strong banks diverge more and more widely. Weak banks receive safety net subsidies from central bank loans and government guarantees that stronger banks and general taxpayers eventually have to pay for.

The longer a silent run proceeds, the more deeply supervisory efforts to retard the exit or to delay the formal recapitalization of inefficient and insolvent deposit

institutions push the net regulatory benefits of other economic sectors into negative territory. The economic and political forces exerted when a large bank suffers open and silent runs are nicely illustrated by the British government's response to the Northern Rock debacle. In September 2007, an open depositor run on this bank was stopped by the government's promise to provide emergency funding to the £114 billion institution and to 'guarantee all existing deposit arrangements'. However, a silent run persisted. By year end, emergency loans from the Bank of England reached about £25 billion and Treasury guarantees had been extended to cover most of the bank's non-deposit obligations as well. Well-publicized efforts to persuade stockholders and outside acquirers to inject private capital into the bank showed little progress. Finally, in February 2008, the bank was 'temporarily' nationalized.

Three exculpatory norms of modern crisis management

A severely overleveraged banking system may be portrayed as an accident waiting to happen. A regulation-induced crisis occurs when misfortune impacts a banking system whose managers have made their institutions vulnerable to this amount and type of bad luck. Figure 12.1 breaks the evolution of a regulation-induced banking crisis into five stages. The 2007–8 breakdown of arrangements for financing structured securitizations in the US and Europe, and banking crises that rolled through Latin America, Japan, Korea, the Philippines, Malaysia, Indonesia, Thailand, and Russia during 1997–8 passed through the first three-and-one-half stages of this model of crisis generation and response.

In 2007–8, German, British, and American authorities showed again that politicians are reluctant to move beyond the stopgap partial recapitalization stage (stage 4A). As long as the hopelessness of an institution's situation can be covered up, outsiders cannot easily distinguish a wave of financial institution insolvencies from a transitory shortage of aggregate liquidity. In either circumstance, a group of economically significant firms find it exceedingly difficult to roll over their liabilities on profitable terms. It is an accepted first-response practice for central bankers and other regulators to provide liquidity to distressed institutions as a way to buy time for their supervisory staff to investigate the extent to which irreparable insolvencies might underlie the distress. This time-buying strategy is supported by three exculpatory norms whose ethical force intensifies in times of political, market, or institutional turmoil: a mercy norm; a nationalistic norm; and a non-escalation norm.

The mercy norm holds that it is bad policy and unacceptably cruel behavior for regulators to abandon the employees, creditors, and stockholders of institutions

1. Rent-seeking generates aggressive loss exposures at highly leveraged institutions
   - Pursuit of safety net subsidies tied to government-promoted forms of lending
   - Pursuit of subsidies tied to other kinds of leveraged risk taking
2. Adverse events and industry problems upset financial markets
   - Banks and regulators keep losses from registering on bank books by accounting trickery and cover-up
   - Large-denomination creditors test the strength of the safety net
   - Fragility of system rises as good assets are collateralized and endgame incentives induce go-for-broke gambling
   - Threat of shortages in safety net funding rises over time
3. Supplementation of traditional safety net support mechanisms
   - Loans from central bank discount window cannot carry the load
   - Inventive accounting loopholes and forms of public credit support expand
4. Recapitalization of troubled banks and safety net institutions
   A. Stopgap partial recapitalizations: half-measures move the financial sector back into stage two of the cycle
   B. Transformation of bank losses into explicit taxpayer obligations or explicit nationalization of zombie banks
5. Final clean-up of the mess
   - Reprivatization of zombie institutions
   - Blame heaped on designated scapegoats
   - Credible safety net reforms are adopted

Fig. 12.1. Five stages of a regulation-induced banking crisis

they oversee before they can convincingly establish whether the distress is too fundamental to be remedied by subsidized loans. This norm gives regulators the discretion (if not the duty) to alleviate the initial pains of any client institution that experiences a silent run.

The nationalistic norm presupposes that regulators should help domestic institutions and marketmakers to cope with foreign competition. In practice, this norm is reinforced by community resistance to foreign control of national credit decisions and by lobbying pressure from politically favored sectors who suspect that foreign banks will not serve their interests very well.

The non-escalation norm allows authorities to lend on subsidized terms to distressed institutions as long as they can popularize the view that doing anything else would invite a national or global financial disaster. In invoking this norm, officials must spread fear. They must argue that, without a large injection of subsidized funds, markets will set prices for troubled assets that are unreasonably low and prices for emergency credit to institutions that hold these assets that are unreasonably high and that these price movements would sweep strong and healthy institutions into the turmoil.

It is dangerous for government officials both to make these exaggerated claims and to deny the increasingly transparent flow of subsidies that partial recapitalization entails. For high-ranking regulators to keep churning out safety net subsidies, two further conditions must hold. First, they must be able to control the flow of information, so as to keep taxpayers and the press from convincingly assessing either the magnitude of the implicit capital transfer or the antiegalitarian character of the subsidization scheme. Second, the self-interest of top regulators must be continually nourished by praise and other forms of tribute from the bankers, borrowers, and investors whose losses are being shifted to less influential parties.

Authorities are reluctant to move to full recapitalization until overwhelming losses reveal themselves in the form of strongly resurging crisis pressures. The longer the game goes on, the greater the risk that the reputations of incoming policymakers and the particular politicians that appoint them will be saddled unfairly with the sins of their predecessors. Although it is unwise to draw inferences from a sample of two, the US savings-and-loan mess and the most recent Argentine crisis cast some light on how costs are allocated during the final stages in the life-cycle of a regulation-induced crisis.

Formally, continuations and breakdowns in the burden-shifting process may be analyzed as two states of an evolutionary process. Though small on any given day, the probability (p) of a breakdown during an incentive-conflicted regulator's term in office increases with the fragility of the system for making good on implicit and explicit safety net guarantees. It is convenient to represent the value of these guarantees as G and the cumulative size of the taxpayer's hidden responsibility for supporting the liabilities of troubled institutions as T. T and G increase with system fragility (F). In turn, whenever F grows, p also rises. During the early stages of an incipient crisis, increments in the probability of breakdown depend on the informativeness (A) of the accounting principles that banks and safety net officials use to report losses and loss exposures:

$$p = p[G,T,F;A].$$

During these early stages, banks and their regulators are tempted to seek and provide 'accounting relief'. However, once market participants begin to recognize partial recapitalizations and cover-ups as half-measures, weaknesses in A compound the problem and improvements in A become a critical part of the crisis resolution process.

Rolling and incompletely resolved crises sound at least three alarms. First, the frequency and geographic extent of banking crises convincingly demonstrate that, around the world, numerous banks have found it reasonable to book potentially ruinous risks. Looking at the period 1977–95, Caprio and Klingebiel (1996) cite fifty-eight countries in which the net worth of the banking system was almost or entirely eliminated. Second, in country after country, domestic (and sometimes foreign) taxpayers have been billed to bail out banks, depositors, and deposit insurance

funds. Honohan and Klingebiel (2003) confirm that, in recent crises, taxpayers' bill for making good on implicit and explicit guarantees typically ran between 1 and 10 percent of GDP. The size of these bailouts establishes that, at least in crisis countries, banks managed to put large bets on the table and were able to shift a substantial amount of the downside of these bets to taxpayers. In many cases, authorities were eventually blamed for the size of the bills taxpayers were asked to pay. Officials were seen to have shirked their duties to expose and stop loss-causing patterns of credit allocation and to have compounded the damage from credit losses by not addressing individual bank insolvencies until their situation had deteriorated disastrously.

In times of financial turmoil, weaknesses in ethical controls on the job performance of government regulators responsible for protecting the safety and soundness of financial institutions encourage regulatory forbearance. The high cost of modern crises indicates how far the risk-taking preferences of officials responsible for managing taxpayer risk exposures diverge from those of large-denomination creditors in private financial markets. Although institutional mechanisms for financing safety net loans and guarantees differ across countries, poor information flows and incentive conflict in government policymaking complicate the treatment of banking crises everywhere.

Special problems of accountability and incentive conflict arise in managing cross-country risk exposures. Financial regulators subject foreign banks and the foreign operations of domestic banks to patterns of regulation that differ in two important ways from those that apply to strictly domestic banking operations. First, most developed countries are willing to allow their domestic banks to book a wider range of risks in foreign subsidiaries than they are prepared to tolerate in home country offices. This is because relationships with internationally active customers are a geographically footloose part of the banking business and because government officials don't expect to confront responsibility for foreign banking losses in domestic political arenas. This creates incentives for offshore banks to 'overlend' into foreign markets. Second, though greatly weakened by technological change and outside political pressure, obstacles to the entry of foreign financial firms in most banking markets still exist.

Globalization and securitization of bank funding opportunities

Contemporary theories of industrial organization seek to explain how a product's market structure evolves through time to permit *efficient firms* and *efficient contracting instruments* to reshape or displace relatively less efficient alternatives.

The force of these theories is particularly easy to grasp when we focus on hypothetical markets that meet a set of ideal conditions that Baumol, Panzar, and Willig (1986) call 'perfect contestability'.

A market is perfectly contestable when entry and exit costs are each zero *and* incumbent firms exit quickly whenever they find themselves faced with negative profits. In perfectly contestable markets, low-cost firms readily displace high-cost firms and incumbent competitors are prevented from setting monopoly prices by the threat of hit-and-run entry by other equally efficient firms. Financial markets are never perfectly contestable. New entrants must adapt and expand their information systems before they can safely expand their customer base. Incumbents cannot easily abandon the pipeline of loan commitments they have promised to customers and the regulatory foundations on which inherently non-transparent financial markets must be built are burdened with inescapable entry and exit costs.

During the last thirty years, particularly in wholesale banking markets, technological change has steadily lowered entry costs for foreign and non-traditional competitors. Most of these firms undertook banking activities in innovative ways, making creative use of substitute products, substitute organizational forms, and substitute offshore locations. In some countries, the viability of a new entrant's business plan was temporarily enhanced by long-standing restrictions on how banks could compete domestically.

Chief among the innovative methods of doing business was structured securitization. With help from investment banks, credit rating firms, mortgage insurers, and hedge funds, banks sliced and securitized titles to the cash flows from their loans in ways that assigned the slicing (or 'tranching'), reslicing, and servicing of flows of interest and principal to separately capitalized conduit vehicles. By placing important tranches of their loans through and with foreign and non-bank firms, banks permanently layered the institutional character and broadened the geographic span of bank funding arrangements.

Innovative funding technologies benefited borrowers by integrating bank loan pricing within and across countries. However, outsourcing the funding side of a bank's balance sheet weakened its staff members' due diligence by severing the link between the income a lender could make from originating securitizable loans and the quality of its system for underwriting the loans it originated. Investors in a securitized pool of loans did not rely on either the lender's or their own due diligence. Instead, they expected private profit-making credit rating organizations to assess the risks in the positions they were offered and they expected investment banks and mortgage insurers to make sure that the returns offered would respond appropriately to differences in loan quality. Unfortunately, the confidence with which these expectations were held undermined agents' incentives to meet them. Compensation for rating and pricing individual securities was collected as soon as the securities were floated, with little exposure to *ex post* blowback for personnel that might later be

shown to have made a serious rating or pricing mistake. With supervisors closing their eyes to the erosion of this chain of agents' contractual incentives to execute faithfully their duties of loyalty, competence, and care, investors presumed that they were purchasing titles to well-rated and well-priced securities.

Securitization also brought firms that were supervised in different regulatory cultures and jurisdictions into sharper competition with one another. This mutual invasion of traditional markets by institutions headquartered in different regulatory cultures put pressure on particular regulatory enterprises (especially at enterprises whose leaders' remaining terms in office promised to be short) to relax vigilance as a way of defending their bureaucratic turf. In retrospect, it is clear that banking supervisors did this by regularizing and legitimating cutting-edge ways to hide or transfer risk without fully exploring the threats that these complex new contracting structures imposed on individual country safety nets.

Whenever a regulator acquiesced in innovative entry by a foreign or non-traditional firm, it had to relax restraints that might make it hard for its traditional clients to compete with the new entrants. Institutions pressed politicians to make this happen sooner rather than later.

Authorities' positive response to this competitive pressure has been labeled *financial deregulation*, but our ethical perspective makes it clear that the response is better described as *desupervision*. In most countries, regulatory competition and defects in accountability led banking supervisors to assess the risks of innovative instruments of risk transfer with less watchfulness than these instruments deserved. With respect to structured securitizations, banking supervisors and mortgage-insurance firms outsourced their duty of vision to accountants and credit rating firms without adequately bonding the obligations they were asking them to perform. They did this despite these firms' obvious conflicts in goals and evidence of outsized delays in downgrading distressed securities in past downturns (Portes, 2008).

The contestability of banking markets is greatly reduced by the political clout that domestic banks enjoy and by the ability of supervisory entities to bill government safety nets for the losses their heedlessness might engender. In crises, safety net subsidies disadvantage less subsidized competitors and unreasonably sustain the operations of decapitalized banks. The contestable-markets portrayal of market structure evolution helps us to understand that in most countries deregulation focused on unblocking entry without addressing supervisory incentives to resist the exit of important domestic banks. Bank and supervisory exit resistance attenuates the benefits to society that entry relaxation would otherwise produce. Banking crises teach foreign and non-traditional competitors the need to estimate the extent of supervisor-supported exit resistance. By standing ready to absorb the losses of unprofitable clients, a regulator (especially a central bank) can prevent low-cost entrants from earning the profits needed to justify hit-and-run entry.

Dialectics of a regulation-induced banking crisis

For any policymaker, a crisis may be described as a time of upheaval that generates strong pressure for decisive changes in policy strategy. Figure 12.2 portrays a regulation-induced banking crisis as an evolutionary process that is driven in Hegelian fashion by dialectical collisions of irreconcilable market and regulatory adjustments.

For any regulated institution, change—not rest—represents the path of profit-making equilibrium. The Hegelian model of regulation assumes that the conflict between regulated parties and their regulators can never be completely eliminated. The contradictory forces at work in each round of adjustments are labeled the 'thesis' and the 'antithesis'. Every sequence of adjustment and response produces a temporary 'synthesis' that serves in turn as the 'thesis' for a new round of action and response.

In the US, policies designed to promote home ownership encouraged borrowers and lenders alike to operate with a 'perilously high degree of leverage' (Shadow Financial Regulatory Committee, 2008). For borrowers, the value of the subsidies that they could derive both from tax deductions for mortgage interest and from federal programs supporting mortgage credit increased with the amount they borrowed. For lenders, federal programs supported the securitization of home mortgages by offering cheap guarantees and by making it possible for banks to

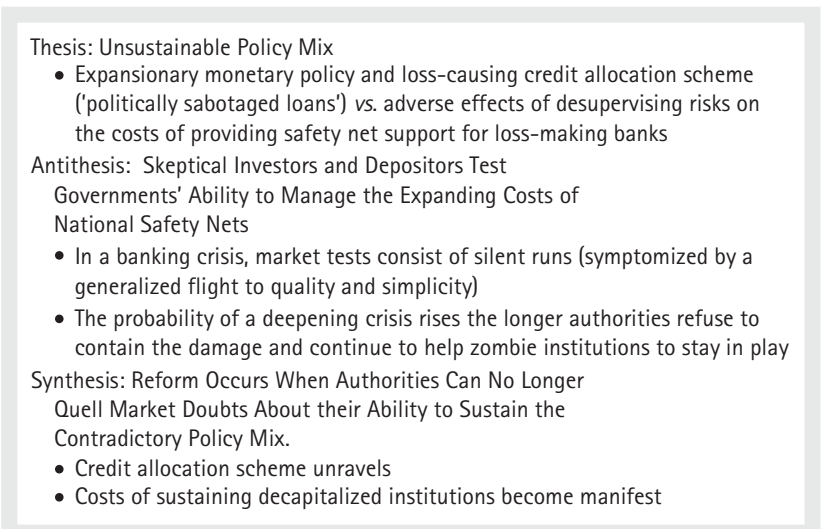

Fig. 12.2. Dialectics of a regulation-induced crisis

avoid capital requirements on mortgages that they chose to securitize. Bank supervisors did not require banks either to estimate or to hold capital against the implicit obligations that structured securitization vehicles passed through to a sponsor's net worth. The high degree of leverage on borrower positions meant that, if and when housing prices declined by more than a few percent, marginal borrowers would be unable to service their obligations. Once a sharp increase in delinquencies and foreclosures by subprime borrowers occurred, savvy investors revalued and cut back their positions in securitized mortgage pools. When this revaluation wiped out the equity of mortgage securitization conduits, reputational concerns persuaded bank sponsors to move a good portion of conduit losses back onto their balance sheets. Besides being billed for conduit losses, banks that had been heavily involved in originating mortgages for sale to conduits were stuck with losses on pipelines of ongoing mortgage commitments that they could no longer profitably securitize. Inevitably, silent runs on these banks tested the ability of safety net managers to manage a spreading crisis.

The appropriate policy response to crisis pressures depends on the nature of the policy contradictions that occasioned the crisis. A perennial issue is to assess the potential insolvency of troubled banks and to determine how rapidly bank net worth is being undermined by falling prices on crisis-creating loans. Asset price meltdowns are most likely to occur when incentives for overlending by domestic and offshore institutions confront a host-country policy regime that offers incentives for overborrowing at domestic households and firms. In such cases, pressure on asset prices is apt to generate a crisis-intensifying run from claims issued by the insolvent borrowers and lenders.

It is superficial to conceive of the silent runs that triggered the US securitization crisis as manifestations of an underprovision of aggregate 'liquidity'. In fact, the US central bank had for many years accommodated overspending in the favored sector and also financed a long run of current account deficits. A central bank can prolong a payments deficit by letting its currency decline and by drawing down the country's foreign exchange reserves and foreign lines of credit. In any consumption-driven currency devaluation, the need to rebuild the central bank's currency reserves may or may not be urgent. If it is, authorities can shrink the current account deficit in two complementary ways: (1) by allowing the exchange rate to decline even further and (2) by tightening their mix of fiscal and monetary policies.

But when a financial center country is experiencing a banking crisis, this prescription is unattractive. These policies would impose a sizeable opportunity loss on foreign and domestic holders of the country's financial assets. The currency adjustment half of this strategy would put inflationary pressure on domestic prices. To pile on the tight-money half of the prescription would induce a decline in aggregate economic demand, whose effects would reduce the real value of a country's financial assets in general and the net worth of its banking system in

particular. This would further undermine asset values by raising prospective rates of default and delinquency on troubled assets. In crisis circumstances, it is politically impossible for authorities to ignore the effects that these adjustments would have on safety net loss exposures.

In a financial center country, authorities face a three-way policy dilemma about how to control a silent run:

- *Choice one.* Try to finance the runs with minimal adjustment in the loss-causing parts of the policy mix. We may describe this strategy as disinformational 'hardball'. Authorities may temporarily nationalize one or more insolvent institutions and deny that any other zombies exist. They may or may not soften the potential decline in their exchange rate by drawing down reserves or borrowing from private and official foreign sources.
- *Choice two.* Rebalance the policy mix to make it more sustainable, but only with respect to a narrowly defined window of time (e.g., until after the next election). Authorities may resolve or strengthen some of the weakest institutions and may slow monetary growth. We have described this as a strategy of 'partial recapitalization'.
- *Choice three.* (unlikely to be chosen unless prior efforts to use one or both of the other strategies have failed dramatically). Face up to and eliminate the most obvious contradictions in the policy mix. The new policy regime should aim for a full clean-up of insolvent institutions and to establish a more incentive-compatible supervisory system going forward.

Leaving bank and corporate insolvencies unresolved fosters further malinvestment and enhances the likelihood that a deeper crisis will emerge down the line. Still, it is dangerous to acknowledge and resolve corporate and banking insolvencies in the midst of a national recession. In crisis circumstances, politicians are strongly tempted to reflate demand and to strengthen the credibility of safety net guarantees, without doing much to resolve the incentive distortions that widespread insolvency creates.

The role of regulatory competition in banking crises

Contradictory policies misallocate capital in the household, financial, corporate, and government-planning sectors. The result is that asset values and bank net worth eventually have to be written down. Had asset values either been supported by a sustainable expansion in productive capacity or been written down promptly

as unfavorable information surfaced, silent runs would not have become large enough to test the safety nets of financial center countries.

The seeds of the 2007–8 subprime crisis were sown over many decades. They did not flower into a crisis until doubts began to surface about authorities' willingness and ability to measure and absorb the losses and loss exposures confronting a suddenly decapitalized banking system. Measurement is important. As in the 1980s savings-and-loan mess, crisis costs were intensified by openly delaying loss recognition at loss-making institutions.

What the press describes as a 'banking crisis' may be more accurately described as the surfacing of tensions caused by the continuing efforts of loss-making banks to force the rest of society to accept responsibility for their unpaid bills for making bad loans. In US mortgage markets, long-standing systems for subsidizing poorly underwritten loans to builders and overleveraged households imposed unbooked losses both on banks and on supporting national safety nets.

Around the world, financial institutions and markets are supported by regulatory systems that show numerous country-specific features (Wilson, 1986; Dermine, 2003; Barth, Caprio, and Levine, 2006). Differences in patterns of financial regulation address differences that exist in the various economic, political, and bureaucratic deficiencies and inefficiencies that each country's regulatory system is overtly or covertly expected to correct (Garcia and Nieto, 2006; and Herring and Schuermann, 2006).

However, the survival of differences in regulatory patterns is limited by the tendency of funding and loan-making opportunities to flow to markets and institutions that offer their customers the best deal. The extent to which net regulatory burdens on financial markets and institutions differ across countries is narrowed by the regulatory arbitrage that interjurisdictional deal flows entail. When and as technological change in information processing and telecommunications lowers the cost of transacting with foreign entities, adverse flows of capital and financial deal-making help to persuade a nation's authorities to lower the net burdens that their regulatory framework imposes on the savers and investors that book deals in its financial markets.

In recent years, rolling banking and currency crises have become frequent for three reasons. First, advances in information and communications technology have simultaneously globalized banking markets and markets for government guarantees. Second, the globalization of markets for banking and guarantee services has made it less costly for domestic corporations and wealthy investors to mount silent runs on a country's zombie institutions. Third, lenders, securitizers, credit rating organizations, and supervisory authorities are not compensated in ways that make them accountable for the slow-developing but inevitable losses that their policies engender.

In 1997–8, crises in Korea, Indonesia, Malaysia, the Philippines, and Thailand were hastened by the technologically driven absorption of these countries into an

international market for loanable funds that allowed large depositors to protect themselves against the burdens of inefficient or discriminatory patterns of national regulation. Globalization put the costs and benefits of banking regulation in these countries into closer competition with the regulatory systems of offshore financial centers.

Offshore banking competition shortened in two ways the crisis-gestation period featured in traditional crisis models (such as Krugman, 1979). First, even limited entry by outside banks expanded the stock of well-priced domestically available substitutes for deposits that local citizens had previously held in host-country banks. This lowered the cost to Asian depositors of participating in a silent run on domestic banks. Second, the relative safety of foreign bank deposit substitutes demonstrated the greater reliability of the performance guarantees written for each offshore entrant by the regulatory systems of its homeland.

Each crisis constitutes an exit cost that society incurs to shrink the domain of a high-cost or discriminatory regulator. Regulation-induced crises are triggered by efforts to avoid the inefficiencies and inequities that political maneuvering interjects into particular markets for regulatory services. Squeezing the equilibrium rents that short-sighted or corruptible officials can extract and distribute to their supporters disciplines incumbent regulators, but only at the margin. To improve public service contracting in the longer run will require authorities to expose themselves to blowback for the delayed effects of policy mistakes by accepting a performance-based scheme of deferred compensation.

Exploitive regulation drives sophisticated depositors, unsubsidized borrowers, and other bank stakeholders to book at least some of their business elsewhere: either abroad or in informal or differently regulated domestic markets. Such regulatory arbitrage limits the extent to which politicians can promote a distribution of regulatory burdens that arbitrarily narrows opportunities for important sectors of a national economy to accumulate and manage their wealth.

The normative goal of financial reform should be to induce non-discriminatory and efficient patterns of regulation and supervision. Regulators should be made accountable not just for producing a stable financial economy, but for providing this stability fairly and at minimum long-run cost to society. In practice, this means establishing contractual incentives that would lead authorities to follow market-mimicking standards of supervisory performance. In the absence of explicit or implicit government guarantees, markets would insist that any bank that experiences a spate of opportunity–cost losses do one or more of three things: shrink, raise more equity capital, or pay higher interest rates for its debt. The public policy problem is to design employment contracts that would make it in supervisors' self-interest to invoke 'market-mimicking' disciplines when and as a country's important institutions weaken.

Although officials understand that strengthening bank supervision is part of crisis resolution, they seem reluctant to identify the behavioral norms and

incentive structures that made a crisis country's supervision weak in the first place or to recommend public service contracting and reporting reforms that would be strong enough to make tougher supervision serve an incentive-conflicted regulator's self-interest.

For any regime, the size of tolerable deviations from a fair and efficient distribution of net regulatory burdens increases with the opportunity costs its citizens face in engaging in capital flight. In turn, the benefits and costs of capital flight evolve with information technology, the volatility of the real economy, and the fluidity of the political environment. The information revolution that is underway in finance today makes it short-sighted and inequitable to adopt credit-allocation schemes that inexorably eat away at the capital of a country's banks and that require taxpayers to subsidize weak banks and uneconomic patterns of real investment. Credit rating organizations and the Basel Committee on Banking Supervision would be well-advised to abandon sampling procedures that set aside the costs of adverse tail events and models that presume that asset risks are relatively stationary over time. They should focus also on finding ways both to bond the scrupulousness with which staff members perform their supervisory duties and to enlist forward-looking betting and derivatives markets to help them track the changing odds of defaults in individual countries and industries (Kane, 2003).

References

Barth, J. R., G. Caprio, and R. Levine (2006). *Rethinking Bank Regulation: Till Angels Govern*. New York: Cambridge University Press.

Baumol, W., J. C. Panzar, and R. D. Willig (1986). 'On the Theory of Contestable Markets', in C. F. Matthewson and J. E. Stiglitz (eds.), *New Developments in the Theory of Industrial Structure*. Cambridge, MA: MIT Press, 339–65.

Caprio, G. and D. Klingebiel (1996). 'Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking', *Annual Conference on Development Economics*, Washington: The World Bank, 79–104.

Dermine, J. (2003). 'European Banking: Past, Present, and Future', in V. Gaspar, P. Hartmann, and O. Sleijpen (eds.), *The Transformation of the European Financial System*. Frankfurt: European Central Bank, 31–95.

Garcia, G. and M. Nieto (2006). 'Banking Crisis Management in the European Union: Multiple Regulators and Resolution Authorities', *Journal of Banking Regulation*, 6: 215–19.

Herring, R. and T. Schuermann (2006). 'Capital Regulation for Position Risk in Banks, Securities Firms and Insurance Companies', in Hal Scott (ed.), *Capital Adequacy: Law, Regulation, and Implementation*. Oxford: Oxford University Press.

Honohan, P. and D. Klingebiel (2003). 'The Fiscal Cost Implications of an Accommodating Approach to Banking Crises', *Journal of Banking & Finance*, 27: 1539–60.

KANE, E. J. (1977). 'Good Intentions and Unintended Evil: The Case Against Selective Credit Allocation', *Journal of Money, Credit, and Banking*, 9: 55–69.

—— (1981). 'Accelerating Inflation, Technological Innovation and the Decreasing Effectiveness of Banking Regulation', *Journal of Finance*, 36: 355–67.

—— (1988). 'Interaction of Financial and Regulatory Innovation', *American Economic Review*, 78: 328–34.

—— (1989). 'Changing Incentives Facing Financial Services Regulators', *Journal of Financial Services Research*, 2: 263–72.

—— (1998). 'Capital Movements, Asset Values, and Banking Policy in Globalized Markets', in Lee Seongtae (ed.), *The Implications of Globalization of World Financial Markets*. Seoul: Bank of Korea, 278–98.

—— (2003). 'What Kind of Multinational Arrangements Might Best Enhance World Welfare?', *Pacific Basin Finance Journal*, 11: 413–28.

KRUGMAN, P. (1979). 'A Model of Balance-of-Payments Crises', *Journal of Money, Credit, and Banking*, 11: 311–25.

PORTES, R. (2008). 'Ratings Agency Reform', posted on *www.voxeu.org*, January 12.

SCHÜLER, M. (2003). *How do Banking Supervisors Deal with Europe-wide Systemic Risk?*, Discussion Paper 03-03, Centre for European Economic Research.

Shadow Financial Regulatory Committee (2008). 'Statement on Facilitating Mortgage Renegotiations: The Policy Issues', Statement 255, Chicago, February 11.

WILSON, J. A. G. (1986). *Banking Policy and Structure: A Comparative Analysis*. New York University Press, New York.

CHAPTER 13

DEPOSIT INSURANCE

ROBERT A. EISENBEIS
GEORGE G. KAUFMAN

Introduction

Deposit insurance has become a central component of every country's financial safety net, in both developed and developing countries. Demirgüç-Kunt, Karacaovali, and Laeven (2005) in a recent World Bank survey indicate that at least eighty-three of 181 countries queried in 2004 have explicit deposit guarantee schemes, and, of these, fourteen were established between 2000 and 2004. The remainder de facto provide implicit guarantees. Thus, some form of deposit insurance protection is now effectively a political reality in every country that has a banking system.

Although widely adopted to reduce both the probability and the costs of bank failure, recent evidence suggests poorly implemented deposit insurance schemes have contributed at least as much to increasing both the probability and costs of bank failure as to reducing them (Kane, Hovakimian, and Laeven, 2003; Kane, 2000; Kane and Yu, 1996; and Kane, 1993).

As the events of the 2007–9 worldwide financial crisis have shown, even a well-designed deposit insurance will not always prevent a crisis from occurring. Nevertheless, because all crises come to an end, it is important to consider how deposit insurance and the financial safety net should be structured in non-crisis periods.

The recent crisis also suggests that some large non-bank financial institutions can be as important as commercial banks as potential propagators and sources of systemic risk and may require access to parts of the safety net even in normal times.

This chapter briefly discusses the objectives of deposit insurance and how it should function in non-crisis situations. It also analyzes evidence on how well various deposit insurance systems have functioned, and notes critical conceptual weaknesses in the design of typical systems that have been recently put in place. Finally, the chapter proposes key features that any well-functioning and efficient guarantee system should have if it is to be an effective part of a country's broader financial safety net system in non-crisis periods.

Why deposit insurance for banks as a matter of public policy?

While the adoption of deposit insurance schemes is largely a late-1900s' phenomenon, the concept is not new. The first recorded evidence of government-sponsored 'deposit' insurance scheme, at least in the US, was the New York State safety fund. It was created in 1829 to insure bank notes which were then the major outstanding component of bank liabilities (Chaddock, 1911; Thies and Gerlowski, 1989; and Calomiris, 1989). Five other states created funds before the Civil War of 1865, but closed them before the turn of the century (Calomiris, 1989). State-sponsored systems were again created after 1908, following the panic of 1907, but all fell by the wayside by 1930 owing to economic downturns and accompanying banking panics (English, 1993). Some privately funded and managed but publicly perceived as state-sponsored funds were resurrected for smaller state-chartered banks and thrifts after World War II. But, they also eventually failed. For example, a second Ohio fund for thrifts failed in the 1980s and Rhode Island's fund failed in the early 1990s (Pulkkinen and Rosengren, 1993; Kane, 1987; and Todd, 1994).

The first central-government-sponsored deposit insurance system was established in Czechoslovakia in 1924 followed in 1933 in the US with the establishment of the Federal Deposit Insurance Corporation (FDIC) (Camara and Montes-Negret, 2006). The FDIC was created as one of a number of public initiatives instituted in response to the calamity of the Great Depression and the inability of the Federal Reserve to prevent widespread bank failures, which totaled nearly 10,000 between 1929 and 1933.

Deposit insurance was intended to mitigate the perceived negative externalities associated with bank failures—namely, bank runs, credit losses to depositors,

disruptions to the payments' system, spillovers to other banks, and reductions in the funding of economic activity. The insurance system was designed especially to assure small (retail) depositors of banks that they would be protected from loss should their bank become insolvent. In this way, depositors would have few incentives to run on their banks, thereby preventing banking panics that might transmit problems in one bank to others and possibly the banking system as a whole, and adversely impact on economic activity (Demirgüç-Kunt and Detragiache, 1998). It was felt necessary to create a targeted safety net for banks. This is despite the fact that many other firms were also experiencing difficulties at that time but were not afforded similar treatment because there was and still is the perception that banks are 'special'.

Are banks 'special'?

The rationale for a government-supported deposit insurance scheme is rooted in the view that banks were and are today 'special' when compared to other business entities because of greater positive externalities that flow to the economy from their activities (Corrigan, 1982). That is, healthy banks are critical to a well-functioning and growing real economy for a long list of possible reasons that includes:

- banks provide a large share of financing for consumers, business firms, and governments;
- banks operate much of the nation's payments' system that transfers payments from buyers to sellers on a timely and certain basis; and
- banks function as the primary transmitter of monetary policy to the economy.

It follows that a healthy and efficient banking system is a prerequisite for a healthy and efficient macroeconomy. Problems in the banking system that generate large credit or other losses and interfere with the smooth operation of the bank activities above have important adverse effects, not only directly on the failed bank's customers but also on the health of the macroeconomy in the bank's market area and possibly beyond. Thus, bank failures, which de facto represent official acknowledgment of legal insolvencies, are widely perceived as more costly both to bank customers and to the economy at large than the failure of other firms of similar size, and thereby require special prudential public policy attention. Moreover, unlike most other firms, banks tend to be closely interconnected through interbank deposits and loans. Thus, the failure of one bank is perceived to have the potential to spill over and infect other financially healthy banks (Kaufman, 1988; and Benston, et al., 1986). Losses may thus be transmitted along a chain of banks and

the resulting damage may spread well beyond the initial failed bank's market area and its customer base.

Banks are perceived to be more fragile and more susceptible to failure than most other firms because of four characteristics of their balance sheets:

- they have a high proportion of demandable debt (demand deposits) and other short-term debt relative to total debt;
- the duration of their assets is typically longer than the duration of their liabilities;
- they have a relatively small proportion of cash assets to total assets;
- they are highly leveraged with little capital relative to assets.

The first characteristic makes banks highly susceptible to runs by depositors. It permits a large number of depositors simultaneously to attempt to withdraw funds with little if any advance notice. The threat of quick withdrawal of funds can serve as a useful source of market discipline on bank management (Calomiris and Kahn, 1991). However, problems may arise when large number act on this threat and actually try to withdraw their funds at the same time, because they believe that the bank may be insolvent or heading toward insolvency and may not be able to repay all depositors in full and on time. Thus, if there is uncertainty about availability of funds, the incentive is for depositors to withdraw funds. Those that are at the head of the line are more likely to get paid in full than those at the back of the line. The second characteristic makes banks highly susceptible to both losses should interest rates rise and to 'fire-sale' losses should depositors withdraw funds before the assets mature. The third characteristic defines fractional reserve banking and implies that banks experiencing runs are unlikely to have sufficient cash immediately to pay all the claims of the running depositors. The banks may have to sell some of their earning assets. But, the less liquid are these assets and the faster the sales need to be made to meet the depositor demands, the larger would be any 'fire-sale' losses. The last factor implies that banks are not able to absorb very large 'fire-sale' losses before running through their entire capital buffer and becoming insolvent and unable to repay the remaining deposits on time and in full. At that point, the bank fails.[1]

In this scenario, the appeal of deposit insurance is evident. When deposits are fully insured by a credible insurance agency then depositors will have fewer incentives to run, even when a bank's assets are perceived to be insufficient to repay all depositors in full and on time. This reduces the need for banks to sell assets quickly and suffer 'fire-sale' losses that could drive a bank into insolvency, generating losses to their customers, potentially transmitting the problem to other banks, and interfering with the efficient operation of the payments' system (Rochet, 2004).

[1] Minsky (1977) argues that banks are inherently unstable. See also Diamond and Dybvig (1983).

The downside of deposit insurance

But, deposit insurance has a downside.[2] It can induce moral hazard behavior by insured institutions and poor agency behavior by regulators. Unless deposit insurance is priced and administered correctly by the insurance agency, it is likely to encourage increased risk-taking behavior by banks by enabling them to operate with less capital and/or to adopt riskier portfolios than they would in the absence of insurance. In particular, insurance can exacerbate endgame gambles when banks approach or enter insolvency. Should these gambles pay off and the institution returns to financial health, managers and shareholders reap the benefits. Should the institution fail, which is likely on average, the insurer suffers the losses. When the insurance is provided or ultimately supported by central governments, poor behavior in the form of forbearance by banking supervisors exacerbates the moral hazard problem, and by permitting insolvent banks to continue to operate increases the losses that must ultimately be borne by taxpayers (Demirgüç-Kunt and Kane, 2001). Knowing that depositors are less likely to run and withdraw funding from their banks permits regulators not only to delay imposing sanctions on troubled institutions but also to postpone legally closing banks by withdrawing their charters and placing them in receivership that would otherwise occur as soon as insolvency became evident and depositors began withdrawing their funds. In the absence of deposit insurance, depositors, like creditors of any firm, would monitor the financial performance of their banks and discipline them if they fail to maintain agreed-upon measures of financial health either by charging a higher interest rate for their deposits or by withdrawing their deposits. Withdrawing funding may force the bank to suspend its operations. The possibility of either higher funding costs or loss of funding should induce managers to correct their problems well before a run actually begins.

Unfortunately, the record of deposit insurance in almost all countries to date is that it has been on the whole poorly structured and administered. As a result, moral hazard behavior by banks and poor agency behavior by regulators have caused deposit insurance to be excessively costly to the insurance agency and often taxpayers.

The record of deposit insurance has been mixed

In the US, both before and after the introduction of the FDIC, state-sponsored deposit insurance was largely a failure. As noted earlier, all state-sponsored funds either failed or were disbanded. None remains today. The main reason for this was that the funds available were insufficient to deal with a crisis when it occurred, and in some instances funds were inadequate to pay off depositors at even a single large institution failure. Even when the insurance funds had either explicit or implicit state backing, which

[2] Kane, 1989a.

meant that taxpayers in the state would have to make up the difference, state legislatures often lacked either the will or the means to make good on the guarantee. For example, when the Ohio deposit insurance fund collapsed in 1988, the legislators in Ohio failed fully to support the fund, and many of the affected institutions were eventually transferred to the Federal Deposit Insurance Corporation.

On the other hand, federally-sponsored deposit insurance for both banks and thrifts in the US appeared to have worked well for much of its early history. Indeed, federal deposit insurance was hailed by Milton Friedman and Anna Schwartz in 1963, thirty years after its introduction, as one of the most important and beneficial pieces of financial legislation in US history. The ability of federal deposit insurance to prevent runs by insured depositors was evident immediately. However, it was not until the onset of the thrift crisis in the late 1970 and early 1980s that the weaknesses in deposit insurance that were described earlier became exposed. The thrift institution insurance fund at the time, the Federal Savings and Loan Insurance Corporation or FSLIC, became insolvent and the federal government had to step in and bail out the depositors at insolvent thrift institutions. Taxpayers contributed more than $150 billion, or nearly 3 percent of US GDP at the time. In response, among other legislation, Congress enacted the Federal Deposit Insurance Corporation Improvement Act (FDICIA) of 1991. That Act significantly reformed the nature of the US deposit insurance, carefully defined the role and duties of the responsible supervisors in monitoring and resolving trouble institutions, and imposed greater accountability on the regulators for losses imposed upon the taxpayers.

The current US structure

Starting in 1989, a series of legislative acts reorganized the insolvent insurance fund for thrift institutions and housed it in the FDIC, which administered the bank fund (Kaufman, 2002). The two funds were subsequently merged into one in 2006. The fund is funded by a combination of *ex ante* and *ex post* premiums paid by insured institutions. The funds were initially assigned a minimum reserve ratio that they were required to maintain, which was 1.25 percent of insured deposits. If the reserve ratio declined below the 1.25 percent target, the FDIC was required to increase premiums on insured institutions to recapitalize the fund to at least this ratio within a brief period of time. The 2006 legislation eliminated the 1.25 percent hard target percent and established a range between 1.15 percent and 1.50 percent each year. Should the reserve ratio fall below 1.15 percent, the FDIC is required to establish a recapitalization plan to bring it back to 1.15 percent within a longer five years. However, if the FDIC's reserve ratio exceeds 1.35 percent, the FDIC generally

must pay dividends to banks that in aggregate equal to one-half of the amount by which the fund exceeds 1.35 percent of estimated insured deposits. If the reserve ratio exceeds 1.50 percent, the FDIC must pay dividends equal to the full amount by which the fund exceeds 1.50 percent.

Thus, the fund is effectively a mutual organization; losses to the FDIC from bank failures are shared among the surviving banks until the combined capital of the industry is exhausted. Only then, without a change in legislation, would the Treasury and taxpayers be the guarantors of insured deposits. In contrast, before FDICIA in 1991, premiums were difficult for the insurance agencies to increase, even when the FDIC and the FSLIC incurred losses. Thus, the US Treasury and taxpayers served as more immediate guarantors of insured deposits.

FDICIA required insurance premiums to be based on bank risk, rather than levied as a flat percentage of bank asset size, as previously. From 1992 to 2005, the risk metric was a simple 9-cell formula based on the combination of a bank's CAMELS rating[3] from regulators and its capital strength as defined in FDICIA's prompt corrective action (PCA) provisions. Thereafter, the premium formula became more complex. It was based primarily on a probability of supervisory downgrade model developed by the FDIC using capital and other financial ratios and CAMELS ratings. For larger banks, the formula also used the banks' credit ratings from the rating agencies.

The FDIC was also prohibited by the FDICIA from protecting *de jure* uninsured depositors—deposits in excess of $100,000—and other creditors and equity owners. The one exception to this prohibition is the case is the so-called 'systemic-risk exemption', where not protecting these debt claimants would threaten financial and economic instability, and protecting them would prevent this. But, invoking this systemic-risk exemption is not easy. It requires, among other actions, a two-thirds vote in favor by both the board of directors of the FDIC and the Board of Governors of the Federal Reserve, approval by the Secretary of the Treasury after consultation with the President, and the FDIC levying a special assessment on all banks to reimburse the FDIC for its losses.

The financial crisis of 2007–09 changed things dramatically. Deposit insurance components were expanded rapidly in response to deteriorating conditions in financial markets and institutions, first on brief temporary bases but then extended as conditions continued to deteriorate. But, as expansions in coverage once introduced are politically difficult to take back, these changes are likely to be more or less permanent. They include expanding coverage to all transaction deposits and from $100,000 to $250,000 for other deposits and guaranteeing new bank debt issues. Because of large losses to the FDIC, a surcharge was introduced for deposit insurance and a separate charge for debt guarantees. The FDIC's ability

[3] The components of a bank's condition that are assessed: (C)apital adequacy, (A)sset quality, (M)anagement, (E)arnings, (L)iquidity, and (S)ensitivity to market risk.

to borrow temporarily from the US Treasury was increased from $30 billion to $100 billion and to $500 billion under unusual circumstances. Lastly, the systemic risk exemption was invoked a number of times to protect uninsure depositors, creditors, and derivative counterparties. Because the crisis was not over by June 2009 and it was too soon to analyze objectively, this chapter limits itself only to the pre-2007 period.

What were the lessons for from the US thrift crises?

The US experience with both state-sponsored and federally-sponsored deposit insurance plans has several lessons for how insurance schemes both in the US and abroad should be structured, financed, and governed in non-crisis periods. This includes how the regulatory and supervisory process operates and how the resolution of a failure proceeds. Particularly important is the legal structure and ability of the supervisors legally to close and resolve troubled institutions promptly. They do this by revoking the banks' charters, placing them in receivership, and disposing of their assets and liabilities before the losses to the uninsured depositors and the FDIC become large, and without interrupting the provision of their important services.

Lessons for deposit insurance fund structure and funding

Recent events confirm that the existence of deposit insurance per se provides little assurance that either the frequency or costs of banking crises will be reduced, especially without the appropriate processes and procedures in place. Establishing a deposit insurance fund alone does not guarantee that it will be adequate to prevent or deal with a financial crisis or to weather a series of failures or even the failure of a major institution. The insurance fund itself may be viewed as a petty cash fund that provides the managers of the fund the flexibility to deal with limited problems without having to go to the legislature or taxpayers for needed resources to make the guarantees credible to deal with large problems.

How deposit insurance funds have been funded

Countries have adopted various structures for providing resources to fund their deposit insurance schemes. Some are totally private, some are totally public, and others are a mix of the two (see Eisenbeis and Kaufman, 2008 for details of

the deposit insurance structures in the European Union, and Demirgüç-Kunt, Karacaovali, and Laeven, 2005 for a description of systems in other parts of the world). Some guarantee systems have no actual monies, but instead rely upon *ex post* premium levies on surviving participants, while others have actual monies collected from *ex ante* funding. A few have risk-related premiums; while the majority has some form of a flat-rate funding plan. Finally, some have an explicit draw on the government, while other schemes, such as in the US, try to reduce the potential injection of public monies. Germany prohibits the injection of public monies.

As was pointed out earlier, the deposit insurance system in the US is now effectively an industry funded mutual institution, which by law is backed by the equity of the overall banking industry. The Fund also has a small line of credit with the US Treasury that was recently increased. In normal times, the scale of the banking system relative to most of the risks that even large individual banks are likely to pose to the banking system's equity, should a failure occur, should be manageable provided the supervisory agencies follow their mandate to implement prompt corrective action and structured early intervention and prompt legal-closure resolution polices as mandated by FDICIA (Benston and Kaufman, 1994).

Many of the deposit insurance funds being put in place in other countries are both small and often meager relative to the risks that failure of a large institution might pose, especially since most of the countries do not have accompanying PCA or structured early intervention and resolution (SEIR) features. This suggests that making good on their guarantees will require the potential commitment of taxpayer funds. As was pointed out earlier, the US experience with state-sponsored funds shows that legislatures are not always willing or able to live up to their perceived commitments. It is interesting to note that in the case of the individual country deposit insurance systems in the European Union all but one of the accession countries has a smaller GDP than did the State of Ohio when its fund collapsed. With large, cross-border banking organizations operating within the European Union that are often headquartered in relatively small countries, like the Netherlands and Luxembourg, the ability or willingness of those countries to make good on their guarantees to depositors at their banks in other European Union countries may be problematic. The current crisis confirms the need for taxpayer funds.

Lessons for resolution powers and bankruptcy laws

The resolutions powers and processes that enable the administrators of the system quickly to limit the costs of failure to shareholders rather than on either bank depositors or other customers is widely underappreciated. This has become readily evident from the recent experience in the UK with the run on Northern Rock. The UK had a federally mandated deposit insurance scheme that covered 100 percent of

the first £2,000 (about $3,000) and 90 percent of the next £33,000 (about $50,000) as compared with the US's 100 percent coverage of up to $100,000 at that time.

Responsibility for supervising Northern Rock lay with the UK's Financial Services Authority, while lender of last resort and responsibility for financial stability lay with the Bank of England. In addition, the Chancellor of the Exchequer had the authority to commit public funds, should it be necessary. Neither of the first two entities had the legal authority to close Northern Rock and revoke its license should it be actually insolvent. Outside of nationalization its fate was entrusted to the bankruptcy courts, as is the case in virtually all countries outside the US. In many systems, the tension among the interests of insured depositors and uninsured creditors and shareholders all but guarantees substantial delays when resolution of a failed institution is administered by the courts (Bliss and Kaufman, 2007).[4] In most European countries, the laws establishing deposit insurance provide that insured funds must be made available to depositors within three months, but often one or more extensions are permissible. A relatively low full guarantee, combined with considerable uncertainty about when a troubled institution would be turned over to the courts and when even insured depositors would receive their designated funds, meant that it was rational for British depositors to run at the first hint of trouble to avoid either losing part of their funds or ongoing access to their funds. The specter on British television of lines forming at branches of Northern Rock left the UK Treasury little choice but to step in and guarantee 100 percent of the deposits at Northern Rock and to suggest that other institutions would be granted similar guarantees should that be necessary. The bank was eventually nationalized.

The Northern Rock incident shows that runs can occur, even when there is a guarantee scheme in place and even when there are strong assurances from the government that it will make good on the guarantee. The case also points out a fundamental weakness in virtually every deposit guarantee scheme, but particularly in those that rely upon general bankruptcy statutes and the courts to resolve troubled institutions. Faced with long and uncertain delays in obtaining the use of their funds as well as suffering possible credit losses, it becomes rational for both insured and uninsured depositors to run. It also became clear that there is little benefit to co-insurance as a source of market discipline for retail depositors. Small (retail) depositors ran in Northern Rock's case long after funding had dried up to the institution in the interbank and commercial paper markets, when questions arose about the quality of its assets. They were the last to appreciate the depth of the problem and the run was a symptom, not cause of Northern Rock's difficulties.

In the 1800s, the US realized that the most adverse spillover effects from bank failures could be avoided by making banking failures isolated events. Importantly, insolvent banks needed to be legally closed more quickly than other firms to reduce

[4] Goodhart (2005) addresses the issue of too many regulators involved in banking resolutions.

both runs—in those days by note holders trying to convert their notes into gold or silver specie—and losses. Furthermore, the closure decision should be made by regulators not by the courts, to avoid untimely delays. Thus, it enacted a special bank bankruptcy code separate and very different from the general corporate bankruptcy code. Legal closure powers were given to the chartering authority to revoke an institution's charter and to place it in receivership.

Congress also made it clear who was responsible for the resolution of a failed institution. Regardless of which regulatory agency in the US was responsible for supervising an institution, once legal closure procedures are evoked the FDIC is immediately appointed receiver for federally-insured institutions and directs the resolution process. Thus, there is no uncertainty as to where the responsibility lies for the resolution of a failed institution.

The US experience has led to another unique feature of the US system and that is the increased emphasis by the FDIC on providing all depositors at legally closed banks near-immediate access to all or some of their accounts, so that the deposits are not frozen until the bank or its assets are sold. Legal closure is separated from physical closure. The FDIC pays insured depositors the full value of their accounts currently up to $250,000 on most accounts the next business day after legal closure. This keeps most retail depositors politically off the streets and simplifies the resolution process. Combined with minimizing the loss to uninsured depositors, this reduces the incentive for depositors to run. At the same time, the FDIC may pay uninsured depositors an advance against the estimated recovery value of the bank's assets before the proceeds are received. Thus, uninsured depositors also have immediate access to at least some of their funds.

In short, the aim of the bank bankruptcy code in the Federal Deposit Insurance Act is to reopen failed institutions as new or merged institutions with as small credit losses as possible and as quickly as possible—usually over the weekend—so as to make funds available to depositors and loan customers as promptly as possible (Bliss and Kaufman, 2007). To facilitate this process for very large banks, the FDIC has authority to charter a temporary bridge bank to assume the activities of an insolvent large bank, for which a private purchaser could not be found in time. This power is now being advocated for 'systemically important' institutions.

Lessons on monitoring and supervisory responsibilities

The US experience also carries with it important lessons concerning the roles that timely and effective monitoring as well as effective risk-control systems play in an effective deposit guarantee system.

Timely and effective monitoring and attention to valuation of banks assets combined with appropriate remedial responses when problems are discovered are key to protecting the deposit insurance fund and ultimately the taxpayer against

losses. One of the lessons of the US thrift crisis of the 1980s—and also the current 2007–9 crisis—was that supervisory and regulatory responses to difficulties often lead to forbearance and the use of accounting gimmicks to delay or avoid recognition of losses and insolvencies. As noted, the ultimate Congressional response was enactment of the Federal Deposit Insurance Corporation Improvement Act in 1991. That Act emphasized prompt corrective regulatory actions and provided supervisors with stronger tools to turn troubled banks around before insolvency. In particular, the Act attempted to refocus regulatory attention on remediation of troubled institutions as their capital declined rather than after they had become insolvent and imposed mandatory and well as discretionary regulatory sanctions. Banks are encouraged to hold sufficient capital and regulators are incentivized to take actions promptly should capital fall below predefined levels. However, if a bank does not respond and lets its equity capital decline below 2 percent of its assets the regulators are legally obligated to close or otherwise resolve the bank within a maximum of nine months.

Recent evidence from the 2007–9 financial crisis suggests that some regulators apparently failed to learn this lesson. In a letter to the US Senate Finance Committee, Eric Thorson, Inspector General of the US Treasury, reported that the Office of Thrift Supervision (OTS) permitted IndyMac Bank to post-date a capital injection received from its parent holding company in May 2008 to 31 March 2008 (Thorson, 2008). Had this been reported accurately, IndyMac, which failed in July 2008 with an estimated loss of $9 billion to the FDIC fund, would have had insufficient capital to qualify as 'well-capitalized', sanctions may have been imposed earlier, and eventual losses may have been reduced. The Inspector General also noted similar backdating of capital injections in at least five other OTS-supervised institutions.

As noted above, in the event that supervisory efforts to engineer a turnaround of a troubled institution fail, then under PCA in the US system the appropriate federal regulatory agency is legally empowered to close the institution before its capital is depleted and the FDIC is empowered to sell or liquidate the bank. If successful, losses would be confined to shareholders, who are paid to assume risk, while depositors and other creditors would remain whole. Thus, unlike deposit insurance that only shifts credit losses from insured depositors to the insurer or taxpayers, PCA and the associated legal closure rule at positive capital attempts to eliminate credit losses to depositors. However, to be successful the program requires timely and effective implementation by the regulators.

Risk-related deposit insurance premiums are used in the US and only a few other countries. The remainder levy flat-rate premiums. More prevalent, however, are systems of risk-based capital requirements to control bank risk taking.

Much of the bank supervisory community has recently placed greater emphasis on risk-based capital standards—such as under Basel I and Basel II. Risk-based capital standards essentially try to determine the appropriate level of capital for an

institution for assuming credit risk, operations risk, market risk, and interest rate risk exposures (Shadow Financial Regulatory Committee, 2001; and Shadow Financial Regulatory Committee, 2003). However, this emphasis is not only misplaced but also is counterproductive. Worrying about the appropriate allocation of capital in a healthy institution is the job of its management and not of banking supervisors. The focus of supervisors under PCA is on ensuring that an institution whose capital and financial position is deteriorating turns itself around before insolvency occurs and that the regulators have sufficient time legally to close a bank before its capital is fully depleted.[5]

Northern Rock serves as a poster child for why Basel capital standards are misdirected. At the end of June 2007, Northern Rock reported a capital-to-asset leverage ratio only a fraction above the 2 percent critical level that would have required US regulators to initiate a legal closure process (Northern Rock, 2007). Yet, the UK regulators effectively rewarded Northern Rock in June by permitting it to compute its minimum capital requirement according to the Basel II advanced internal rating approach. This reduced its risk-based regulatory capital requirement substantially, much of which was allocated to increasing the bank's dividend payout by 30 percent. With such leverage, and given its heavy reliance upon short-term capital market funding, even a small change in spreads against it would significantly erode the bank's remaining reported capital, and quickly raise solvency questions, if the institution was indeed solvent at the time.

There are many difficulties with how the risk-related premiums and capital standards have been established in the US. They share virtually the same flaws. Both risk premiums and risk-based capital standards have been made unduly complex and mechanistic. They are also not market based, and thus are subject to political manipulation, such as granting favorable risk classifications to sovereign debt or home mortgages. They place too much reliance upon the asset-quality ratings of the external rating agencies, and, in the case of risk-based capital standards, too much credence is given to the risk assessments in institutions' internal models. Furthermore, they focus—especially in the case of risk-based premiums—almost totally on the probability of default and risks to equity and other stakeholders and not on the loss given default, which is the main risk to the deposit insurance fund. Finally, as the recent problems in the subprime mortgage and derivatives markets have suggested, important sources of risks that affect how institutions actually behave when their reputations are at stake have not be considered adequately from either an accounting or a supervisory perspective.

Particularly for small banks, for which fraud is the major cause of failure, risk-based premiums or capital standards are unlikely to be related to the FDIC's actual loss exposure. In fact, since the institution of FDICIA's requirement that

[5] There are other problems with risk-based capital standards as well. See Shadow Financial Regulatory Committee (2001; 2003) for criticisms.

supervisors close troubled institutions once their capital falls below the critical value of 2 percent, the loss to the FDIC from bank failure depends largely on how successful the regulators are in implementing the FDICIA's PCA requirements rather than on the risk exposure of the banks. With a well-functioning PCA legal closure process at positive capital, there is little justification for imposing premiums based on the riskiness of the bank. Rather, the real risk is that the FDIC will not legally close the bank before its capital is depleted so that credit losses occur. For this reason it is not at all clear why insured institutions should be charged for this regulatory risk or that the risk is a function of bank portfolio composition. This view of risk contrasts with the original idea behind risk-based premiums, which was driven by the view that, when properly set, the premiums would serve as a deterrent to risk taking by banks that would supplement the supervisory monitoring process (Pennacchi, 2001; Duan and Yu, 1994; Flood, 1990; Chan, Greenbaum, and Thakor, 1992; Flannery, 1991; and Allen and Saunders, 1993). However, this focus was misplaced because it ignored the fact that there was an important distinction between portfolio risk, capital structure, and the probability of default on the one hand and the losses that the insurance fund might have to bear, should a default occur on the other. Experience has suggested that the chief determinant of the losses to the fund beyond fraud is regulatory risk, which is a function of the examination frequency and policies and how quickly an institution is closed, once it becomes insolvent.[6] That is, if the premiums are to be based on risk, they should be based primarily on the risk of delayed closure by the FDIC and not on the risk of individual banks except to the extent that riskier banks may be more difficult for the FDIC to monitor. Flannery (1991) addresses this problem head-on in a model that includes a separate monitoring charge as well as a risk-related premium.

Summary and conclusions: What should an effective deposit insurance system look like?

Deposit insurance schemes have become an integral feature of the banking landscape and an important component of virtually every nation's financial safety net. In essentially every case, however, the guarantee systems that had been in place suffered from structural deficiencies. In addition, as a consequence of the financial turmoil

[6] There are also other problems with the FDIC's risk-based premium system. For example, premiums should be based upon just-insured deposits and not total deposits, since uninsured deposits share in loss absorption when failure occurs.

of 2007–9, many countries have temporarily provided 100 percent guarantees of bank deposits and other forms of credit. But, these countries will be faced with the need to reform their deposit insurance systems, once the crisis has abated, to avoid the moral hazard that 100 percent guarantees entail. In considering possible reforms, it should be emphasized that an effective system consists not only of a credible guarantee (which in some countries is more dependent upon a government commitment than the existence of a deposit insurance fund or the ability to draw on the industry's equity) but also effective monitoring by supervisors, who are incentivized to protect the insurers and taxpayers against losses, and a legal system that empowers them both to intervene effectively before a troubled institution becomes insolvent and legally to close them before their capital is depleted, and which, moreover, provides for quick resolution when failure occurs. We have argued elsewhere (Eisenbeis and Kaufman, 2008) that efficient resolution of insolvent and near-insolvent banks can be achieved when four conditions exist.

First, a regulatory insolvent institution that could not be recapitalized should be legally closed—charter revoked—when its equity-capital-to-asset ratio declines to a pre-specified and well-publicized positive minimum, so that, if successful, no losses accrue to depositors. Once closed, recovery values should be estimated quickly and any credit losses from delayed closure assigned to *de jure* uninsured bank claimants. There is also loss sharing with the FDIC. Institutions that have been closed by supervisors should be reopened as soon as possible, preferably the next business day, to enable depositors full access to their accounts on their due dates at their insured or estimated recovery values and performing borrowers access to their pre-established credit lines. For large banks that cannot be sold quickly, these activities may be transferred to a newly chartered FDIC-operated temporary bridge bank. Finally, the institution should be re-privatized with adequate capital so that it does not fall into insolvency again quickly. Under such a system both credit and liquidity losses that generate the widespread fear of bank failures are minimized, deposit insurance becomes effectively redundant, and the adverse moral hazard incentives inherent in deposit insurance become benign. Credit losses are minimized, if not eliminated, rather than simply shifted to the insurer. But, properly structured deposit insurance should be maintained in place for retail deposits in case regulators may not be able legally to close institutions in time before their capital turns negative. Redundancy has its uses. Moreover, the system only works to the extent that regulators are faithful agents for their healthy bank and taxpayer principals and enforce the rules in place in a timely manner.

Going forward, it is also clear that before the beginning of the crisis world financial systems were becoming more integrated, more dependent upon capital markets, and less reliant upon banks for the bulk of financial intermediation. As a result, once the crisis is put behind us, deposit insurance for bank liabilities only is likely to again assume a progressively smaller role as a key component of the overall financial safety net. Prudential financial regulatory agencies, to the extent they are

charged with maintaining financial stability, will probably have to rely upon tools and techniques that extend beyond the current bank deposit insurance schemes. In addition, the rapid expansion of cross-border banking arrangements pose special burdens on regulators and deposit guarantee structures. Differing legal structures and deposit guarantee systems complicate the resolution and settlement of claims when large cross-border institutions experience financial stress. Such an expansion of prudential regulatory responsibilities will require greater thought to safety net and regulatory design than is developed in this chapter.

References

Allen, L. and A. Saunders (1993). 'Forbearance and Valuation of Deposit Insurance as a Callable Put', *Journal of Banking & Finance,* 17: 629–43.

Benston, G. J. and G. G. Kaufman (1994). 'The Intellectual History of FDICIA of 1991', in George Kaufman (ed.), *Reforming Financial Institutions and Markets in the United States*, Boston: Kluwer.

—— R. A. Eisenbeis, P. M. Horvitz, E. J. Kane, and G. G. Kaufman (1986). *Perspectives on Safe and Sound Banking*, Cambridge, MA: MIT Press.

Bliss, R. R. and G. G. Kaufman (2007). 'U.S. Corporate and Bank Insolvency Regimes: Comparison and Evaluation', *Virginia Law and Business Review*, Spring, 143–77.

Caamara, M. K. and F. Montes-Negret (2006). 'Deposit Insurance and Banking Reform in Russia', World Bank Policy Research Working Paper 4056, November.

Calomiris, C. (1989). 'Deposit Insurance: Lessons from the Record', *Economic Perspectives*, Federal Reserve Bank of Chicago, May/June, 10–30.

—— and C. M. Kahn (1991). 'The Role of Demandable Debt in Structuring Optimal Banking Arrangements', *American Economic Review*, June, 497–515.

Chaddock, R. E. (1911). 'The Safety Fund Banking System in New York State, 1829–1866', in *Publications of National Monetary Commission,* 2: 227–388. Washington, DC: Government Printing Office.

Chan, Y. S., S. I. Greenbaum, and A. V. Thakor (1992). 'Is Fairly Priced Deposit Insurance Possible?', *Journal of Finance,* 47: 227–45.

Corrigan, G. (1982). 'Are Banks Special?', *Annual Report*, Federal Reserve Bank of Minneapolis.

Demirgüç-Kunt, A. and E. Detragiache (1998). 'The Determinants of Banking Crises in Developing and Developed Countries', *IMF Staff Papers,* 45: 81–109.

—— and E. J. Kane (2001). 'Deposit Insurance around the Globe: Where Does It Work?', World Bank Policy Research Working Paper 2679.

—— B. Karacaovali, and L. Laeven (2005). 'Deposit Insurance around the World: A Comprehensive Database', World Bank Policy Research Working Paper 3628, June.

Diamond, D. W. and P. H. Dybvig (1983). 'Bank Runs, Deposit Insurance and Liquidity', *Journal of Political Economy,* 91: 401–19.

Duan, J. and M. T. Yu (1994). 'Forbearance and Pricing Deposit Insurance in a Multiperiod Framework', *Journal of Risk and Insurance,* 61: 575–91.

Eisenbeis, R. A. and G. G. Kaufman (2008). 'Cross Border Banking: Challenges for Deposit Insurance and Financial Stability in the European Union', in *Building the Financial Foundations of the Euro: Experiences and Challenges,* Lars Jonung, Christoph Walkner, and Max Watson (eds.), New York and London: Taylor & Francis.

English, W. B. (1993). 'The Decline of Private Deposit Insurance in the United States', *Carneigie-Rochester Conference Series on Public Policy,* 38: 57–128.

Flannery, M. J. (1991). 'Pricing Deposit Insurance When the Insurer Measures Bank Risk with Error', *Journal of Banking & Finance,* (September), 975–98.

Flood, M. (1990). 'On the Use of Option Pricing Models to Analyze Deposit Insurance', *Federal Reserve Bank of St. Louis Review,* 52: 217–32.

Friedman, M. and A. Schwartz (1963). *A Monetary History of the United States*, 1867–1960, Princeton, NJ: Princeton University Press.

Goodhart, C. A. E. (2005). 'Multiple Regulators and Resolutions', in *Systemic Financial Crises: Resolving Large Bank Insolvencies,* D. Evanoff and G. Kaufman (eds.), Singapore: World Scientific.

Kane, E. J. (1987). 'Who Should Learn What from the Failure and Delayed Bailout of the ODGF?', 1987 Proceedings of Conference on Bank Structure and Competition, Chicago: Federal Reserve Bank of Chicago, 306–26.

—— (1989). *The S&L Mess: How Did It Happen?*, Washington, DC: Urban Institute Press.

—— (1993). 'Taxpayer Loss Exposure in the Bank Insurance Fund, *Challenge*, 36 (March–April), 43–9.

—— (2000). 'The Dialectical Role of Information and Disinformation in Banking Crises', *Pacific Basin Finance Journal*, 8 (July), 285–308.

—— Armen Hovakimian, and Luc Laeven (2003). 'How Country and Safety Net Characteristics Affect Bank Risk-Shifting', *Journal of Financial Services Research* (June).

—— and Min-Teh Yu (1996). 'How Much Did Capital Forbearance Add to the Cost of the S&L Insurance Mess?', *Quarterly Review of Economics and Finance*, 36: 189–99.

Kaufman, G. G. (1988). 'Bank Runs: Causes, Benefits, and Costs, *Cato Journal*, 7: 559–87.

—— (1992). 'Capital in Banking: Past, Present and Future', *Journal of Financial Services Research*, April, 385–402.

—— (2002). 'FDIC Reform: Don't Put Taxpayers Back at Risk', *Policy Analysis (Cato Institute)*, April 16.

—— (2004). 'Basel II: The Roar that Moused', in Benton Gup (ed.), *The New Basel Capital Accord*, New York: Thomson.

—— (2006). 'Depositor Liquidity and Loss Sharing in Bank Failure Resolutions', *Contemporary Economic Policy*, 22: 237–49.

—— and S. A. Seelig (2002). 'Post-Resolution Treatment of Depositors at Failed Banks: Implications for the Severity of Banking Crises, Systemic Risk, and Too Big to Fail', Federal Reserve Bank of Chicago, *Economic Perspectives,* 2nd Quarter: 27–41.

Minsky, H. P. (1977). 'A Theory of Systematic Financial Fragility', in E. J. Altman and A. W. Sametz (eds.), *Financial Crises: Institutions and Markets in a Fragile Environment.* New York: Wiley.

Northern Rock (2007). 'Northern Rock PLC Interim Results', Six Months Ending 30 June 2007, http://companyinfo.northernrock.co.uk/investorRelations/results/stockEx070725.asp.

Pennacchi, G. G. (2001). 'Estimating Fair Deposit Insurance Premiums for a Sample of Banks under a New Long-Term Insurance Pricing Methodology', in *The Financial Safety*

Net: Costs, Benefits, and Implications for Regulation, *Proceedings of the 37th Annual Conference on Bank Structure and Competition,* 756–76, Federal Reserve Bank of Chicago.

Pulkkinen, T. E. and E. S. Rosengren (1993). 'Lessons from the Rhode Island Banking Crisis', Federal Reserve Bank of Boston, *New England Economic Review* (May/June).

Rochet, J. C. (2004). 'Bank Runs and Financial Crises: A Discussion', in S. Bhattacharya, A. W. A. Boot, and A. V. Thakor (eds.), *Credit Intermediation and the Macro Economy,* Oxford: Oxford University Press.

Shadow Financial Regulatory Committee (2000). 'Reforming Bank Capital Regulation', Washington, DC: American Enterprise Institute, March.

——(2001). 'The Basel Committee's Revised Capital Accord Proposal', Statement 169, February 26.

——(2003). 'The Latest Revision to Basel II and Implementation Plans in the United States', Statement 193, May 5.

Thies, C. F. and D. A. Gerlowski (1989). 'Deposit Insurance: A History of Failure', *Cato Journal*, 8 (Winter).

Thorson, Eric (2008). 'Letter to the Honorable Charles Grassley, Inspector General', Department of the Treasury, London, December 22.

Todd, W. F. (1994). 'Lessons from the Collapse of Three State-Chartered Private Deposit Insurance Funds', *Federal Reserve Bank of Cleveland Economic Commentary.*

CHAPTER 14

RISK-BASED REGULATORY CAPITAL AND BASEL II

MICHAEL B. GORDY
ERIK A. HEITFIELD[1]

Motivation: Benefits of coordinated, risk-based bank capital regulation

Bank solvency regulation is intended to reduce systemic risk and deadweight loss associated with bank failures, and to address moral-hazard problems arising from implicit or explicit government guarantees that interfere with effective market discipline. Bank supervisors in most countries impose minimum capital adequacy standards on banks as an important component of banking regulation. Banks that breach these standards may be subject to supervisory action, and, in extreme cases, liquidation.

[1] The views expressed here are solely those of the authors and do not reflect the opinions of the Board of Governors of the Federal Reserve System or its staff. We thank Patrick de Fontnouvelle, David Jones, Mary Frances Monroe, and John Wilson (the editor) for helpful discussion and suggestions.

The most straightforward approach to regulating bank capital adequacy is to impose a floor on the ratio of banks' book equity to book assets, but this simple leverage ratio test has a number of critical drawbacks. By treating all banks the same regardless of risk, a simple leverage requirement effectively advantages those banks that invest in higher-risk assets. These banks can anticipate earning higher than expected returns on equity, but they are not required to hold extra capital to protect debt holders from the greater volatility of their investments. Thus, a capital rule that is not sensitive to risk can perversely encourage a bank to take on risk. Furthermore, a rule based solely on accounting leverage is relatively easy to circumvent, since sophisticated banks can efficiently gain exposure to risks that do not appear on their balance sheets by securitizing assets, providing credit guarantees, or trading in derivative securities. For these reasons, regulatory capital requirements typically link the amount of capital a bank must hold to measures of the risks associated with its business activities.[2]

In a globally-integrated financial system, several benefits accrue to coordination in setting bank capital adequacy standards. First, coordination can help to solve the 'prisoner's dilemma' faced by national banking authorities. All supervisors prefer a stable, well-capitalized banking system that is protected from adverse systematic shocks, but each country's banking authority also wants to see its own banks grow and compete aggressively in the international marketplace. Thus, each banking authority would prefer to see all other countries increase bank capital standards, but may have an incentive to advantage its own banks by keeping capital requirements relatively low. By working together to set common standards, national banking authorities can mitigate this coordination problem. Second, coordination can help reduce compliance costs and incentive problems that arise when internationally active banks must comply with different capital requirements in different jurisdictions. Third, a single international standard promotes transparency in the markets by facilitating comparison of banks under different jurisdictions. Finally, international coordination can help to facilitate the adoption of best-practice standards by smaller or less experienced regulatory authorities who may have more limited access to technical expertise.

The 1988 Basel accord

Recognizing the benefits of coordinated action, national banking authorities began work to develop common bank capital adequacy rules in the late 1980s

[2] In the United States, both risk-based capital standards and a simple accounting leverage requirement are imposed on banks.

under the auspices of the Basel Committee on Banking Supervision (the Basel Committee), a committee of G-10 banking authorities. In 1988, the Basel Committee reached its first bank capital accord, now called Basel I. This accord, like all subsequent work by the Basel Committee, was not a treaty and carried no force of law. Then, as now, the Basel Committee relied solely on moral suasion to encourage countries to adopt its recommendations. Nonetheless, today, Basel Committee capital standards are used by virtually all countries with well-developed banking systems.

Basel I established standards for computing a measure of bank capital adequacy called a risk-based capital ratio (RBCR). Basel I's RBCR is similar to a standard equity-to-assets ratio, but it relies on more sophisticated definitions of capital and assets than are typically reported on firm balance sheets. The capital figure used in the numerator of the RBCR includes equity capital as well as limited amounts of loan-loss reserves and certain types of equity-like debt instruments such as convertible bonds. More importantly, from a risk management perspective, the denominator—termed risk weighted assets—is calculated by taking a weighted sum of both on- and off-balance sheet exposures. Following consultations with industry and its own analysis, the Basel Committee (1988: para. 44) determined that a minimum RBCR of 8 percent would be 'consistent with the objective of securing over time soundly based and consistent capital ratios for all international banks'.

In a rudimentary way, the weights used in calculating risk weighted assets under Basel I are intended to reflect the underlying risks associated with different types of exposures. Low-risk OECD sovereign debt securities, for example, are given a zero risk weight, so that banks are not required to hold capital against these assets. Residential mortgages, which are generally highly collateralized, are given a 50 percent risk weight while corporate loans are weighted at 100 percent. By weighting different assets differently, Basel I attempts to link a bank's regulatory capital requirement to the riskiness of its asset portfolio. A bank invested primarily in relatively safe assets such as sovereign debt or mortgages has lower risk weighted assets and thus is required to hold less capital than a comparable bank invested primarily in higher-risk corporate debt.

Basel I's simple risk weighting scheme is transparent and easy for banks and supervisors to implement, but it is somewhat arbitrary and is not sufficient to capture important differences in risk across bank assets. For example, under Basel I, loans to highly rated 'blue-chip' corporations attract the same capital requirements as junk bonds, and all home mortgages attract the same level of capital regardless of the borrower's credit score or the amount of home equity backing the loan. The Basel I framework also is not readily applied to loan securitization arrangements which allow banks to move loans off their balance sheets while retaining much of the credit risk associated with those investments.

Over the years since Basel I was implemented, the limitations of its coarse risk weighting approach have become increasingly apparent. As best-practice risk management systems have continued to advance, bankers and their regulators have observed significant and systematic differences between the regulatory capital charges imposed by Basel I, and the economic capital charges generated by banks' internal models. As anticipated by Merton (1995) and explored in detail by Jones (2000), such discrepancies give banks strong economic incentives to shift lending or to engage in regulatory capital arbitrages in order to bring regulatory and economic capital requirements closer together. Indeed, early collateralized debt obligation (CDO) structures were designed expressly for this purpose.

The new Basel accord

The limitations of Basel I led regulators to begin work on broad revisions to the accord in the late 1990s. After an extensive development and public consultation process, the Basel Committee published revised capital adequacy standards in 2004 and updated rules for the treatment of banks' trading and risk mitigation activities in 2005. 'Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework', published by the Basel Committee in 2006, collects these new standards together in a single document. This 'Revised Framework', known as Basel II, sets forth detailed standards which need to be interpreted and implemented by national banking authorities. European and Japanese banks began transitioning to Basel II rules in January, 2007. Transition in the US began in 2008.

Basel II embodies a more comprehensive view of capital regulation than did Basel I. Whereas Basel I presents an unadorned rulebook for minimum capital standards, Basel II puts minimum capital standards in a broader context of supervisory and market discipline. The 'three pillars' of Basel II are intended to be mutually reinforcing. Pillar I establishes minimum risk-based capital requirements intended to cover the credit, trading, and operational risks faced by well-diversified financial institutions. Pillar II establishes guidelines for supervising banks' internal risk management processes and encourages regulators to require that banks hold capital buffers above Pillar I minimums to cover those economic risks not explicitly addressed under Pillar I. Pillar III imposes new public disclosure requirements on banks with an eye toward increasing transparency and facilitating more-effective market discipline of bank capital adequacy. Decamps,

Rochet, and Roger (2004) analyze the complementary roles of the three pillars in a theoretical model of bank behavior under moral hazard. They show how market signals can allow supervisors to apply a lighter hand to inspection and closure. The stringency of Pillar I requirements and the intrusiveness of Pillar II supervision depend on the quality of Pillar III market disclosures. For the remainder of this chapter, as in the great majority of the Basel II literature, we focus on Pillar I standards.

Like Basel I, Basel II requires that a bank maintain a risk-based capital ratio (RBCR) of at least 8 percent. While Basel II modestly updates rules for determining regulatory capital in the numerator of the RBCR, it dramatically changes the way risk weighted assets (RWA) in the denominator or the RBCR are calculated. Basel II relies on a much more detailed and rigorous approach to determining RWA than Basel I. Pillar I capital requirements are specifically designed to cover credit risk embedded in a bank's traditional lending portfolio (the banking book), market and credit risk associated with its trading activities, and operational risk arising from a failure of the bank's internal financial controls. Under Basel II, risk weighted assets are defined as

$$\mathrm{RWA} = \frac{1}{0.08}\left(s\sum_{i} k_i \mathrm{EAD}_i + K_{TR} + K_{OR}\right).$$

k_i is the capital requirement per currency unit associated with the bank's i-th credit exposure, and EAD_i is an estimate of the value of that exposure at default (computation of EAD is discussed in detail below). The Basel Committee (2006: para. 14) has indicated that it may adjust the credit risk capital requirements for banks using advanced approaches to compute credit risk capital charges by a constant scaling factor 's' in order to 'maintain the aggregate level of [overall minimum capital] requirements, while also providing incentives to adopt the more advanced risk-sensitive approaches of the Revised Framework'. K_{TR} and K_{OR} are capital charges that cover a bank's trading risks and operational risks, respectively. All capital charges are expressed as risk weighted assets by dividing through by 8 percent.

Basel II offers a menu of rulebooks of differing complexity and risk sensitivity for each of the main categories of risk covered under Pillar I. In general, it is expected that large internationally active banks in all Basel II jurisdictions will migrate toward the most sophisticated approaches on the menu. Less sophisticated approaches may be preferred for smaller institutions with more traditional bank portfolios. The range of options for such institutions helps to accommodate cross-country differences in approaches to bank supervision. For the more advanced menu options, capital charges are tied to value-at-risk (VAR), a measure of risk widely used by commercial and investment banks. VAR is defined as a specified

percentile of a portfolio loss distribution over a given assessment horizon. For example, if L is a random variable representing portfolio losses over the chosen horizon, then q^{th} percentile VAR is

$$\mathrm{VaR}_q[L] = \inf\{k|P[L \geq k] \leq 1 - q\}.$$

Capital requirements for the banking book are intended to cover the credit risk associated with debt instruments and related hedging instruments.[3] The Revised Framework proposes three separate options for calculating risk weights for positions in the banking book. The simplest approach, called the standardized approach, derives credit risk capital charges from broad loan-type categories that are similar, though more refined, to those used in Basel I. The most significant innovation in the standardized approach is that risk weights may be based on *external* rating, where a public rating is available from a recognized rating agency. The foundation and advanced IRB approaches go a step farther, and rely on banks' own, or *internal*, measures of the credit quality of individual loans or pools of retail loans to determine required capital. Under both IRB approaches, a loan's capital requirement depends on the bank's estimate of the likelihood that the loan will default. The main difference between the foundation and advanced IRB approaches is that in the latter a loan's risk weight also depends on its remaining maturity (except in the case of retail loans) and on the bank's estimate of the loss it expects to incur in the event that the loan defaults. For standardized and both IRB approaches, elaborately detailed rules are provided for recognition of risk mitigation in the form of financial collateral, third-party guarantees, and credit derivatives.

Basel II requires banks with significant trading activities to hold capital to cover potential losses in the market value of their trading positions. Rules for determining this capital requirement closely follow those established through amendments to Basel I in 1996 and 1997.[4] The trading book capital requirement has two components: a general market risk charge covering changes in the market value of a bank's entire trading portfolio arising from moves in market-wide risk drivers such as interest rates and equity valuations, and a specific risk charge for each trading position covering changes in the value of that position arising from

[3] Our reference to 'banking book' and 'trading book' positions is a somewhat imprecise shorthand notation, as certain risks in the trading book are treated according to banking book rules. For example, required capital against counterparty credit risk on derivative positions in the trading book is calculated by banking book rules as if it were a direct exposure to the counterparty.

[4] Relative to Basel I, Basel II places tighter restrictions on eligibility of a position for market risk treatment. The position not only must be marked to market in the trading book, but also must be held with intent to trade. As a practical matter, this may be a significant change from Basel I. Our discussion of the treatment of the 'trading book' should be taken to mean treatment of trading book risks subject to market risk rules.

idiosyncratic factors unrelated to broad market movements. Subject to supervisory approval, a bank may either use a simple standardized approach or a more advanced internal-models' approach to compute these charges. Under the standardized approach, simple risk weights are applied to a bank's trading positions. Under the internal models' approach, a bank uses its own risk management systems to compute a 99th percentile VAR of its trading portfolio's exposure to market risk over a ten-day horizon. The VAR measure for trading losses is multiplied by a factor of at least three to determine the bank's general-market risk-capital charge. A scaling factor of larger than three may be applied if the bank's VAR measure fails *ex post* performance tests (Basel Committee, 2006: para. 718). The Revised Framework provides formulas for computing specific risk charges for trading book positions, but it also allows banks to use internal models for this purpose.

Under Basel II, for the first time, a bank is also required to hold capital to cover the risk of defaults by the issuers of securities held in the trading book to the extent that this risk is not captured in its ten-day VAR calculations. According to the Basel Committee (2005a: para. 260–1) this incremental default risk charge was introduced to address new risks arising from the increasing prevalence of credit-risk-related products such as credit default swaps and less liquid structured credit products in banks' trading portfolios.

Basel II also introduces capital charges for operational risk, which the Basel Committee (2006: para. 644) defines as 'the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events'. Here, too, Basel II offers a menu of options. The simplest, called the basic indicator approach, is based on the intuition that operational risk derives from business activity, and so should increase in proportion with bank revenues. It sets a capital charge of 15 percent of the average gross income over the past three years.[5] The standardized approach is a refined version of the basic indicator approach in which bank activities are divided into eight broad business lines such as commercial banking, retail banking, asset management, and corporate finance. Capital charges that vary across these business lines (from 12 percent to 18 percent) are applied to average gross income by business line.

Under the most advanced option, called the advanced measurement approach (AMA), capital for operational risk is determined by the bank's own operational risk model. As a general principle, minimal structure is imposed by the Revised Framework on the AMA. The capital charge must correspond to 99.9 percentile VAR at a one-year horizon in the bank's internal operational-risk model, but the bank is given broad discretion to tailor model design to its own unique

[5] Gross income is net interest income plus non-interest income. Years with negative gross income are excluded from the average.

organizational structure, business environment, and internal controls. The flexibility of the AMA avoids stifling the development of what is still a nascent science. A potential cost to the approach, however, is that it will be difficult to discipline a bank that designs its AMA model opportunistically, as data constraints on statistical validation are severe and there is, as yet, no consensus among practitioners concerning what modeling assumptions are most appropriate. Supervisory oversight of the model is likely to be based instead on qualitative assessment and ongoing dialogue. Seivold, Leifer, and Ulman (2006) provide a primer on the regulatory treatment of operational risk under Basel II. One popular approach to modeling operational risk is drawn from extreme-value theory (EVT), a branch of statistics widely used in actuarial sciences that examines the properties of extreme-event data. Fontnouvelle, et al. (2006) estimate EVT models on two vendor-provided operational-loss data sets.

Theoretical foundations for the internal-ratings-based approaches

Credit risk in the banking book accounts for the dominant share of required capital under Basel I for the great majority of large commercial banks, and this will remain the case for the foreseeable future under Basel II. In terms of its impact on the practice of banking, both with respect to lending activity and regulatory arbitrage of the current Basel rules, the IRB approach is the most important innovation within Basel II. In this section, we review the seemingly contradictory goals that gave shape to the IRB approach, and the means by which the inherent contradictions have been addressed by the Basel Committee.

The primary motivation for Basel II was to achieve greater risk sensitivity in capital charges. This is desirable at the portfolio level, so that bank capital requirements are commensurate with bank portfolio risk. Banks with greater portfolio risk ought to face higher regulatory capital charges than banks with less risky portfolios, and, similarly, required capital for a bank should increase or decrease over time as the bank changes its risk profile. To reduce regulatory distortions in lending patterns and incentives for regulatory arbitrage, capital requirements should also be risk-sensitive at the *exposure*-level. That is, the marginal capital charge for a particular credit exposure should be broadly consistent with banks' assessments of the risk contribution of that exposure. Banks allocate economic capital for credit risk as contributions to portfolio *unexpected loss*, defined as the

difference between VAR and expected losses (the latter are assumed to be covered by interest income or reserves, and so excluded from required capital). Therefore Basel II capital charges would need to align risk weight formulas with those implied by bank credit VAR models.

The easiest way to align banks' internal risk management systems with regulatory capital requirements would be to allow banks to use their own internal models to calculate capital charges, much as they do for the trading book under the Market Risk Amendment of Basel I. A disadvantage of this internal-models' approach is that competitive pressures could induce banks to choose a model which delivers low capital charges over a more rigorous model that might demand higher capital. Given available data and supervisors' limited experience evaluating these models, it would be difficult for supervisors to impose much discipline on opportunistic behavior. Thus, although the Basel Committee intends to continue to examine the use of internal models for regulatory capital purposes, Basel II does not allow banks to use these models directly (Basel Committee, 2006: para. 18).

As an intermediate step between the broad category-based approach of Basel I and a full internal models' approach, the Basel Committee mandated that Basel II would offer an internal-ratings-based approach in which the risk weight for a banking book exposure would depend on the bank's assessment of the creditworthiness of the obligor and other risk characteristics of the instrument (e.g., collateral protection and maturity), but would not depend on the bank's assessment of how the exposure diversifies or concentrates risk in the context of its portfolio. That is, IRB risk weights would depend on the 'own' characteristics of each exposure, but not the characteristics of the portfolio in which the exposure is held. This property is termed *portfolio invariance*.

A challenge in the design of Basel II was to reconcile portfolio invariance with the desire to align risk weights with contributions to VAR in widely used models of portfolio credit risk. In general, an exposure's marginal contribution to portfolio VAR depends on the composition of the portfolio as a whole. Thus, two banks with different asset portfolios may well assign different marginal capital requirements to the same risk exposure. However, Gordy (2003) shows that contributions to VAR are portfolio-invariant if one assumes that

- the portfolio is *asymptotically fine-grained*, in the sense that no single obligor accounts for more than an infinitesimal share of portfolio exposure;
- a single risk factor is the sole source of systematic risk in the portfolio; and
- realizations of the systematic risk factor are monotonically related with the conditional expected losses associated with most risk exposures.

Under these assumptions, portfolio VAR is equal to the sum across exposures of the expected loss for that exposure conditional on a particular adverse draw of the systematic risk factor.

Letting X denote the systematic factor and x_q denote the q^{th} percentile of the distribution of X, and letting ℓ_i denote the loss on the i^{th} exposure, we have

$$VaR_q[L] = \sum_i E[\ell_i | X = x_q].$$

This linear expression implies that contributions to VAR (both average and marginal) can be computed independently across exposures (i.e., that portfolio invariance is satisfied). Expected loss (EL) for the portfolio is similarly a linear aggregation of exposure EL, and so contributions to unexpected loss (UL) are portfolio invariant as well. This asymptotic-single-risk factor (ASRF) framework serves as the theoretical foundation for IRB risk weights.

For simplicity, consider the treatment of a one-year bullet loan to a commercial enterprise. Let D be an indicator variable that is equal to one if an obligor defaults over a one-year horizon and zero otherwise, and let R be a random variable between zero and one that describes the proportion of an outstanding credit exposure to the obligor that will be recovered in the event of default. In this set-up, R is a latent variable; it exists in all states of the world, but it can only be observed in those states in which the obligor defaults. The default-related loss is $\ell = D \cdot (1-R) \cdot EAD$. Under the ASRF framework, D and R may depend on a systematic risk factor X, but, conditional on X, defaults and recoveries are assumed to be independent across obligors (EAD is here assumed to be non-random for simplicity). The marginal contribution of this position (per currency unit of exposure) to portfolio 99.9 percentile UL is given by the difference between its conditional and its unconditional expected loss:

$$K = E\lfloor D \cdot (1 - R) | X = x_q \rfloor - E[D \cdot (1 - R)]. \quad (1)$$

The conditional expected loss term in equation (1) can be expressed as the product of a term measuring the probability of default under systematic stress conditions

$$SPD = E\lfloor D | X = x_q \rfloor \quad (1a)$$

and a term measuring loss-given-default under the same stress conditions

$$LGD = E\lfloor (1 - R) | D = 1, X = x_q \rfloor. \quad (1b)$$

This decomposition is compatible with results from surveys by the Basel Committee (2000) and by Treacy and Carey (2000), which find that banks commonly evaluate credit exposures using two-dimensional rating systems that separately account for an obligor's likelihood of default and a loan's loss rate, should default

occur. The particular models used to derive stress probability of default (SPD) and loss given default (LGD), and the calibration of embedded parameters is described below.

The actuarial perspective of equation (1) does not capture the full credit risk associated with exposures of maturity beyond the model's one-year horizon. Any credit migration short of default that is incurred within the year will imply a gain or loss on the exposure's market value. The longer the loan's remaining maturity, the greater the sensitivity of its market value to a rating change. In general, higher-quality loans are less likely to experience default-related losses over a one-year horizon but they are more likely to lose market value because of a ratings downgrade short of default. The Revised Framework includes a maturity adjustment function that rescales capital charges for corporate, bank, and sovereign credit exposures to reflect the effects of credit-related changes in market value on an exposure's marginal contribution to UL. The IRB maturity adjustment is derived within a generalized mark-to-market version of this model that accounts for migration risk. While the particular form of the IRB maturity adjustment is a piecewise-linear approximation to the model-implied relationships, Gordy (2003) emphasizes that the principles of the ASRF framework apply in a mark-to-market setting in exactly the same fashion as in an actuarial setting.

Equation (1) describes the ASRF capital charge for a whole loan, but the logic of the ASRF framework applies equally to any sort of credit exposure in a bank portfolio. Heitfield (2003) shows that the ASRF framework can also be used to derive capital requirements for loans that include a third-party credit guarantee or bonds whose default risk is hedged with a derivative instrument such as a credit default swap. Pykhtin and Dev (2002) and Gordy and Jones (2003) use the ASRF framework to derive capital requirements for structured finance products in which credit performance of a security depends on the performance of an underlying pool of assets. These models are the basis of the IRB treatment of securitization exposures.

The three assumptions of the ASRF framework enumerated above are not inconsequential. Pillar II requires that banks and their supervisors consider ways in which these assumptions might be violated, and, if necessary, hold additional capital beyond that implied by the Pillar I risk weight formulas.

The assumption of monotonicity of the conditional expected loss rate (Assumption 3) may be violated for hedging instruments and certain asset-backed securities. However, this assumption need apply only for the portfolio as a whole and not for each position, and furthermore can be significantly relaxed.

For the largest banks, characterization of the portfolio as asymptotically fine-grained (Assumption 1) may be a reasonable approximation. To the extent that the IRB approach is applied to less well-diversified institutions, there will be a residual of undiversified idiosyncratic risk in the portfolio that is ignored by the IRB approach, and so regulatory capital requirements may understate economic capital requirements. Analytic and semi-analytic approximations developed by Wilde

(2001) and Martin and Wilde (2002) can be used to measure the effect of name concentration on capital requirements. Gordy (2004) reviews the mathematical foundations for granularity adjustment, and Gordy and Lütkebohmert (2007) develop simple algorithms for application of the granularity adjustment in the IRB context.

The 'single factor' assumption (Assumption 2) is the more serious limitation of the ASRF framework. Empirical research by Zeng and Zhang (2001), Carling, Ronnegard, and Roszbach (2004), Heitfield, Burton, and Chomsisengphet (2006), and McNeil and Wendin (2006) find that credit losses associated with exposures to obligors in the same industry sectors are more highly correlated with one another than those associated with exposures to obligors in different sectors. So long as credit conditions across countries and industries do not move together in lockstep, diversification in the portfolio will depend not only on name concentration (i.e., granularity), but also on diversification across sectors. While this assumption limits the validity of *any* ratings-based method for assessing capital charges, Pykhtin (2004) shows how analytic methods can be used to adjust ASRF capital requirements for sector concentration effects and Garcia Cespedes, et al. (2006) describe a practical approach to measuring the impact of sectoral diversification in a ratings-based capital framework.

The IRB risk weight formulas

The IRB capital formula for exposures of one-year maturity is derived from a simple one-factor Merton (1974) model similar to that underlying the popular CreditMetrics model (Gupton, Finger, and Bhatia, 1997).

Let Y be a random variable describing the return on asset value of the obligor over the model horizon. Under the simplest Merton-style default model, the obligor is assumed to default if Y falls below some critical default threshold γ. Thus, if we assume that Y has a standard normal distribution, the probability of default (PD) for the obligor is given by

$$\mathrm{PD} = \mathrm{E}[\mathrm{D}] = \Phi(\gamma)$$

where $\Phi(\mathrm{x})$ is the standard normal cumulative distribution function (CDF). Since the normal CDF has a well-defined inverse function, if the obligor's PD is known γ can be calculated using the transformation $\gamma = \Phi^{-1}$ (PD).

To account for the influence of systematic risk, which gives rise to correlations in defaults across obligors, one needs to decompose Y into common and obligor-specific components. Let

$$\mathrm{Y} = \mathrm{U}\sqrt{1-\rho} - \mathrm{X}\sqrt{\rho}$$

where X is the systematic risk factor that affects all obligors and U is an idiosyncratic risk factor that is unique to the obligor in question and is uncorrelated with X. X and U are both assumed to be standard normal random variables. The asset correlation parameter ρ measures the relative importance of systematic risk in determining the obligor's likelihood of default. A value of ρ close to one implies that default by the obligor is determined primarily by the systematic risk factor X, while a value of ρ close to zero implies that defaults are largely independent across obligors.

Since X has a standard normal distribution, $x_q \equiv -\Phi^{-1}(0.999)$. Thus, we can write

$$\mathrm{SPD} = \Phi\left(\frac{\Phi^{-1}(\mathrm{PD}) + \sqrt{\rho}\cdot\Phi^{-1}(0.999)}{\sqrt{1-\rho}}\right). \tag{2}$$

Equation (2) provides a formula for computing an obligor's conditional probability of default given x_q from a bank's estimate of its PD.

Equation (1b) implies that the LGD parameter in the ASRF gross capital charge should be consistent with the same systematic stress conditions ($X = x_q$) as applied to the PD component. To the extent that loss rates on defaulted loans tend to be higher during times when default rates are also high, capital requirements need to capture this correlation. Evidence on this point is provided by Frye (2000; 2003), Altman, et al. (2005), and Acharya, Bharath, and Srinivasan (2007), who find that loss rates on defaulted corporate debt are elevated during times of industry or economy-wide stress, and Qi and Yang (2008), who finds that mortgage loss severities during distressed housing markets are significantly higher than under normal housing market conditions.

Pykhtin (2003) and Düllmann and Trapp (2004) propose parametric models that extend the one-factor Merton default model described in the previous section to incorporate correlation between the systematic risk factor and defaulted loan loss rates. In principle, a model could be used to derive a functional relationship between a bank's unconditional expectation of default-related losses to a stressed LGD parameter consistent with equation (1b). However, since models of systematic recovery risk are not widely used in practice, the Revised Framework does not require that banks' LGD estimates explicitly be tied to an adverse draw of the systematic risk factor. Instead, advanced IRB banks must report LGD estimates that 'reflect economic downturn conditions where necessary to capture the relevant risks' (Basel Committee, 2006: para. 468). This qualitative requirement and clarifying guidance issued by the Basel Committee (2005b) gives banks a great deal of flexibility in determining how to incorporate the effects of systematic risk in their LGD estimates.

Plugging equation (2) into equation (1a) shows that, all else being equal, the higher an exposure's asset correlation parameter ρ, the greater will be its required capital. To understand why, consider two credit exposures that each have the same PD but different asset correlations. If a portfolio is well diversified, only adverse systematic shocks that do not average out across exposures can lead to higher than expected portfolio credit losses. Thus, the exposure with a lower asset correlation should contribute less to UL because a larger proportion of the uncertainty associated with the credit performance of that exposure can be diversified away. Given the importance of asset correlation parameters in determining risk-based capital requirements, it is not surprising that a large and growing body of research has sought to quantify these parameters.

The most direct approaches to estimating asset correlation parameters rely on historical data, on firm asset valuations, or the credit performance of debt instruments. Early research simply used historical correlations in equity returns for public companies as proxies for asset correlation. A shortcoming of this approach is that equity values do not map directly to asset values, since, for example, different firms have different capital structures. More recently, research by Heitfield, Burton, and Chomsisengphet (2005), Düllmann, Scheicher, and Schmieder (2006), and others have estimated asset correlation parameters using North American and European firm asset values imputed by Moodys KMV from equity valuations and leverage information. This approach is simple and direct, but Zhu, et al. (2007) show that estimated correlations may be sensitive to measurement errors in imputed asset values. Gordy (2000), Hamerle, Liebig, and Rösch (2003), and others estimate asset correlation parameters from data on the credit performance of bonds or other debt instruments by exploiting the fact that, all else being equal, more volatile observed default rates imply higher underlying asset correlation parameters. This approach does not rely on imputed firm asset values, but, as Frey and McNeil (2003) point out, results are sensitive to the functional form assumptions embedded in the single-factor Merton model. In general, studies that rely on imputed firm asset values tend to find higher asset correlations than those that employ historical default data.

A second strand of research infers average asset correlations for the simple single-factor Merton model from a more sophisticated multi-factor portfolio risk model. Under this approach, one fixes a benchmark portfolio, and then solves for a single-factor Merton asset correlation parameters that equate VAR across the two models. Lopez (2004) uses this method to infer asset correlation parameters from VAR estimates produced by the Moody's KMV Portfolio Manager model. Instead of just one systematic factor, Portfolio Manager includes over 100 factors that capture country- and industry-specific shocks. Lopez finds that for a portfolio of loans within a single country, the single-factor Merton model with asset

correlations between 0.14 and 0.26 produce results similar to those generated by Portfolio Manager, and that portfolios of loans with higher default probabilities tend to have lower implied asset correlations. Calem and Follain (2003) infer asset correlations from mortgage default models developed by the Federal Reserve Board, the Mortgage Insurance Corporation of America, and Fitch Ratings. Depending on the model used and other details of the specification, they find mortgage-asset correlation parameters ranging from 0.12 to 0.25 with estimates centered near 0.15.

The asset correlation parameters embedded in Basel II vary across type and size of borrower (e.g., lower for retail and SME borrowers than for corporate borrowers). In setting these parameters, the Basel Committee made use of information from a variety of sources, including empirical research of the sort described above and comments received through the Committee's consultations with banks, rating agencies, and other market participants.

Bank-supplied inputs to the IRB risk weight formula

Under both the foundation and advanced IRB approaches, the risk weight for a banking book exposure is a function of four bank-supplied parameters: the PD, downturn LGD, expected exposure at default (EAD), and, in some cases, the expected maturity (M).

PD is the parameter that has received the most attention. A prerequisite for IRB adoption is that a bank have an internal rating system that assigns a credit rating to each corporate obligor commensurate with that obligor's 'ability and willingness to contractually perform despite adverse economic conditions or the occurrences of unexpected events' (Basel Committee, 2006: para. 415). The PD associated with the obligor is a function of the credit rating grade assigned to that obligor. The rules require that the PD associated with a rating grade be 'a long-run average of one-year default rates for borrowers in the grade' (Basel Committee, 2006: para. 447). Although the IRB rules are fairly specific about how PDs should be computed, they give banks substantial latitude in determining how ratings are assigned.

In practice, approaches to slotting obligors into rating grades may differ widely across banks. For example, Treacy and Carey (2000) find that some banks rapidly update obligor ratings as business conditions change, while others use so-called 'through-the-cycle' rating systems designed to produce ratings that remain stable

even as aggregate business conditions change. Because an obligor's PD depends on the rating it is assigned, there is no guarantee that the PD associated with an obligor will be consistent across banks with different rating systems. Using samples of obligors rated by multiple banks in the US and Sweden, respectively, Carey (2002) and Jacobson, Linde, and Roszbach (2006) find frequent cross-bank differences in the default probabilities associated with banks' ratings for the same obligors. Rösch (2005) and Heitfield (2006) show that differences in banks' rating philosophies can affect the level and volatility of portfolio-wide regulatory capital requirements.

For banks on the foundation IRB approach, LGD parameters are specified by the Revised Framework. For example, senior, uncollateralized loans to corporations or government entities are assigned LGDs of 45 percent, while subordinated, uncollateralized loans are assigned LGDs of 75 percent. If a loan is secured, the LGD may be reduced by an amount tied to the value and quality of the collateral. Banks on the advanced IRB approach are required to estimate LGD parameters using information from their credit rating systems. To ensure that these LGD parameters reflect recoveries that may be depressed during downturn conditions, the Basel Committee (2005) guides banks to have rigorous and well-documented systems for identifying appropriate downturn conditions, identifying any adverse dependencies between default rates and recovery rates, and incorporating dependencies so as to produce LGD parameters consistent with downturn conditions.

EAD is the bank's expected legal claim on the borrower in the event of default. For bonds and term loans, EAD is the loan's face value plus expected accrued but unpaid interest. For undrawn commitments, EAD is more difficult to estimate. For revolving lines, banks are expected to specify EAD as the current drawn balance plus a 'credit conversion factor' (CCF) applied to the remaining undrawn balance. Under the foundation IRB approach, the CCF is given in the Revised Framework and depends on the type of credit facility. Advanced IRB banks are permitted to use their own estimates of CCFs. Banks must consider their ability and willingness to prevent drawdowns of unused commitments in the event of borrower distress. Treatment must reflect not only legal enforceability, but also the bank's systems and procedures for monitoring drawdowns. The CCFs used in practice are often based, at least in part, on the estimates of Asarnow and Marker (1995).

The maturity parameter, M, is calculated as a cashflow-weighted duration, and is subject to a floor of one year and a ceiling of five years.[6] For banks on the foundation IRB approach, maturity is fixed at 2.5 years for corporate, bank, and sovereign exposures. Retail credit exposures, such as mortgages and credit

[6] More precisely, M is a Macauley duration calculated with a discount rate $r = 0$. The one-year floor does not apply to certain short-term exposures such as repo agreements.

cards, do not include an explicit maturity adjustment. For these exposures, average maturity effects are implicit in the calibration of the IRB risk weight formulas.

Looking forward

In the approach to adoption of Basel II, much has been written on the challenges of implementation. For regulators, the validation of rating systems and PD and LGD quantification (particularly for investment-grade obligors) may be the most persistent challenge in Basel II implementation. For some banks, the required investment in information systems may be substantial. In many cases, this reflects deficiencies in bank information systems that should be addressed for internal risk management purposes, and therefore does not necessarily constitute a direct cost of regulatory compliance. Indeed, wherever a bank-supplied input is based on data that the bank does not itself find useful for internal management, the data should be viewed as inherently suspect. The theme of the 'use test' runs throughout the Revised Framework. Nonetheless, in any regulatory system as complex as Basel II, it is inevitable that some requirements will not mesh naturally with internal bank processes, and thereby give rise to compliance costs.

Another topic that has received considerable attention is the potential procyclical effects of Basel II. So long as bank rating systems are responsive to changes in borrower default risk, capital requirements under the IRB approach will tend to increase as an economy falls into recession and fall as an economy enters an expansion. To the extent that banks curtail or expand lending in response, Basel II could make it more difficult for policymakers to maintain macroeconomic stability. Daníelsson et al. (2001) elaborate on this critique (among others) of Basel II. Kashyap and Stein (2004) and Gordy and Howells (2006) suggest modifications to Basel II rules that would mitigate procyclicality without sacrificing the risk sensitivity of capital requirements or the quality of information in Pillar III market disclosures.

Basel II is designed to evolve over time along with financial innovation and improvements in best-practice risk management. At least in principle, the model foundation for the IRB approach permits capital treatment of new instruments to be derived in a manner consistent with that of existing instruments. The IRB framework itself can evolve to some degree. For example, as data improve, the simple supervisory rules for asset correlation could be replaced with bank-supplied inputs. However, the stylized nature of the IRB model does impose certain limits. Integration of interest rate risk with the treatment of credit risk in the banking book would probably be very difficult. The single-factor assumption drastically

limits our ability to recognize sectoral concentration or to reward international diversification, yet is deeply embedded in the IRB framework and, indeed, is unavoidable in *any* ratings-based capital rule. When these limitations can be exploited for regulatory capital arbitrage—a prospect not currently in sight, but perhaps inevitable in the long run—the stage will be set for transition to a full internal-models' approach to regulatory capital.

References

Acharya, V. V., S. T. Bharath, and A. Srinivasan (2007). 'Does Industry-Wide Distress Affect Defaulted Firms?: Evidence from Creditor Recoveries'. *Journal of Financial Economics*, 85: 787–821.

Altman, E. I., B. Brady, A. Resti, A. Sironi (2005), 'The Link Between Default and Recovery Rates: Theory, Empirical Evidence, and Implications'. *Journal of Business* 78: 2203–28.

Asarnow, E. and J. Marker (1995). 'Historical Performance of the U.S. Corporate Loan Market: 1988–1993'. *Commercial Lending Review*, 10: 13–32.

Basel Committee on Banking Supervision (1988). *International Convergence of Capital Measurement and Capital Standards*. Bank for International Settlements.

——(1996). *Amendment to the Capital Accord to Incorporate Market Risks*. Bank for International Settlements.

——(2000). *Range of Practices in Banks' Internal Rating Systems*. Bank for International Settlements.

——(2005a) *The Application of Basel II to Trading Activities and the Treatment of Double Default Effects*. Bank for International Settlements.

——(2005b). *Guidance on Paragraph 468 of the Framework Document*. Bank for International Settlements.

——(2006). *International Convergence of Capital Measurement and Capital Standards. A Revised Framework, Comprehensive Version*. BCBS Publication 128, Bank for International Settlements.

Calem, P. and J. Follain (2003). 'The Asset-Correlation Parameter in Basel II for Mortgages on Single-Family Residences'. White paper. Board of Governors of the Federal Reserve System.

Carey, M. (2002). 'Some Evidence on the Consistency of Banks' Internal Credit Ratings', in *Credit Ratings: Methodologies, Rationale, and Default Risk*, Michael Ong (ed.), London: Risk Books.

Carling, K., L. Ronnegard, and K. Roszbach (2004). 'Is Firm Interdependence within Industries Important for Portfolio Credit Risk?', Sveriges Riksbank Working Paper 168.

Daníelsson, J., P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, and H. S. Shin (2001). 'An Academic Response to Basel II'. Special Paper 130. London School of Economics Financial Markets Group.

Decamps, J.-P., J.-C. Rochet, and B. Roger (2004). 'The Three Pillars of Basel II: Optimizing the Mix. *Journal of Financial Intermediation*, 13: 132–55.

Düllmann, K., M. Scheicher, and C. Schmieder (2006). 'Asset Correlations and Credit Portfolio Risk: An Empirical Analysis'. Discussion Paper Series 2: Banking and Financial Studies 2007, 13, Deutsche Bundesbank Research Centre.

—— and H. Scheule (2003). 'Determinants of the Asset Correlations of German Corporations and Implications for Regulatory Capital', Working Paper.

—— and M. Trapp (2004). 'Systematic Risk in Recovery Rates: An Empirical Analysis of US Corporate Credit Exposures. Discussion Paper Series 2: Banking and Financial Studies 2004, 02, Deutsche Bundesbank Research Centre.

Fontnouvelle, P. de, V. DeJesus-Rueff, J. S. Jordan, and E. S. Rosengren (2006). 'Capital and Risk: New Evidence on Implications of Large Operational Losses'. *Journal of Money, Credit, and Banking*, 38: 1819–46.

Frey, R. and A. McNeil (2003). 'Dependent Defaults in Models of Portfolio Credit Risk'. *Journal of Risk*, 6: 59–92.

Frye, J. (2000). 'Collateral Damage'. *Risk*, April: 91–4.

—— (2003). 'A False Sense of Security'. *Risk*, August: 63–7.

Garcia Cespedes, J. C., J. A. de Juan Herrero, A. Kreinin, and D. Rosen (2006). 'A Simple Multifactor "Factor Adjustment" for the Treatment of Credit Capital Diversification'. *Journal of Credit Risk*, 2: 57–86.

Gordy, M. B. (2000). 'A Comparative Anatomy of Credit Risk Models'. *Journal of Banking & Finance*, 24: 119–49.

—— (2003). 'A Risk Factor Model Foundation for Ratings-Based Bank Capital Rules'. *Journal of Financial Intermediation*, 12: 199–232.

—— (2004). 'Granularity Adjustment in Portfolio Credit Risk Measurement', in *Risk Measures for the 21st Century*. G. P. Szegö (ed.). Hoboken, NJ: John Wiley & Sons.

—— and B. Howells (2006). 'Procyclicality in Basel II: Can We Treat the Disease without Killing the Patient?'. *Journal of Financial Intermediation*, 15: 395–417.

—— and E. Lütkebohmert (2007). 'Granularity Adjustment for Basel II. Deutsch Bundesbank Discussion Paper Series 2, 01/2007.

—— and D. Jones (2003). 'Random Tranches'. *Risk*, March: 78–83.

Gupton, G., C. C. Finger, and M. Bhatia (1997). *CreditMetrics: Technical Document*. New York: J.P. Morgan & Co.

Hamerle A., T. Liebig, and D. Rösch (2003). 'Benchmarking Asset Correlations', *Risk*, 16: 77–81.

Heitfield, E. (2003). 'Using Credit Derivatives and Guarantees to Reduce Credit Risk Capital Requirements under the New Basel Capital Accord', in *Credit Derivatives: The Definitive Guide*. J. Gregory (ed.). London: Risk Books.

—— (2006). 'The Use of Internal Rating Systems in Credit Risk Managements'. Working Paper.

—— S. Burton, and S. Chomsisengphet (2005). 'Risk-Sensitive Regulatory Capital Rules for Hedged Credit Exposures', in *Credit Risk Modeling: Pricing, Risk Management and Regulation*. M. Pykhtin (ed.). London: Risk Books.

—— —— —— (2006), 'Systematic and Idiosyncratic Risk in Syndicated Loan Portfolios'. *Journal of Credit Risk*, 2: 3–31.

Jacobson, T., J. Linde, and K. Roszbach (2006). 'Internal Ratings Systems, Implied Credit Risk and the Consistency of Banks' Risk Classification Policies'. *Journal of Banking & Finance*, 30: 1899–926.

Jones, D. (2000). 'Emerging Problems with the Basel Capital Accord: Regulatory Capital Arbitrage and Related Issues'. *Journal of Banking & Finance*, 24: 35–58.

Kashyap, A. K. and J. C. Stein (2004). 'Cyclical Implications of the Basel-II Capital Standards'. *Federal Reserve Bank of Chicago Economic Perspecives*. First Quarter: 18–31.

Lopez, J. (2004). 'The Empirical Relationship between Average Asset Correlation, Firm Probability of Default, and Asset Size'. *Journal of Financial Intermediation*, 13: 265–83.

Martin, R. and T. Wilde (2002). 'Unsystematic Credit Risk'. *Risk*, 15: 123–8.

McNeil, A. and J. Wendin (2006). 'Dependent Credit Migrations'. *Journal of Credit Risk*, 2: 87–114.

Merton, R. C. (1974). 'On the Pricing of Corporate Debt: The Risk Structure of Interest Rates'. *Journal of Finance*, 29: 449–70.

—— (1995). 'Financial Innovation and the Management and Regulation of Financial Institutions'. *Journal of Banking & Finance*, 19: 461–81.

Pykhtin, M. (2003). 'Unexpected Recovery Risk'. *Risk*, August: 74–8.

—— (2004). 'Multi-Factor Adjustment'. *Risk*, 17: 88–90.

—— and A. Dev (2002). 'Credit Risk in Asset Securitizations: An Analytical model'. *Risk*, May: S16–S20.

Qi, M. and X. Yang (2009). 'Loss Given Default of High Loan-to-Value Residential Mortgages'. *Journal of Banking & Finance*, 33: 788–99.

Rösch, D. (2005). 'An Empirical Comparison of Default Risk Forecasts from Alternative Credit Rating Philosophies'. *International Journal of Forecasting*, 21: 37–51.

Seivold, A., S. Leifer, and S. Ulman (2006). 'Operational Risk Management: An Evolving Discipline'. *Supervisory Insights*, 3: 4–11.

Treacy, W. and M. Carey (2000). 'Credit Risk Ratings at Large U.S. banks'. *Journal of Banking & Finance*, 24: 167–201.

Vasicek, O. (1991). 'Limiting Loan Loss Probability Distributions'. White Paper. KMV.

Wilde, T. (2001). 'Probing Granularity'. *Risk*, 14: 103–6.

Zeng, B. and J. Zhang (2001). 'An Empirical Assessment of Asset Correlation Models'. Moody's/KMV White Paper.

Zhu, F., B. Dvorak, A. Levy, and J. Zhang (2007). 'Using Asset Values and Asset Returns for Estimating Correlations'. Moody's/KMV White Paper.

CHAPTER 15

MARKET DISCIPLINE IN BANK SUPERVISION

MARK J. FLANNERY[1]

Introduction

Most companies in capitalist countries are subject to 'market discipline'. Profits are maximized by producing the most valued products at the lowest attainable cost. Market forces determine input and output prices, and investors determine a firm's access to new capital on the basis of the likely return on investments. Capitalist governments generally remain indifferent to what firms produce, how much they pay for their inputs, or how much they lever their equity. In contrast, banking firms are subject to government restraints beyond those imposed by the product and capital markets. Stable banking firms are considered important for the public welfare. In addition, they are often considered unusually 'opaque' firms—difficult for outsiders to understand or control. Financial supervisors therefore assert that a

[1] I would like to thank the Markit Group for providing the data underlying some of the figures, and the editors for extremely helpful comments on a prior draft of this chapter. Any errors or omissions remain my responsibility.

bank's true value can only be properly understood by gathering confidential, 'inside' information about the firm's operations.

Banking firms are somewhat unusual because many banking services require that the customer also become a creditor of the bank (Merton, 1995). Small depositors probably find it most difficult to assess a bank's creditworthiness. Many governments offer explicit (or implicit) deposit insurance to relieve unsophisticated investors of that burden. Such government guarantees probably exacerbate the challenge of assuring bank stability. Insured banks' funding costs do not fully reflect their risk exposures, while riskier investments tend to offer higher returns. Government-insured banks thus have an artificial incentive to increase risk ('moral hazard'). Government supervision seeks to replace (or strengthen) this lost market discipline by controlling bank risk via 'prudential' regulation (Flannery, 1982).

Financial institutions have grown increasingly complex and new financial instruments permit an institution to change its risk exposures rapidly. As a consequence, supervisors' historical methods of identifying and controlling bank risks have become less effective and the value of timely, forward-looking information has increased. For most types of firms, financial economists believe that investor preferences do influence firm behavior. Some academics and regulators therefore suggest that market forces could supplement government efforts to control bank risk taking. In this context, 'market discipline' entails two components (Bliss and Flannery, 2002). First, investors must identify the condition of banking firms in a timely and accurate manner. If outside bank stakeholders accurately *monitor* their bank's condition, security prices and the bank's liability choices will reflect that market information. The second component of market discipline requires that investors' reactions to a bank's credit developments must *influence* how the firm behaves. Surprisingly, little academic research supports this hypothesis.

Even if market discipline works for most firms in a textbook fashion, it remains an open question whether outside stakeholders can *directly* control bank risks. However, the structure of bank supervision creates a second, *indirect* path of influence if supervisors use market information to guide their policies toward individual firms (Kwast, 1999). Some writers simply argue that market prices provide a valuable source of information, which should be incorporated into bank supervision. Others have proposed formal rules linking supervisory actions to market conditions. For example, if a bank's subordinated debenture yields exceed some threshold, supervisors would be required to constrain the bank's activities (Evanoff and Wall, 2002). Some proposals link debenture yields to the prompt corrective action (PCA) steps created in FDICIA.[2]

[2] Appendix A in Kwast (1999) provides an overview of fourteen specific proposals dating back to 1981.

This chapter describes the research suggesting that bank securities' prices (including stock, bonds, and CDSs) can be used to improve government supervisory processes. The next section defines the concept of market discipline and explains the importance of its two main components: market *monitoring* of firm value and market *influence* over a firm's behavior. The third section explains how the view of banking firms as unusually 'opaque' may create a need for government regulation, and discusses whether banks are truly opaque. The evidence on investors' ability to monitor bank condition accurately is described in the fourth section, which concludes that bond yields do reflect bank risks, and that depositors withdraw funds when a bank's riskiness increases. In short, the evidence indicates that the market can monitor banking firms quite well. The fifth section discusses some of the practical issues related to the use of market information and proposes the best way to incorporate market assessments into the supervisory process. The sixth section evaluates the concept of market discipline in light of the financial crisis that began in August 2007. The final section summarizes and concludes.

'Market discipline' defined

The ability of outsiders to monitor a firm's condition is the required first stage of market discipline. Financial markets are generally considered efficient, in the sense that investors collectively set prices that accurately reflect a firm's true condition. For non-financial firms, most financial economists believe that equity and debt prices lead firm managers to make optimal resource allocation decisions. This belief is so ingrained that it is rarely mentioned, much less tested—unless the firm in question is a bank (see below).

If 'monitoring' is accurate and timely, market discipline then requires investors to 'influence' firms' (managerial) behavior. Such influence can occur either *ex ante* or *ex post*. *Ex ante*, a bank considering a new line of business would understand that investors will charge more for funding if the bank's portfolio risk rises (or vice versa for a risk decrease). In considering whether to enter a new line of business, therefore, managers would incorporate changed financing costs into their assessment of the new project's costs and benefits. Now, consider *ex post* discipline. Imagine that a firm has received a negative shock—for example, a large increase in loan losses. The losses deplete capital and hence reduce the value of outstanding bonds and equity. New bonds can be sold only with a higher coupon rate. Old bondholders hope that those increased costs will convince the bank to reduce its default risk by replacing the lost equity capital. *Direct* influence would occur if higher interest expenses lead managers to recapitalize. Direct influence might also

	Monitoring	Influence	
		Direct	Indirect
'Regular' firms	Stock and bond prices are good estimates of a firm's true value and risk exposures.	Actual or anticipated changes in security prices signal appropriate actions to managers.	No such concept
Banking firms	Same as above	Same as above	Supervisors use security price changes to identify banks that might require oversight or corrective action.

Fig. 15.1. Aspects of market discipline

occur through a bond covenant mandating a minimum capital ratio. A capital-deficient bank would be forced to issue new shares or shrink its asset portfolio—either of which would presumably reduce default risk.[3]

Even if banks prove insensitive to direct market influence, government supervision provides a source of *indirect* influence. Effective market monitoring permits supervisors to extract forward-looking information from the bank's stock and bond prices, CDS premiums, etc. If supervisors use security prices to identify banks that might need supervisory attention, those prices have indirect influence (Kwast, 1999). Some writers have suggested formal rules linking supervisory actions to market prices. Figure 15.1 summarizes the various aspects of 'market discipline', to which I return below.

Market monitoring: The information in bank security prices

To some extent, the banks' special regulatory oversight reflects a belief that outside investors find it unusually difficult to identify bank asset values and risk exposures. If banks are unusually 'opaque', their debt and equity prices will not accurately

[3] Goyal (2005) argues that deregulation made US banking potentially more risky in the early 1980s, and shows that investors included more restrictive covenants in bank debentures following deregulation. This is a good example of direct influence at work. However, the US implementation of the

reflect how managerial decisions and external events affect the bank's value and riskiness.

Are banks unusually opaque?

Banks historically specialized in purchasing and financing the debt of firms that were unable to borrow in public bond markets. If we think that a bank's value reflects primarily its assets in place—the current stock of specific loans—opacity does appear potentially important. If a customer's creditworthiness were readily understood by outsiders, why would she choose to borrow from a bank? Large banks recently began to sell many of the loans they originate, which makes it even more difficult to know which loans are still on the books. This aspect of bank valuation is clearly illustrated by the 2007–9 'subprime crisis', during which banks' holdings of complex securities made it difficult for many institutions to borrow at usual risk spreads.

Although this view contains an element of truth, it seems quite different from the way we value other firms. Consider an automobile manufacturer. Would analysts value the company by calculating the market value of its assets in place (metal stamping machines, engine plants, etc.)? Or would they predict the company's future cash flows from using these assets? Similarly, banking firms should be evaluated on the basis of the profits they can generate using their underwriting and trading skills, which the current assets reflect incompletely. If banking firms' underwriting standards and information systems can be assessed accurately by outsiders, their market value should not depend so completely on the valuation of their opaque assets in place at any one time.

The greater emphasis on banks' assets in place may reflect their ability to change the nature of their risk exposures. While an auto company always makes autos, a bank might hold primarily commercial loans one year and subprime mortgage loans the next. Risk exposures can also be shifted rapidly via off-book derivatives and market-making activities. Commercial banks share this characteristic with investment banks, insurance companies, and hedge funds.[4]

The evidence on banking firms' relative opacity takes two forms, whose conclusions largely contradict one another. Flannery, Kwan, and Nimalendran (2004) compare the trading properties of bank stock against those of non-bank firms. A stock's bid-ask spread should be positively related to that amount of private

Basel I Accord weakened this source of direct influence in 1988: debentures were excluded from Tier 2 capital if they included provisions to accelerate bond repayment or to vary the bond's coupon rate with changes in bank condition.

[4] A small number of very large US banks operate proprietary trading desks, which invest like hedge funds. Hildebrand (2007) reports that the largest five (investment and commercial) bank proprietary trading desks controlled more assets than the entire hedge fund industry at the end of 2005.

information 'informed' investors might have about the stock's value (Kyle, 1986). Flannery, Kwan, and Nimalendran (2004) reasoned that relatively opaque firms' stock would trade with higher bid-ask spreads because outside investors were wary of trading with investors who might have important, private information about the firm's value. They find that the largest banks' equity (traded on the NYSE) had very similar bid-ask spreads to a set of equity-size-matched non-financial firms. Smaller banks (traded on the NASDAQ) appeared *easier* for the typical investor to understand than their non-financial matched firms. In neither subsample did the evidence indicate that banks were more opaque.

Morgan (2002) investigates relative opacity by comparing the ratings assigned by Moody's and Standard & Poor's to newly issued bonds during the period 1983–93. He proposes that split ratings imply greater opacity, because split-rated firms were viewed differently by two specialized analysts (the bond raters). Morgan compares split ratings for firms from various US industries and finds that banking and insurance firms' bonds are substantially more likely to carry split ratings. Among the banks, those with relatively high asset concentrations in loans and trading account assets were more likely to have a split rating, particularly if the bank's capital ratio was low. Iannotta (2006) undertakes a similar analysis for bonds issued in Europe from 1993 to 2003, and also concludes that bank bonds are more likely to carry split ratings. However, the European evidence differs from Morgan's US findings in several ways. Iannotta does not find insurance companies to be more opaque. He also identifies three industries (construction, 'energy and utility', and 'other') with more splits than the banking industry (see his Table 3), raising some question about the true implication(s) of split bond ratings.

The available evidence about bank opacity is limited and inconclusive. However, even if banks are not unusually opaque, their opacity may be more costly for society if depositors (or other stakeholders) are unusually subject to irrational runs or cascades of opinion.

Do bank investors (depositors) make inappropriate inferences?

Aharoney and Swary (1983) identify two reasons why one bank's failure might influence the market's assessment of other banks' value. First, if other banks share a risk exposure with the failed bank, their claimants would *rationally* perceive lower values. Such a revaluation constitutes a normal and desirable aspect of financial markets. Second, other banks' claimants might *mistakenly* infer that their own banks' quality had deteriorated. In a banking panic, many depositors withdraw their funds because they believe that their bank is *or might be* insolvent. Even a hint of insolvency might generate a rational run if outsiders find it difficult to value

their bank and the cost of withdrawing deposits is small (Diamond and Dybvig, 1983). Contagious runs impose unwarranted costs on other banks and may pose a special threat for firms that fund long-term, illiquid assets with shorter-term deposits—such as the UK mortgage lender Northern Rock.

How important is this phenomenon? Should government regulation or deposit insurance be implemented to reduce the possibility of costly bank runs? Because we lack data on day-to-day deposit flows, some researchers have examined contagious inferences among other outside claimants, bank-equity holders. Here, the available research indicates that bank stakeholders generally *do not make* unfounded inferences about other banks when one bank announces a problem. Aharoney and Swary (1996) identify five large bank holding companies (BHCs) that failed in the southwestern US between 1986 and 1989. If investors were inclined toward irrational inferences, we should observe unrelated banks losing market value as information emerged about the five failures. To test whether this occurred, Aharoney and Swary computed the abnormal equity returns for thirty-three surviving southwestern banks on each of ten 'critical event' dates associated with the five failures. They then regressed each critical event's vector of abnormal returns on three characteristics of the surviving bank holding companies: the physical distance between the BHCs' headquarters and that of the troubled institution; the BHCs' total assets; and the BHCs' book equity ratios. Aharoney and Swary found that bank-specific characteristics explained part of the cross-sectional variation in surviving banks' abnormal returns (see their Table 2), and concluded that investors were making rational inferences about the value of surviving banks. In other words, Aharoney and Swary found no evidence consistent with irrational, contagious runs.

Empirical evidence about market monitoring

Numerous studies have concluded that bank security prices reflect changes in 'true' condition quite accurately, at least on average. Academic and regulatory research began with studies of subordinate debentures and uninsured deposits in the US, and then proceeded in four directions: examining liability rates in other countries; studying additional financial instruments; incorporating quantity effects; and recognizing the endogeneity of prices and security issuance decisions. This empirical literature has previously been reviewed by Berger (1991), Flannery (1998), and Flannery and Nikolova (2004). Here, I will briefly summarize the evidence, without seeking to cite all of the relevant research.

The first strong conclusion: Subordinated debenture pricing

The first studies investigating market discipline of banking firms sought merely to determine whether bank share prices accurately reflected the bank's apparent financial condition. This limited objective indicates how differently investors were thought to treat banks vs. other firms. No researcher would have thought to ask this narrow question about non-financial firms! Not surprisingly, these studies (Murphy, 1979; Pettway, 1976; Pettway, 1980; and Pettway and Sinkey, 1980) indicated that investors held generally accurate views about banking firms' values. However, share prices were initially thought to be poor instruments of market discipline because they would not necessarily reflect changes in bank creditworthiness. Although share values unambiguously increase when a bank's asset value rises, the share price might also increase if its asset *risk* rises substantially.[5] Given this ambiguous interpretation of a share-price change, researchers' attention soon turned to the pricing of subordinated debentures, whose payoffs closely resemble those of a deposit insurance fund. If a bank does well, the debentures are paid off at par and the insurance fund suffers no loss. If the bank fails, the debentures and the insurance fund both suffer losses.

Two studies of bank-subordinated debenture pricing concluded that the required debenture yield was *not* closely related to the issuing bank's financial ratios in 1983–4 (Avery, Belton, and Goldberg, 1988; and Gorton and Santomero, 1990). Although this work seemed to support the concept of bank opacity, the presence of (conjectured) government guarantees confounded the results. If government protects a failed bank's debenture holders from loss, they have less incentive to monitor their bank's condition. The US government provided extensive support for weak financial institutions in the 1980s, although most of that support was withdrawn by the early 1990s. Taking a longer time series of debenture yields, Flannery and Sorescu (1996) show that a bank's default risk affected its equilibrium debenture yield after the government support had been curtailed. Implicit government guarantees had initially muted investors' concerns about possible bank defaults, but risk-sensitive pricing emerged once those guarantees were weakened. By the end of the 1990s, many other studies had confirmed the correspondence between debenture- or deposit-risk premiums and the issuer's 'true' financial condition.

[5] Higher asset risk can raise a firm's equity value because it increases the shareholders' payoffs in profitable times without commensurately reducing their payout in unprofitable times. (A failed firm's shareholders fare equally poorly if its assets fall one dollar below its liabilities, or many dollars.) Gavish and Kalay (1983) discuss the shareholders' ability to extract value from a firm's debtholders by substituting riskier assets after the debt's terms have been negotiated. As a practical matter, the positive relation between default risk and bank equity value occurs only with unusually high leverage.

Multinational tests of market monitoring

The initial evidence about market monitoring derived from the US, for which data on publicly traded bank obligations were most readily available. Subsequent studies looked further afield. Sironi (2003) studied the credit spreads paid by sixty-five European banking institutions on 290 bonds issued during the period 1991–2000, and concluded that the spreads reflected bank risk. As in the US in the 1980s, however, European spreads reflected both the bank's inherent riskiness and the likely extent of external (government) support if the bank encountered financial difficulties. Sironi found that bank-specific risk measures affected debenture spreads more prominently later in his sample period, consistent with the hypothesis that investors became less confident that 'TBTF' support would protect them from possible default losses. Pop (2006) also studied non-US risk premiums, using secondary-market debenture prices for ninety-five banks from fourteen countries during the period 1995–2002. Even though secondary market bond prices are sometimes characterized as noisy and inaccurate, Pop reports that the implied risk premiums significantly reflect the banks' stand-alone credit ratings and their apparent access to external support.

Rather than examining the cost of bank funds, Nier and Baumann (2006) study the capital ratios of 729 banks in thirty-two countries during the 1990s. They argue that capital market discipline requires banks with more portfolio risk to operate with larger capital cushions. Nier and Baumann find evidence consistent with this theory. Banks whose governments provide relatively little support for troubled institutions hold relatively more capital when their liabilities include a higher proportion of uninsured claimants. In contrast, the relation between asset risk and leverage is weak in countries with stronger government support for troubled banks.[6]

Monitoring with other bank liabilities

Furfine (2001) extended the documentation of market monitoring to the US interbank market. Starting with information about all Fedwire℠ transactions during the first quarter of 1998, he selected those most likely to be interbank loans and repayments.[7] The interest rate charged on the transaction could then be inferred by comparing the repayment amount with the amount initially transferred. Lower interbank loan rates were paid by banks with higher capital, lower delinquent loans, and higher income. Moreover, Furfine documents a quantity

[6] Flannery and Rangan (2008) also find a close connection between risk and capital ratios at large US bank holding companies, but only after policymakers had weakened TBTF policies.

[7] These were assumed to be round-sum transfers that could be associated with reverse transactions the following business day, in an amount similar to the initial transfer plus a day's interest.

effect in interbank lending: when banks become more than slightly risky, their federal fund borrowings shrink substantially. A recent study by King (2008) confirms these results using a different methodology over the period 1986–2005. Further, he concludes that the risk sensitivity of interbank loans increased in the second half of his sample period, reflecting the fact that 'regulatory reforms in the early 1990s . . . shifted more of the costs of failure to fed-funds sellers, thus raising the expected costs of fed-funds defaults and providing lenders stronger incentives for caution' (King, 2008: 207).

Despite the demonstrated value of bank fixed claims in assessing creditworthiness, relatively few banks have observable debenture or uninsured deposit rates and the market for uninsured obligations is less liquid than equity markets. Because equity and debt prices should reflect the same information about an issuer's condition, it seems unwise to ignore equity returns when studying market discipline (Saunders, 2001).[8] Several recent studies have evaluated the information content of bank equity prices, despite the fact that equity prices have an ambiguous relation with bank risk under some circumstances. Krainer and Lopez (2004) investigate how much information is contained in equity return measures beyond that in a ratios-based model designed to predict official rating changes.[9] The augmented model predicts more true changes in bank condition, but it also generates a larger number of false positives. Despite this limited payoff, Krainer and Lopez conclude that incorporating equity market information into formal supervisory processes is cheap and therefore worth doing.

Gropp, Vesala, and Vulpes (2006) compare the predictive abilities of debt and equity market signals for a sample of European banks.[10] With no supervisory ratings for European banks and virtually no outright failures, they choose to predict banks whose Fitch[11]/ International Credit Brokers Alliance[12] (ICBA)

[8] Caldwell (2007) observes that there has been little research aimed at determining the theoretical benefits of the market disciplinary effects of debt vs. equity. Gonzalez-Rivera and Nickerson (2006) use a vector of senior debt, debenture, and equity prices for Fannie Mae and Freddie Mac (two housing financers sponsored by the federal government) to infer the (common) underlying assessment of firm condition.

[9] The Federal Reserve inspects bank holding companies and rates their condition. Before 2005, this rating was the BOPEC (an acronym of the five key areas of supervisory concern: the condition of the bank holding company's (B)ank subsidiaries, (O)ther non-bank subsidiaries, (P)arent company, (E)arnings, and (C)apital adequacy. An aggregation of these five points yielded an overall assessment of BHC condition, called the BOPEC rating). On 1 January 2005, the BOPEC framework was replaced with an 'RFI(C)' system, which evaluated a BHC's: risk management (R); financial condition (capital adequacy, asset quality, earnings, and liquidity) (F); impact of the parent company and non-depository entities on the subsidiary depository institutions (I); and a composite rating (C).

[10] Gropp, Vesala, and Vulpes (2006) have equity data for 86 banks and debenture spreads for 53.

[11] Fitch Ratings is 'a global rating agency committed to providing the world's credit markets with independent and prospective credit opinions, research, and data' <http://www.fitchratings.com/corporate/index.cfm>.

[12] ICBA is 'the world's largest team of independently-owned, specialist trade credit insurance brokerages' <http://www.icba-online.com/>.

'individual rating' falls to 'C' or lower. (Most of their 'C' sample banks received some sort of external support.) Gropp, Vesala, and Vulpes find that equity prices alone predict a substantial proportion of negative credit events twelve to eighteen months before they emerge. The debenture spread alone predicts negative credit events only for a subset of the banks, and not more than twelve months in advance.[13] Gropp, Vesala, and Vulpes then build a prediction model based on four accounting variables (capital ratio, efficiency, problem loans, and ROE) and ask how much explanatory power is *added* by the market variables. Like Krainer and Lopez, they find a statistically significant, but small, increase in explanatory power. However, the authors note that their basic prediction model is sparse, and supervisors probably have further information not included in the accounting variables. Hence, the improved predictions based on either debt or equity prices provide a relatively modest *upper* bound on the possible value of market information in predicting bank downgrades.[14]

Quantity effects of bank condition

Although researchers first associated default risk premiums with bank risk, there may also be a quantity effect: depositors might withdraw funds from banks they perceive as riskier, even if they are offered a higher interest rate to stay. Two studies using data from the Great Depression document this quantity response. Saunders and Wilson (1996) evaluate deposit flows at failed banks during the period 1929–33. They match each failed bank with a nearby bank that survived past 1933, and find that the failed banks had lower deposit growth as much as three years before their demise. Apparently, depositors responded to perceived weaknesses well before the bank's failure date. Calomiris and Mason (1997) analyze the failure of twenty-six Chicago commercial banks during one week in June 1932. They label this incident a 'panic' because substantial deposits were withdrawn from both sound and unsound banks. However, Calomiris and Mason conclude that the failed banks were truly insolvent, while the survivors were not. Moreover, objective disparities between the failures and the survivors existed before the crisis struck: the failed banks had lower equity prices (Tobin's Q ratio) as long as eighteen months before the panic, and their total interest expenses during the second half of 1931 significantly

[13] In order to make the debenture spread a reliable predictor, Gropp, Vesala, and Vulpes (2006) discard banks with a relatively high chance of receiving external support. Surprisingly, they also need to omit UK banks, which pay much higher spreads (an average of 208 bps) than the banks on the Continent (44 bps). Sironi (2003) found a similar, unexplained relation between UK and Continental banks. This book's editors suggest that higher UK bank debenture rates may reflect the fact that they are callable.

[14] It would be worth checking whether market information could be more valuable in countries with more limited supervisory capabilities.

exceeded those for surviving banks. Martinez-Peria and Schmukler (2001) identify similar negative relationships between deposit volumes and bank condition for Argentina, Chile, and Mexico in the mid-1990s: banks with riskier-looking financial ratios pay higher deposit rates *and* attract less deposits.

Billett, Garfinkel, and O'Neal (1998) show how quantity responses to changes in bank risk reflect an interplay between market and government discipline. They identified a set of bank holding companies whose bond rating fell. During the quarter of the downgrade and the subsequent quarter, downgraded banks tend to replace some of their large, uninsured deposits with insured retail balances. Upgraded banks made the opposite change. This behavior implies that relatively weak bankers viewed government supervision (discipline) as less restrictive than market supervision. A similar phenomenon occurred in the US during the subprime credit crisis of 2007. When credit quality uncertainty made it difficult for banks to borrow in the private markets, many banks simply increased their advances from the Federal Home Loan Bank System (Bech, 2007). These two examples illustrate that market discipline cannot substantially affect bank behavior if bankers can easily shift to less disciplinary (that is, government-regulated) funding sources when problems arise.

Using market information in bank supervision

Figure 15.1 identifies three elements of market discipline for banking firms: monitoring, direct influence, and indirect influence. The empirical evidence described in the preceding section indicates that large investors can reliably monitor the condition of banking firms. (The market's identification is imperfect, of course, but neither can supervisors perfectly assess a bank's risk exposure.) However, market prices do have limitations that are important to recognize.

Limitations of market information

I now discuss some features of market information that may limit its value in the supervisory process.

Supervisory effects on public information

Effective market monitoring relies on managers (or auditors) providing accurate, timely information about the firm's condition. Even under applicable accounting

standards, however, managers retain considerable discretion about how and when they report new developments. For example, managers are more likely to share positive news than negative news with outsiders,[15] which makes independent information-gatherers particularly valuable for discovering negative information. Studies of bond rating changes indicate that share prices respond to rating downgrades much more strongly than to upgrades (Dichev and Piotroski, 2001).[16] Berger, Davies, and Flannery (2000) find a similar result for supervisory ratings: stock prices tend to fall following a supervisory ratings downgrade, but do not change significantly following an upgrade. Managers' reluctance to share 'soft' negative information with the market thus limits the ability of bank security prices to flag problem situations.

Several studies have shown that supervisors affect the quality of information released by bank managers. For example, the integrity of banks' loan loss provisioning appears to depend in part on supervisory pressure applied during on-site examinations (Gunther and Moore, 2003). Flannery and Houston (1999) similarly report that large bank holding companies' book and market values correspond more closely to one another when a Federal Reserve inspection has occurred recently.

Another challenge to accurately reflecting bank condition in market prices comes from the complex corporate structures employed by modern financial institutions. This is particularly true in the US, where bank holding company regulations have tried to maintain legal distinctions between commercial banks and other sorts of financial firms. Assuming that supervisors care more about some parts of the financial conglomerate than others, they must assure that securities prices reflect the condition of those specific components. For example, the Shadow Regulatory Committee proposes that mandatory subordinated debentures be issued by a holding company's *banking* subsidiary, because they feel that the non-bank subsidiaries pose no (at least, less) threat to financial stability.

The importance of accurate information provision is recognized in the Third Pillar of the Basel II Accord, 'Market Discipline'. Rather than specifying ways in which supervisors hope to rely on market discipline, Pillar III seeks 'to *encourage* market discipline by developing a set of disclosure requirements which will allow market participants to assess key pieces of information on the scope of application, capital risk exposures, risk assessment processes, and hence the capital adequacy of the institution' (Basel Committee on Bank Supervision, 2006: 226, emphasis added). One wonders whether a single set of supervisory provisions will elicit the

[15] For example, Nier and Baumann (2006) find that banks holding more capital per unit of measured risk also disclose more details about their risk exposures. In other words, riskier banks disclose less information.

[16] The rating agencies themselves provide a type of market discipline when they publicize their assessments of a firm's condition.

information investors really need to assess bank condition—particularly as the bank's condition deteriorates.

Closure rules

Stern and Feldman (2004) argue that supervisors' willingness to support large ('systemically important') firms constitutes the most important impediment to effective market discipline. Explicit government guarantees weaken investors' incentives to gather information, which may leave security prices relatively uninformative. In other words, bank claimants must feel 'at risk' if market discipline is to function. Even without formal *ex ante* arrangements, investors might rationally conjecture that government support will occur. Mishkin (2006) argues that US supervisory reforms, particularly 'prompt corrective action', have reduced the probability of government bailouts. At the same time, he concedes that 'the "too-big-to-fail" problem has the potential to get more serious over time' (Mishkin, 2006: 1003). The main point here is that insured claimants have little incentive to evaluate the creditworthiness of their banks, unless they doubt the insurer's ability to honor its obligations.

Early in the current financial crisis (September 2007), the British Treasury guaranteed all deposits at Northern Rock, despite the fact that formal deposit insurance was limited to only £31,700 per account.[17] Was Northern Rock a special case, or does it reflect a general reluctance of supervisors to permit large failures? We now know that Northern Rock was merely a harbinger of things to come. The Fed intervened to facilitate JPMorgan Chase's acquisition of Bear Stearns in March 2008, and during September of 2008 governments in England, Europe, and the US broadly guaranteed deposits (large and small) and other bank obligations. At least for the near future, market-monitoring incentives have been weakened by government efforts to control the financial crisis.

Even if a government were *willing* to close large financial firms, operational difficulties remain. Failures must be resolved quickly so that large (relatively informed) liability holders and shareholders cannot escape losses. Rapid resolution requires credible *ex ante* policies for closing a bank quickly (Stern and Feldman, 2004). At present, these policies do not exist in the US even for relatively simple institutions. For example, banks do not generally organize their deposit information by insurance status, yet FDIC needs this information to implement a rapid resolution. Without being able to separate insured from uninsured balances, granting extended coverage to all claimants will sometimes seem the best way to maintain liquidity.[18]

[17] Prior to 1 October 2007, the Financial Services Compensation Scheme (FSCS) limited deposit compensation to 100% of the first £2,000 and 90% of the next £33,000. The 10% 'haircut' for even moderate saving balances probably encouraged the run on Northern Rock. Subsequently, coverage was raised to 100% of the first £35,000.

[18] Part of the depositors' incentive to run on Northern Rock derived from uncertainty about how long it would take to receive insurance payments. Even after the ensuing deposit insurance reforms, the FSCS does not aspire to maintain a depositor's liquidity. Rather, it 'generally aims to pay

Resolution problems would be even more daunting for a large, international bank. Can an integrated, worldwide institution's liabilities be separated from its assets held in the same jurisdiction? What sort of netting or offset rights exist? Which ones have been tested? How might ring-fencing affect settlements and therefore customers' incentives to run at the first sign of a problem? The Federal Reserve and other supervisors have been working with large banks to clarify these questions, and to make over-the-counter (OTC) trading and settlement systems more reliable. Still, it remains unclear whether a large, complex financial institution can be reorganized quickly enough that 'uninsured' deposit or debt investors are seriously at risk of loss.

Endogenous prices

Studies confirming the effectiveness of market monitoring have treated market prices as exogenous estimates of firm value. Two recent lines of research show that this assumption may be incorrect in a way that substantially complicates how supervisors should interpret market price information. First, Covitz, Hancock, and Kwast (2004) show that new debenture issues tend to occur when investors are most optimistic about the bank's prospects.[19] The observed spread on new debentures tends to understate the issuers' average future risk. If this effect is important, *secondary* market debenture prices will reflect bank risk more accurately than *issue* prices, despite assertions that corporate bond prices are noisy because they trade in thin markets (Pop, 2006). Some analysts would minimize market timing by requiring banks to issue new debentures on a fixed schedule.

A second, potentially more serious, endogeneity in market prices arises because anticipated indirect discipline may cause a security's price to reflect more than the firm's current condition. If supervisors use share prices to determine how they will treat a bank, the share price may become an ambiguous indicator of the bank's true condition. Is a share price high because investors believe the bank *is* sound, or because they believe that supervisors will take steps to *make it* sound?[20] The problem is readily illustrated by the following example. Suppose all large banks had outstanding debentures, and that supervisors forced banks to sell equity whenever their debenture spreads exceeded some fixed amount (e.g., 2 percent). If the banks could always issue new equity and supervisors always forced such an issue, spreads would never exceed 2 percent, regardless of the bank's true condition. In other words, *indirect influence* may be subject to the Lucas critique or Good-

compensation within six months of receiving a completed application form' (Financial Services Compensation Scheme, n.d.: 11).

[19] Such market timing is not unique to banking firms. For example, all types of firms tend to issue new shares following a run-up in their stock prices.

[20] Birchler and Fachinetti (2007) report that the problem was previously recognized in Krugman's (1991) analysis of foreign exchange rate target zones.

hart's law: when policy begins responding to market prices, those prices no longer behave as they did in the past.

DeYoung, et al. (2001) find some support for this sort of price endogeneity when they study how bank debenture spreads change in the wake of revised supervisory assessments. An odd thing happens after supervisors downgrade a bank holding company to problem status (BOPEC[21] rating = 4 or 5): the banking firm's debenture spreads soon *narrow*. It appears that bank bond holders believe supervisors have a comparative advantage in forcing banks to enhance their solvency. Once the market knows that supervisors have recognized the problem, investors feel that the chances for a good outcome have improved.

Several recent theoretical studies model how policy reactions might affect the informativeness of market prices (Lehar, Seppi, and Strobl, 2007; Bond, Goldstein, and Prescott, 2006; and Birchler-Facchnietti, 2007). These papers reach a similar conclusion: if bank security prices are used to identify problems that supervisors may be able to reverse, the security prices may diverge substantially from investors' assessment of the firm's present condition. This does not mean that market prices cannot provide useful information; rather, those prices must be evaluated in light of the potentially confounding effects of indirect influence.

Finally, supervisors' use of market prices may compromise private incentives to collect new information. Private investors spend resources to gather new information about a firm because they hope to benefit from trading the firm's debt or equity before other investors learn the same news. Yet if policymakers act on this information—as it is transmitted through security prices—they might curtail the value of collecting private information.[22] In turn, this might make observed security prices less informative (Lehar, Seppi, and Strobl, 2007). Gropp, Vesala, and Vulpes make the point quite directly: if 'supervisors . . . conditioned their intervention on market prices, market participants may reduce their monitoring effort and, hence, stop providing useful signals' (2006: 425).

Best uses for market information

I have located four regulatory staff documents describing the current uses of market information by US supervisory agencies (Feldman and Schmidt, 2003; Feldman and Levonian, 2001; Burton and Seale, 2005; and Furlong and Williams, 2006). All maintain that market information is often cited in assessing a banking

[21] An acronym of the five key areas of supervisory concern: the condition of the bank holding company's (B)ank subsidiaries, (O)ther non-bank subsidiaries, (P)arent company, (E)arnings, and (C)apital adequacy. An aggregation of these five points yielded an overall assessment of BHC condition, called the BOPEC rating.

[22] In a similar context, Leland (1992) considers the general equilibrium effects of insider trading, which discourages outside investors from gathering information about the firm's true value.

firm's condition, yet all find it difficult to summarize the impact of market signals on supervisory assessments or actions. It is therefore informative to observe that market information plays only a limited role in the FDIC's latest formula for assessing bank risks. The Federal Deposit Insurance (FDI) Reform Act of 2005 permitted the FDIC to design a new system of risk-based deposit insurance premiums. The assessment procedure implemented in November 2006 defines four risk categories (I–IV) on the basis of regulatory assessments (the bank's CAMELS rating[23]) and book capital ratios.[24] Market information affects the insurance premium only for large banks with low risk. Low-risk institutions are identified by their CAMELS ratings and their book-value capital ratios. No firm's annual premium can change by more than 2 bps as a result of this market information. Banks with lower capital ratios or CAMELS ratings might pay premiums as high as 40 bps, but market information has no place in assessing risk outside of the safest category.[25]

The Basel II Accord also includes only a limited role for the type of market discipline discussed so far in this chapter. Its third pillar is labeled 'market discipline' but primarily concerns standardized information disclosure. Market forces have their most prominent role in Basel II in the advanced internal ratings based (AIRB) approach to capital adequacy, which relies quite extensively on rating agency assessments. Although ratings-based capital standards can reasonably be considered a form of indirect influence—using market information to assess capital adequacy—one can question the wisdom of this reliance when the rating agencies appear to operate under a crippling set of conflicting interests. These conflicts have been revealed with startling clarity in the agencies' recent treatment of structured securities based on subprime mortgages.

Perhaps it is appropriate to place little confidence in market (counterparty) discipline. We know very little about the market's ability to discipline financial firms directly because direct influence is difficult to identify for any sort of firm. It has been studied for banks only by Bliss and Flannery (2002), who cannot draw

[23] The components of a bank's condition that are assessed: (C)apital adequacy, (A)sset quality, (M)anagement, (E)arnings, (L)iquidity, and (S)ensitivity to market risk (analogous to BOPEC).

[24] A summary of the initially proposed system is provided in Federal Deposit Insurance Corporation (2006: Table 4). The final rules differ from what was proposed primarily in having uniformly higher premiums for all risk classes. When FDIC proposed to raise overall deposit insurance premiums on 16 October 2008, it did not change the extent to which market information can affect those premiums.

[25] The new FDIC risk assessment procedures include the possibility of 'Consider[ing] other market information, such as subordinated debt prices, spreads observed on credit-default swaps related to an institution's non-deposit obligations, equity price volatility observed on an institution's parent company stock, and debt rating agency "watch list" notices. These additional market indicators would be especially beneficial in assessing whether the insurance score accurately reflected the relative level of risk posed by an institution' (Federal Deposit Insurance Corporation, 2006: 419–24). Yet a process that incorporates many sources of market information may in fact place little weight on any of them.

strong conclusions. Nor is this necessarily surprising. Atomistic market agents, such as stock and bond investors, suffer from well-known collective action problems. Even when some action would benefit a group of investors—for example, by renegotiating downward a debt issue's required repayment—it may not occur because individual actors try to free-ride off the actions of others. Dewatripont and Tirole (1994) argue that depositors' collective action problem necessitates government supervision. DeYoung, et al. (2001) report some evidence consistent with government's comparative advantage in influencing bank actions.

The preceding list of problems associated with market discipline might seem to imply that government should play an unusually active role in supervising financial firms. However, government supervision also has shortcomings. Special features of government action make it difficult for supervisors to act quickly once a problem has been identified. Agents must be able to defend their assessments, perhaps in court. Those assessments must be justified by GAAP accounting, which often lags changes in a firm's true condition. Delay weakens market discipline by affording informed liability holders an opportunity to escape the results of their investments. Delay also encourages shareholders to gamble for resurrection. By contrast, market prices quickly reflect changes in firm condition. Market information can be used to confirm supervisory assessments and also to limit supervisory tendencies to forebear. Bliss (2001) argues that supervisors should specify critical values for debenture spreads or stock price declines, and be required either to take action when those values are crossed or to explain publicly why no action is required. This seems like a workable rule linking market prices to supervisory actions.

All things considered, it appears that market information can best be used to reinforce supervisory assessments and to constrain the supervisors' ability to forebear. Additional research on the efficacy of indirect influence would yield valuable information.

Implications of the financial crisis[26]

The preceding sections of this chapter were substantially completed in April 2008. JPMorgan Chase had just purchased Bear Stearns, but the extent of the worldwide financial disruption was not yet apparent. Figure 15.2 plots the mean CDS premium for three large banks (Bank of America, JPMorgan Chase, and Wells Fargo) for the year ending 30 April 2008. The CDS premium initially lay below 20 bps, but rose rapidly to 50 bps in August 2007. That autumn's continuing turmoil generated still

[26] Much of the material for this section is based on Flannery (2009).

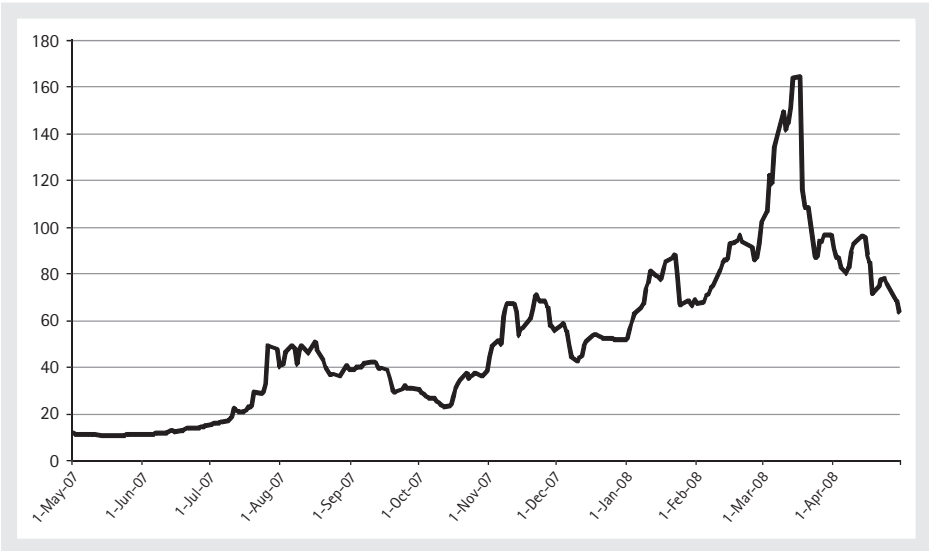

Fig. 15.2. The average annual premium (in basis points) of a five-year credit default swap for three strong commercial banks (BAC, JPM, WFC)

higher CDS premiums through mid-November. Going into the new year, the mean CDS spread for these three banks stood near 60 bps. There is a clear spike (to 160 bps) around the date of the Bear Stearns failure (17 March), but the premium fell back to about 60 bp by the end of April. Perhaps the Fed's extraordinary lending efforts might had stabilized the financial system!

This hope faded, of course, in the late summer of 2008. Amid worsening economic conditions and falling house prices, the US government placed into receivership the two GSE mortgage lenders (Fannie Mae and Freddie Mac) on Monday, 8 September. Following several months of increasing liquidity problems, Lehman Brothers filed for bankruptcy on 15 September. That same weekend, Bank of America (BAC) agreed to acquire Merrill Lynch. Money market mutual funds came under severe withdrawal pressures after the Reserve Primary Fund's position in Lehman Brothers commercial paper forced it to 'break the buck' on 16 September, and the Treasury agreed to insure money market funds that requested such protection for a premium of 10 bp per year. During the week of 22 September, Morgan Stanley and Goldman Sachs converted to bank holding companies, presumably to access the Fed's various lending programs. Washington Mutual became the largest thrift institution ever to fail, on 26 September. Citigroup announced its intention to acquire Wachovia Bank with substantial federal aid, only to have Wells Fargo (WFC) make a dominant bid the following week. The FDIC deposit insurance limit was raised from $100,000 to $250,000 on 3 October. Later in October, Treasury Secretary Hank Paulsen mandated that ten large financial institutions

would accept government capital injections to improve their creditworthiness. All the while, stock market values were falling.

Before September 2008, UBS was the only European bank recognized as heavily involved in the subprime crisis. However, the crisis spread to large parts of Europe in the second half of that month. Fortis, ING, and Dexia were supported by multiple governments' actions. Iceland's three largest banks were taken over by financial supervisors. Ireland guaranteed all bank deposits, as did Denmark, Germany, Britain, and a number of other European countries.

What do these latest events imply about the efficacy of market discipline in the financial sector?

Market monitoring

Ideally, effective market monitoring would have signaled possible problems at one or more financial institutions before those problems became realized disasters. To assess this possibility, I constructed an index of financial firms' share values by averaging four Fidelity sector funds: brokers, financial services, insurance, and banks (ticker symbols FSLBX, FIDSX, FSPCX, and FSRBX, respectively). Figure 15.3 plots this index and the S&P 500 index, with each index scaled by its own level in April 1988. The financial index rose more or less continuously from 2003 (not shown) to a peak in May 2007, when two Bear Stearns hedge funds announced the first serious problems associated with pricing subprime mortgage securities. The vertical line in Figure 15.3 marks the market turbulence of August 2007. Clearly,

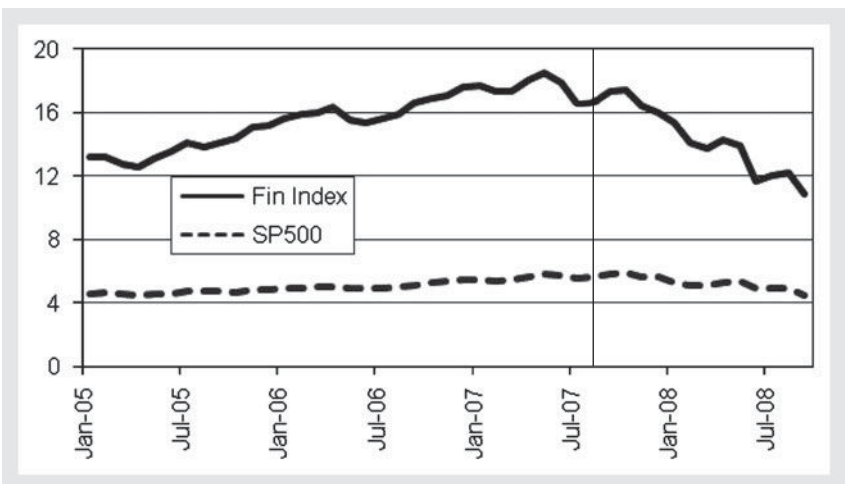

Fig. 15.3. Return to a 'financial index' and to the S&P 500, each deflated by its April 1988 index value

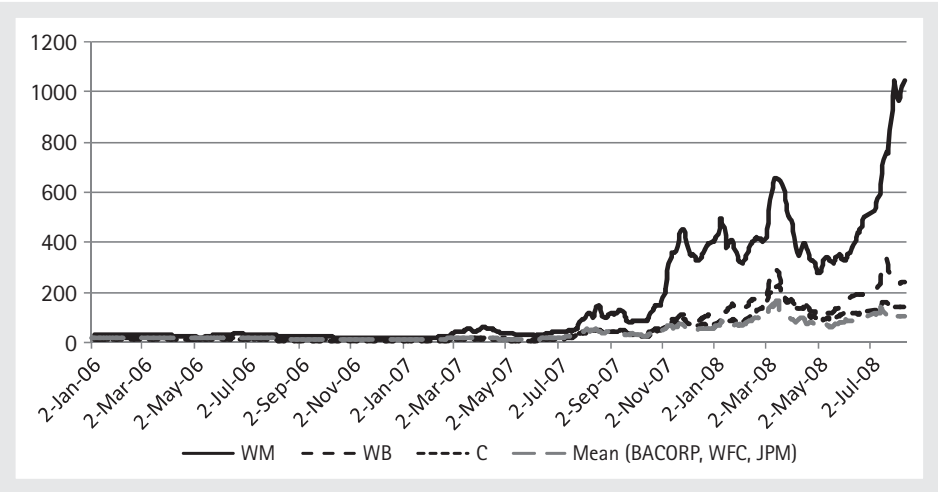

Fig. 15.4. Credit default swap premiums (five-day moving average) for three troubled institutions and the mean of three banks considered relatively sound (at the time)

the financial firms' share prices provided little, if any, advance warning of the subprime crisis. In fact, financial share prices rose again in the autumn before starting their sharp decline in November.

Nor did CDS premiums indicate a pending problem. Figure 15.4 plots the annual premium on a five-year CDS for three individual banks (Washington Mutual, Wachovia, and Citigroup) and for the average of three banks that were considered relatively sound at the time (Bank of America, JPMorgan Chase, and Wells Fargo). Prior to August 2007, we see no significant fluctuations in CDS premiums except for WaMu, whose premium rose in the spring of 2007. All the banks' CDS premiums rose in August and remained elevated thereafter. WaMu's premium substantially exceeded those of the other banks. The month before it closed, WaMu's CDS spread had risen to 1,000 bps.

Figure 15.5 plots CDS premiums for five large commercial banks from 1 July 2007 to 12 August 2008. Note that the vertical scale is narrower than in Figure 15.4, because Figure 15.5 does not include the riskiest firm, WaMu. Beginning in August 2007, these banks' CDS spreads rose by similar amounts until November, when clear differences emerged. The sharp peak in all premiums occurred around the Bear Stearns failure, and premiums again rose after April. Wachovia (WB) had the largest risk premium by a substantial margin, followed by Citigroup (C). At least through mid-August, we see no indication that BAC would encounter serious difficulties at year end.[27]

[27] To be fair, though, BAC's greatest problems derive from Merrill Lynch's portfolio losses, and BAC did not announce its intention to acquire Merrill until mid-September.

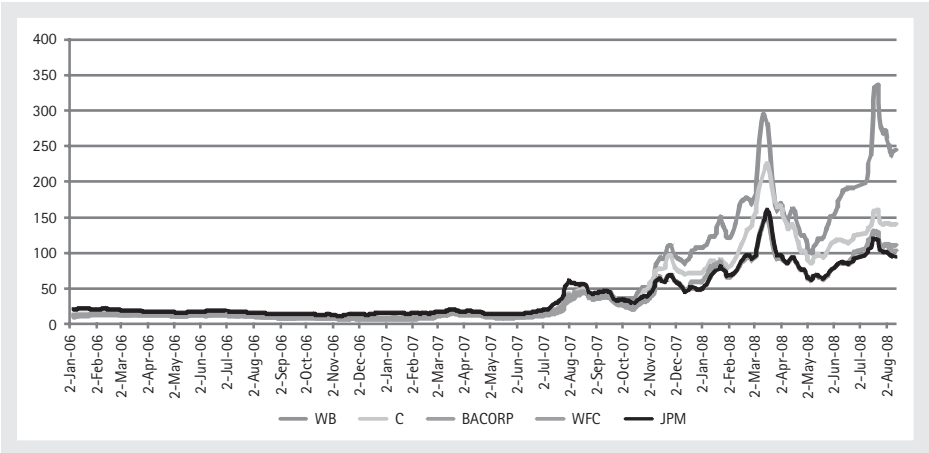

Fig. 15.5. Credit default swap premiums (five-day moving average) for five large commercial banks

In summary, the data indicate that market monitoring worked relatively well *across financial institutions* at any point in time, but it provided *virtually no advance warning* of the financial crisis. Going forward, even this extent of market monitoring is threatened by the broad government guarantees extended during September of 2008 to address the crisis. As noted above, stakeholders have no incentive to monitor their banks if the government insures them against default loss. It remains to be seen whether supervisors can restore a sense of credit risk exposure for large firms when we return to normal financial conditions.

Faith in market monitoring has not been buttressed by the failure of market prices to predict the 2007–9 financial crisis. But this failure does not necessarily support special supervisory treatment of banks, for two reasons. First, supervisors—who can supposedly see through the opacity better than investors—were equally surprised when the subprime crisis emerged. Second, non-financial firms have also fooled outside investors about their value for protracted periods—e.g. Enron or WorldCom.

Financial firms' governance

The financial crisis was initiated by careless actions in the subprime lending market. Although some authors (Bhardwaj and Sengupta, 2008) contend that there was no deterioration in subprime underwriting standards during 2005–6, the evidence more strongly indicates that fraud flourished and underwriting standards collapsed in part because institutional investors were willing to purchase assets with little regard for their actual quality. Instead, bonuses based on short-term performance motivated mortgage originators to 'do deals' and sell the product, while institutional investors appeared to substitute credit ratings for their own due diligence.

Much of this behavior would have been avoided by firms with strong corporate governance and directors whose interests were aligned with shareholders'. Had investors been concerned with the long-run, reputational effect of their actions, they would have been less willing to purchase the sort of subprime securities they were offered. Without the easy securitization, loan originators would have been more careful about fraudulent borrowers and biased appraisals. It seems that direct market influence failed, and that supervisory influence provided no substitute.

This failure of direct market influence is consistent with the discussion above. Atomistic shareholders and bondholders cannot exercise much control, at least without great expense. Although corporate governance is problematic at many firms, this failure seems more serious in the financial sector. First, the effects of an employee's actions are more difficult to evaluate in the short-run. When a manufacturing firm's salesman sells X units of product during the year, few of those sales will be reversed in future years. For a lender or trader, however, the full effect of one year's actions may become apparent only later. This requires a different sort of incentive contracting in financial firms—one that is more long-term than what we commonly see. So, why are incentive payments structured so inappropriately at financial firms? Perhaps competition for the best personnel saddles the financial industry with a collective action problem in changing from short-term to longer-term performance measures. Second, statutes impose relatively severe ownership restrictions on many types of financial firms. It is more difficult to acquire a large block of shares, and hostile takeovers are inhibited by the need for supervisory acceptance of the new owners. Finally, reputational effects may be weaker than commonly supposed. Firms only worry about their reputations if customers understand how their agents have performed and also remember that performance when it comes time to allocate new business. This seems to work less well in practice than it does in theory.[28]

With relatively ineffective stakeholder governance at financial firms, government supervisors must be more prominent in imposing external discipline. The recent financial crisis has done nothing to reverse the conclusion that direct market influence is weak. The results in DeYoung, et al. (2001) suggest that the market understood this situation, but recent events suggest that supervisors did not.

Market 'values'

The 1980s' savings and loan crisis was substantially exacerbated by the Federal Home Loan Bank Board's willingness to ignore the asset values lost by many thrifts with fixed rate loan portfolios. In the current crisis, many firms have argued that 'mark to market' accounting requirements have forced them to overstate their true losses on subprime-related securities. Their argument has been that current market

[28] Indeed, Alan Greenspan confessed to a Congressional committee in Oct. 2008 that he had overestimated the ability of firms to protect 'their own shareholders and their equity in the firms'.

prices reflect 'fire-sale' values, and that assets held to maturity (or, at least, for a long time) will eventually generate cash flows whose present value exceeds the current market prices. This is a complex issue, and permitting a financial institution to value assets above their current market values is fraught with dangers. Arguably, however, this crisis has illustrated a new form of financial externality, in which many firms holding the same asset type might all suffer if they are, as a group, pressured to sell that asset at the same time. 'Limits to arbitrage' may indeed drive asset values away from 'fundamentals'. This is an area that requires further research in order for policymakers to treat it appropriately.

Summary and conclusions

Substantial empirical evidence indicates that market investors can identify the risks associated with investing in complex financial organizations. Valuable information can be derived from equity prices, despite the fact that other liabilities (like subordinated debt) have payoffs more similar to those of the supervisor. Quantity information may be as important as price information in inferring market assessments of bank condition. Through all these channels, market investors appear to 'monitor' reasonably well, although investors' information apparently overlaps substantially with what supervisors know. However, the prominence of conjectural government guarantees may severely inhibit stakeholders' incentives to monitor, particularly after the extensive government interventions associated with the current financial crisis. It will be difficult to re-establish (much less to reinforce) a culture of risk assessment after financial conditions become less fragile. Supervisors and legislators would do well to consider new mechanisms for designing financial firms' capital structures in a way that permits rapid resolution and imposing losses on 'uninsured' claimants. (One such device is 'contingent capital certificates' that convert from debt to equity when a firm's condition worsens (Flannery, 2005).)

Investors' ability to 'influence' bank behavior is severely compromised by collective action problems. Influence may be most effectively implemented by government supervisors, although their understandable tendency to act slowly limits their effectiveness. The most important use of market information may therefore be to overcome supervisory tendencies toward forbearance. Implementing such effects requires that specific market signals be examined and interpreted at regular intervals. 'Considering all available information' is not a sufficiently specific policy to support this role for market information. As Bliss (2001) recommends, we need tripwires that require either explicit supervisory action or a public explanation of why no action is appropriate in the case at hand.

References

Aharoney, J. and Swary, I. (1983). 'Contagion Effects of Bank Failures: Evidence from Capital Markets', *Journal of Business*, 56: 305–22.

—— (1996), 'Additional Evidence on the Information-Based Contagion Effects of Bank Failures, *Journal of Banking & Finance*, 20: 57–69.

Avery, R. B., Belton, T. M., Goldberg, M. A. (1988). 'Market Discipline in Regulating Bank Risk: New Evidence from the Capital Markets'. *Journal of Money, Credit and Banking*, 20: 597–610.

Bhardwaj, G. and Sengupta, R. (2008). 'Where's the Smoking Gun? A Study of Underwriting Standards for U.S. Subprime Mortgages', Federal Reserve Bank of St. Louis Working Paper.

Basel Committee on Bank Supervision (2006). *International Convergence of Capital Measurement and Capital Standards: A Revised Framework.*

Bech, Morten, L. (2007), 'The Federal Home Loan Bank System: The Lender of Next to Last Resort?', Conference Presentation, Federal Reserve Bank of New York. http://www.newyorkfed.org/research/conference/2007/liquidity/FHLB_LiquidityConference.pdf.

Berger, Allen, N. (1991). 'Market Discipline in Banking', Federal Reserve Bank of Chicago Bank Structure Conference Proceedings, 419–37.

—— Davies, S., and Flannery M. J. (2000). 'Comparing Market and Regulatory Assessments of Bank Performance: Who Knows What When?', *Journal of Money, Credit and Banking*, 32: 641–70.

Billett, M., Garfinkel, J., and O'Neal, E. (1998). 'The Cost of Market versus Regulatory Discipline in Banking', *Journal of Financial Economics*, 48: 333–58.

Birchler, Urs W. and Fachinetti, M. (2007). 'Self-Destroying Prophecies? The Endogeneity Pitfall in Using Market Signals as Triggers for Prompt Corrective Action', Swiss National Bank Working Paper, March 2, 2007.

Bliss, Robert, R. (2001). 'Market Discipline and Subordinated Debt: A Review of Some Salient Issues', *Federal Reserve Bank of Chicago Economic Perspectives,* 25: 34–45.

—— and Flannery, M. J. (2002). 'Market Discipline in Governance of U.S. Bank Holding Companies: Monitoring vs. Influencing', *European Economic Review*, 6: 361–95.

Bond, P., Goldstein, I., and Prescott, E. S. (2006). 'Market-Based Regulation and the Informational Content of Prices', Federal Reserve Bank of Richmond Working Paper.

Burton, S. and Seale, G. (2005). 'A Survey of Current and Potential Uses of Market Data by the FDIC', *FDIC Banking Review,* 17: 1–18.

Caldwell, Greg (2007). 'Best Instruments for Market Discipline in Banking', Bank of Canada Working Paper 2007–9.

Calomiris, C. W. and Mason, J. R. (1997). 'Contagion and Bank Failures during the Great Depression: the June 1932 Chicago Banking Panic', *American Economic Review,* 87: 863–83.

Covitz, D. M., Hancock, D., and Kwast, M. L. (2004). 'A Reconsideration of the Risk Sensitivity of U.S. Banking Organization Subordinated Debt Spreads: A Sample Selection Approach', *Federal Reserve Bank of New York Economic Policy Review*, September, 73–92.

Dewatripont M. and Tirole, J. (1994). *The Prudential Regulation of Banks.* Cambridge, MA: MIT Press.

DeYoung, R., Flannery, M., Lang, W., and Sorescu, S. (2001). 'The Information Content of Bank Exam Ratings and Subordinated Debt Prices', *Journal of Money, Credit and Banking*, 33: 900–25.

Diamond, D. and Dybvig, P. (1983). 'Banks Runs, Deposit Insurance and Liquidity', *Journal of Political Economy*, 91: 401–19.

Dichev, I. D. and Piotroski, J. D. (2001). 'The Long-Run Stock Returns Following Bond Ratings Changes', *Journal of Finance*, 56: 173–203.

Evanoff, D. D. and Wall, L. D. (2002). 'Measures of Riskiness of Banking Organizations: Subordinated Debt Yields, Risk-Based Capital, and Examination Ratings', *Journal of Banking & Finance*, 26: 989–1009.

Federal Deposit Insurance Corporation (2006). 'Assessments', *Federal Register*, 71: 41910–48.

Feldman, R. and Levonian, M. (2001). 'Market Data and Bank Supervision: The Transition to Practical Use'. Federal Reserve Bank of Minneapolis, *The Region*, 15: 46–54.

—— and Schmidt, J. E. (2003). 'Supervisory Use of Market Data in the Federal Reserve System', Working Paper, Federal Reserve Bank of Minneapolis.

Financial Services Compensation Scheme, 'A Guide to the Work of the Financial Services Compensation Scheme', http://www.fscs.org.uk/files/documents/pdfs/vcrckxvpjmrzmku.pdf.

Flannery, M. J. (1982). 'Deposit Insurance Creates a Need for Bank Regulation', *Federal Reserve Bank of Philadelphia Business Review*, January/February: 17–27.

—— (1998). 'Using Market Information in Prudential Bank Supervision: A Review of the U.S. Empirical Evidence', *Journal of Money, Credit and Banking*, 30: 273–305.

—— (2005). 'No Pain, No Gain? Effecting Market Discipline via "Reverse Convertible Debentures"', in Hal S. Scott (ed.), *Capital Adequacy beyond Basel: Banking, Securities, and Insurance*, Oxford: Oxford University Press.

—— (2009). 'The Subprime Crisis: Lessons About Market Discipline, in Douglas D. Evanoff, Philipp Hartmann, and George G. Kaufman (eds.), *The Credit market Turmoil of 2007–2008: Implications for Public Policy*, Hackensack, NJ: World Scientific Publishing Co.

—— and Houston, J. F. (1999). 'The Value of a Government Monitor for Firms with Hard-to-Value Assets', *Journal of Money, Credit and Banking*, 31: 14–34.

—— Kwan, S., and Nimalendran, M. (2004). 'Market Evidence on the Opaqueness of Banking Firms' Assets', *Journal of Financial Economics*, 71: 419–60.

—— and Nikolova, S. (2004). 'Market Discipline of U.S. Financial Firms: Recent Evidence and Research Issues', in William C. Hunter, George G. Kaufman, Claudio Borio, and Kostas Tsatsaronis (eds.), *Market Discipline across Countries and Industries*, Cambridge, MA: MIT Press, 87–100.

—— and Rangan, K. P. (2008). 'What Caused the Bank Capital Build-Up of the 1990s?', *Review of Finance*, 12: 391–429.

—— and Sorescu, S. M. (1996). 'Evidence of Bank Market Discipline in Subordinated Debenture Yields: 1983–1991', *Journal of Finance*, 51: 1347–77.

Furfine, C. H. (2001). 'Banks Monitoring Banks: Evidence from the Overnight Federal Funds Market, *The Journal of Business*, 74: 33–58.

Furlong, F. T. and Williams, R. (2006). 'Financial Market Signals and Banking Supervision: Are Current Practices Consistent with Research Findings?', *Federal Reserve Bank of San Francisco Economic Review*, 17–29.

GAVISH, B. B. and KALAY, A. (1983). 'On the Asset Substitution Problem', *Journal of Financial and Quantitative Analysis,* 18: 21–30.

GONZALEZ-RIVERA, G. and NICKERSON, D. (2006). 'Dynamic Monitoring of Financial Intermediaries with Subordinated Debt', *The Journal of Risk Finance*, 7: 463–87.

GORTON, G. and SANTOMERO, A. (1990). 'Market Discipline and Bank Subordinated Debt'. *Journal of Money, Credit and Banking*, 22: 119–28.

GOYAL, V. K. (2005). 'Market Discipline of Bank Risk: Evidence from Subordinated Debt Contracts', *Journal of Financial Intermediation*, 14: 318–50.

GROPP, R., VESALA, J., and VULPES, G. (2006). 'Equity and Bond Market Signals as Leading Indicators of Bank Fragility', *Journal of Money, Credit and Banking*, 38: 399–428.

GUNTHER, J. W. and MOORE, W. R. (2003). 'Loss Underreporting and the Auditing Role of Bank Exams', *Journal of Financial Intermediation,* 12: 153–77.

HILDEBRAND, PHILIPP, M. (2007). 'Hedge Funds and Prime Broker Dealers: Steps towards a Best Practice Proposal', in Banque de France *Financial Stability Review*, 67–76.

IANNOTTA, G. (2006). 'Testing for Opaqueness in the European Banking Industry: Evidence from Bond Credit Ratings', *Journal of Financial Services Research*, 30: 287–309.

KING, T. B. (2008). 'Discipline and Liquidity in the Interbank Market', *Journal of Money, Credit and Banking,* 40: 295–318.

KRAINER, J. and LOPEZ, J. A. (2004). 'Incorporating Equity Market Information into Supervisory Monitoring Models', *The Journal of Money, Credit and Banking*, 36: 1043–67.

KRUGMAN, P. (1991). 'Target Zones and Exchange Rate Dynamics', *Quarterly Journal of Economics*, 101: 669–82.

KWAST, M. et al. (1999). 'Using Subordinated Debt as an Instrument of Market Discipline', Study Group on Subordinated Notes and Debentures, Federal Reserve Board of Governors.

KYLE, A. S. (1986). 'Continuous Auctions and Insider Trading', *Econometrica*, 53: 1315–35.

LEHAR, A., SEPPI, D., and STROBL, G. (2007). 'Using Price Information as an Instrument of Market Discipline in Regulating Bank Risk', University of North Carolina Working Paper.

LELAND, H. E. (1992). 'Insider Trading: Should it be Prohibited?', *The Journal of Political Economy*, 100: 859–87.

MARTINEZ-PERIA, M. S. and SCHMUKLER, S. L. (2001). 'Do Depositors Punish Banks for Bad Behavior? Market Discipline, Deposit Insurance, and Banking Crises'. *Journal of Finance,* 56: 1029–51.

MERTON, R. (1995). '*Continuous-Time Finance*'. Cambridge: Basel Blackwell.

MISHKIN, F. S. (2006). 'How Big a Problem is Too Big to Fail?', Review of Gary Stern and Ron Feldman's *Too Big to Fail: The Hazards of Bank Bailouts*, *Journal of Economic Literature*, 44: 988–1004.

MORGAN, D. P. (2002). 'Rating Banks: Risk and Uncertainty in an Opaque Industry'. *American Economic Review*, 92: 874–88.

MURPHY, N. B. (1979). 'Disclosure of the Problem Bank Lists: A Test of the Impact'. *Journal of Bank Research,* 10: 88–96.

NIER, E. and BAUMANN, U. (2006). 'Market Discipline, Disclosure and Moral Hazard in Banking', *Journal of Financial Intermediation*, 15: 332–61.

PETTWAY, R. H. (1976). 'The Effects of Large Bank Failures upon Investors' Risk Cognizance in the Commercial Banking Industry', *Journal of Financial and Quantitative Analysis*, 11: 465–77.

PETTWAY, R. H. (1980). 'Potential Insolvency, Market Efficiency and Bank Regulation of Large Commercial Banks', *Journal of Financial and Quantitative Analysis*, 15: 219–36.

—— and SINKEY, J. F. Jr. (1980). 'Establishing On-site Bank Examination Priorities: An Early-Warning System Using Accounting and Market Information', *Journal of Finance,* 35: 137–50.

POP, A. (2006). 'Market Discipline in International Banking Regulation: Keeping the Playing Field Level', *Journal of Financial Stability*, 2: 286–310.

SAUNDERS, A. (2001). 'Comments on Evanoff and Wall/Hancock and Kwast', *Journal of Financial Services Research,* 20: 189–94.

—— and WILSON, B. (1996). 'Contagious Bank Runs: Evidence from the 1929–1933 Period', *Journal of Financial Intermediation,* 5: 409–23.

SIRONI, ANDREA (2003). 'Testing for Market Discipline in the European Banking Industry: Evidence from Subordinated Debt Issues', *Journal of Money, Credit and Banking,* 35: 443–72.

STERN, G. H. and FELDMAN, R. J. (2004). *Too Big to Fail: The Hazards of Bank Bailouts.* Washington: Brookings Institution Press.

CHAPTER 16

COMPETITION AND ANTITRUST POLICY IN BANKING

ASTRID A. DICK
TIMOTHY H. HANNAN[1]

Introduction

Policies designed to insure competition in the banking industry date to at least 1963, when a US Supreme Court decision held that the banking industry, like other industries, was subject to previously existing antitrust laws (US vs. Philadelphia National Bank, 374 US 321). In Europe and elsewhere, such policies were introduced considerably later. For the past twenty years, a substantial wave of consolidation has been occurring in the banking industry worldwide, usually following significant deregulation in the sector. This structural change raises anew the importance of the assessment of competition in the industry, as well as the policies relevant to its maintenance.

In this chapter, we begin by describing the various methods proposed in the academic literature to assess the state of competition in the banking industry,

[1] We thank Robin Prager for helpful comments.

discussing in detail the issues and the practical limitations encountered in their use. Importantly, we discuss several new methodologies in competition analysis that stem from recent developments in the industrial organization literature. Among the main approaches to measuring competition, we begin our review with the Structure–Conduct–Performance (SCP) paradigm, which has been extensively used in the literature. Over time, researchers developed new methods as a way to counter the problems posed by the SCP paradigm, such as its reliance on accounting data and endogeneity issues pervading the relationships among the variables of interest. These new methods have come to be known as the New Empirical Industrial Organization (NEIO). The new methods focused on a single industry, and introduced theoretical models to analyze firm behavior such that the degree of market power could be identified and estimated. It is in this vein that recent structural econometric models came to be developed. While we review the traditional approaches, we highlight these recent methodologies as a way to set the stage for a small but growing banking literature that uses these recent econometric developments. We end this section with a discussion regarding the potential of the new methodologies to improve actual antitrust practice. Next, we turn to the description of the antitrust policies actually employed in the US and the European Union. These policies, we find, are similar and continue to converge over time. We discuss the rationale behind their use and the empirical challenges in their application. We end by drawing attention to the link between antitrust regulation, bank competition, and risk taking, which gives rise to questions especially pertinent after the financial crisis that began in 2007.

Measuring competition in banking

In banking, as in other settings, the less competition faced by a firm, the greater its market power, as reflected in its ability to set price above marginal costs. Thus, measuring market power is central to the analysis of competition. The banking literature has produced a large body of work using more traditional industrial organization techniques in the measurement of market power. These are reviewed in parts 1–3 of this section.[2] More recently, banking researchers have applied newer methodologies following developments in the industrial organization literature. We review those in part 4.

[2] In our discussion, we will not attempt to conduct an exhaustive review of the literature applying the traditional methods. Instead, we direct our readers to the reviews by Berger, et al. (2004b) and Degryse and Ongena (2008) for a wider coverage of the literature.

The SCP paradigm in banking

Most studies relating to competition in the banking industry have followed the SCP paradigm, which asserts a causal relationship running from market structure to bank competitive conduct, and to overall performance. These studies seek to infer competitive conduct by tracing the relationship between measures of market structure and measures of performance, arguing that greater market concentration allows banks to enjoy greater market power. The approach has been thought to be quite relevant to policy, since (if one accepts its premise) it suggests the changes in structure that can be pursued to change competitive behavior (e.g., preventing or allowing a merger between two banks in the same market).

In this setting, traditional measures of bank performance have been based on: (1) the rate of return (2) observed price–cost margin, and (3) Tobin's Q ratio. Price, per se, is not a measure of performance, but its relationship to concentration has also been a major issue in research on banking competition, as discussed more fully below. The rate of return is usually calculated as the ratio of net income to assets or to equity. While it may be true that a bank that exercises market power exhibits higher calculated rates of return, all else being equal, it is hard to infer anything about the state of competition from it. The reasons include the well-known difference between calculated rates of return and true economic profits. This can result from differences in measuring capital at replacement cost and in accounting for depreciation, in the treatment of the long-lasting impacts of advertising and Research and Development, and in the adjustments of values made to account for inflation, to name only a few of the relevant issues.

Because of difficulties in inferring the state of competition from accounting rates of return, economists have sought in some cases to use the observed price–cost margin, defined as $(p - mc)/p$, where p denotes price and mc denotes marginal cost, also referred to as the Lerner Index. A competitive firm would exhibit an index value of zero, while, by definition, a firm exercising market power would register a positive value. Unfortunately, accurate measures of true marginal costs are difficult to come by. Moreover, determining the relevant price and cost associated with the given product or service can be quite difficult in banking, given the multiplicity of products.

Defined as the ratio of the value of a firm's assets to the replacement cost of its assets, Tobin's Q ratio has been used in a few studies of the banking industry as an alternative measure of performance (see, e.g., Keeley, 1990). The rationale for this measure is that, if the firm is worth more than it would cost to rebuild it, then excess economic profits are being earned. The advantage of the measure is that the problems of estimating true rates of return and marginal costs are avoided. The primary disadvantage is the need to obtain accurate measures of both market value and the replacement costs of the bank's assets. In banking, as in most industries, it is quite problematic to obtain an accurate measure of the firm's replacement cost. A

further problem is the fact that most banking organizations (particularly in the US) are not publicly traded, making it difficult to obtain reasonable estimates of the market value of a bank's assets as well.

For market structure, the most commonly used measures employed in SCP banking studies are measures of market concentration, such as the Herfindahl-Hirschman Index (HHI), calculated as the sum of squared market shares of all banks in the defined market (under Cournot competition, concentration moves in the same direction as price cost margins). Underlying market shares are most typically calculated as deposit shares (which are readily available for the US case and may be more relevant to competitive behavior as it applies to deposit pricing), but are sometimes calculated on the basis of loan shares (which may be more relevant to competitive behavior as it applies to loan pricing). In general, the banking industry offers substantial advantages over other industries in testing the relationship between concentration and performance, as (to be discussed in more detail below) the markets for certain types of loans and deposits are thought to be local in nature, implying cross-sectional differences in concentration that can be exploited to conduct tests of the relationship between structure and conduct at one point in time and within one industry.

Barriers to entry are also an aspect of market structure, and a number of studies in the banking industry have sought to employ measures of entry barriers as well as to explain bank competitive conduct and performance. Here too, the banking industry offers advantages in assessing the role of such measures, because differences in regulatory restrictions on entry can be exploited to assess their relationship to bank competitive conduct both cross-sectionally and over time (see Hannan and Prager, 1998 and Jayaratne and Strahan, 1998 for the US, and Berger, 2007 and Berger, et al., 2004b for international studies of regulatory entry barriers).

As in the case of SCP studies in general, SCP studies of the banking industry have been criticized because of the potential endogeneity of the measures of market structure that they typically employ. This is particularly true in the case of most of the earlier SCP studies, which sought to relate measures of bank rates of return to measures of market concentration. A well-known critique of this type of study was leveled originally by Demsetz (1973). His argument is that the frequently observed positive relationship between rates of return and market concentration may not reflect the exercise of market power in more concentrated markets, but may instead be the result of differentials in efficiency among market participants. If more efficient firms come to dominate the markets in which they operate as a result of their greater efficiency, then one may observe both greater profitability and greater concentration in such markets. This positive relationship, however, would not be due to a causal relationship running from market concentration to profitability, as proponents of the SCP paradigm assert.

One response to this criticism, as it applies to the banking industry, was to obtain measures of firm differences in efficiency and use them to test the differing

implications of the SCP paradigm and the 'efficient-structure hypothesis', as Demsetz's argument has come to be known. Berger (1995) found support for the efficient structure hypothesis, and also some for the relative market hypothesis (which controls for efficiency) (see also Berger and Hannan, 1997). A more common response, however, was to focus on the relationships between bank prices and concentration, rather than bank profits and concentration. Given that the geographic markets relevant to many types of bank loans and deposits are thought to be local in nature, cross-sectional studies of the relationship between bank loan rates or deposit rates and market concentration are easily conducted. Berger and Hannan (1989) reported such a study using survey data on bank deposit rates, while Hannan (1991) reported such a study using survey data on bank commercial loan rates. Many of these studies had been conducted by the mid-1990s. For the most part, they reported positive relationships between loan rates and market concentration and negative relationships between deposit rates and concentration, consistent with predictions of the SCP paradigm (see reviews by Berger, Demirgüç-Kunt, et al., 2004b and Degryse and Ongena, 2008).

While such findings may not be as vulnerable to the Demsetz critique, as traditionally interpreted, criticisms of these 'price-concentration' studies have been registered. Evans, Freob, and Werden (1993) point out that in these studies as well, measures of market concentration cannot be considered exogenous, both because of a possible feedback effect from prices to concentration and because of difficult-to-measure aspects of demand, factor prices, and basic conditions that affect both price and the measure of concentration. They note that both of these sources of endogeneity are likely to bias results in the negative direction (i.e., against the finding of a positive relationship between price and concentration), as their case study of pricing in the airline industry shows. Nonetheless, they note that positive biases are possible, and the endogeneity of concentration, whatever biases it imparts, implies that price-concentration studies cannot be used to predict the changes in prices that would result from a merger.

NEIO approaches to measuring market power

Over time, researchers developed new methods as a way to counter the problems associated with the SCP. These came to be known under the name of New Empirical Industrial Organization.[3] They reject the more traditional measures of market performance and seek to estimate market power using more flexible models of profit-maximizing firm behavior (relative to Cournot competition, which has been used to derive the HHI).

[3] Bresnahan (1989) provides a review of the old and NEIO methods.

The Panzar-Rosse approach

The Panzar-Rosse (1987) approach exploits the fact that changes in input prices should affect revenues differently, depending on the degree of competition. In particular, the approach relies on the relationship between input prices (or marginal costs) and equilibrium gross revenue, derived from the theory of the firm under alternate assumptions about competitive conditions. Let reduced-form revenue be represented as:

$$R = R^*(W_1, \ldots W_k, Z, Y),$$

where W_i represents the price of input i, and Z and Y denote vectors of exogenous cost and demand shifters, respectively. The test statistic proposed by Panzar and Rosse is

$$H_R = \sum_i W_i(\partial R/\partial W_i)/R^*(.),$$

or the sum of the elasticities of the reduced-form revenue with respect to each factor price. Monopoly theory implies that the revenue of a pure monopolist must fall as marginal cost rises (since it operates where demand is elastic), implying that under pure monopoly, the sum of the revenue elasticities is negative such that $H_R < 0$. Thus, failure to find that $H_R < 0$ may be taken as a rejection of the hypothesis that the firm behaves as a monopolist. Another hypothesis that may be tested with the statistic is long-run competitive equilibrium, in which case $H_R = 1$. The reason is that increases in input prices lead to proportional increases in both marginal costs and revenues. Potential shortcomings of this approach are that pure competition yields $H_R = 1$ only in long-run equilibrium (not in short-run equilibrium or in disequilibrium) and $H_R < 1$ can arise under scenarios other than monopoly. Other problems may include the inability to account accurately for all input prices and the necessity of dealing with the identification problem by finding appropriate exogenous shifters.

Applying the Panzar-Rosse approach to a cross-section of banks in New York state, Shaffer (1982) reported that both the hypothesis of monopoly and the hypothesis of long-run perfect competition could be rejected. This seems a reasonable finding, since in the case of the banking industry, we are unlikely to find the extremes of monopoly or perfect competition, but rather something in between. Other studies using the Panzar-Rosse methodology soon followed. In their study of Canadian banks, Nathan and Neave (1989) reported that they could reject the hypothesis of monopoly power. Studies applying the methodology to European banking systems generally reject both monopoly and perfect competition in favor of monopolistic competition (with an H value between 0 and 1) (see, e.g., Molyneux, Lloyd-Williams, and Thornton, 1994; Bikker and Groeneveld, 2000; and de Bandt and Davis, 2000).

For Japan, however, Molyneux, Thornton, and Lloyd-Williams (1996) were not able to reject monopoly behavior. In a recent application, Claessens and Laeven (2004) employ the methodology to obtain estimates of competitiveness for banks in fifty different countries and then relate this competitiveness measure to indicators of the country's banking structure and regulatory regime. They find that banks in countries with greater foreign bank entry and fewer restrictions on entry and services tend to behave more competitively, but they fail to find a negative relationship between banking concentration in the country and this measure of competitiveness (see also Goddard and Wilson, forthcoming, who account for the mispecification bias in the Panzar-Rosse approach when adjustments toward the new market equilibrium take time).

An issue relevant to all these studies concerns the geographic extent of the true market for banking services. Studies of the US typically use local market definitions (Metropolitan Statistical Area (MSA) or rural county), following antitrust policy, while studies of other nations typically view the entire country as the market. If the relevant market is smaller in geographic scope than the country, as is clearly the case in the US and probably many other countries as well, then a country-wide measure of competitiveness would represent an average of values that could differ widely across local areas within the country. This would also imply that country-wide measures of market structure do not accurately measure the market structure that is relevant to the behavior observed.

Conjectural variations approach

Another procedure, proposed by Bresnahan (1982) and Lau (1982), has been used in a number of studies of the banking industry. It focuses on the fact that the first-order condition for profit maximization can be expressed as:

$$P = mc(Q,Z) - D_1(Q,Y)Q\lambda,$$

where P and Q represent price and quantity, respectively; D_1 denotes the derivative of the inverse demand relation with respect to Q; Z and Y denote cost and demand shifters, respectively, and λ is a conduct parameter, which takes on the value of zero in the case of perfect competition (in which case $P = mc$) and the value of one in the case of monopoly. This parameter can also be expressed in terms of the conjectures that firms have regarding the response of rivals to a change in price or quantity (hence the 'conjectural variations' terminology). Intermediate values of λ correspond to various degrees of imperfect competition. Bresnahan (1982) and Lau (1982) note that, together with estimation of the demand relation, estimation of the first-order condition above can identify the conduct parameter λ under certain conditions.

Shaffer (1989; 1993) applied this technique to both the US and Canadian banking industries, strongly rejecting collusion, but not perfect competition, in both cases. Berg and Kim (1998) also used this procedure to compare results obtained for

retail vs. corporate banking markets for the Norwegian banking industry, and, in a further study, Shaffer (2001) used the analysis to assess competition in fifteen countries in Europe, North America, and Asia, finding varying results for the different countries.

A potential shortcoming inherent in the use of the procedure is that it requires that the researcher know the nature of the demand relationships and that the functional form be of a certain type. Further, accurate determination of the geographic scope of the market would seem to be of particular importance with the use of this procedure, based as it is on the estimation of market demand relations. If the geographic scope of the relevant market is smaller than that of the country, then one is left to wonder how to interpret the results when the whole country is assumed to be the relevant market.

Direct measurement of the pricing effects of mergers

A few studies have sought to measure directly the impact of mergers on prices. In essence, they compare the before-and-after change in the prices of merger participants with the same-period change in the prices of equivalent firms that are not affected by the merger. The local nature of markets for some banking products makes this kind of study more feasible for the banking industry than for other industries. Prager and Hannan (1998), the first such study conducted for the banking industry, found that deposit rates offered by participants in substantial horizontal mergers (and by their local market rivals) declined by a greater percentage than did deposit rates offered by banks not operating in markets in which such mergers took place. This finding of adverse price effects from mergers conforms to results reported in similar studies of other industries (see Kim and Singal, 1993 for the case of airlines). Focarelli and Panetta (2003) note, however, that these studies look only at the short-run pricing impact of consolidation, ignoring effects that take longer to materialize. Using detailed data on the deposit rates offered by Italian banks over several years, they found strong evidence that, although consolidation does generate adverse price changes in the short run, efficiency gains dominate over market power effects in the longer run, leading to more favorable prices for consumers.

Newer developments: Structural estimation in differentiated product markets and other recent studies

More recently, banking researchers have adopted new and creative methodologies that take advantage of certain developments in the industrial organization literature. By far the most significant introduction has been structural demand estimation, though other innovations include accounting for sunk costs and entry decisions.

Structural estimation under differentiated products

Much work has been reported in recent years on methodologies to measure market power that rely on the estimation of structural models of demand under the more realistic assumption (for most industries) of differentiated products. These are potentially important, given that, once demand parameters are known, estimates of own-price elasticities and cross-price elasticities can be used to determine markups and the effects from changes in policy or the market environment. These models, based on microeconomic foundations, are potentially quite useful in understanding market outcomes. Yet, their estimation does not require detailed micro data, unlike earlier demand models presented in the literature, but can be based on observable aggregate price and quantity data. These models allow for counterfactual exercises designed to address policy questions, such as those involving the effects of allowing a merger or removing a ban on ATM surcharges. Most pertinent to antitrust practice, estimation of demand is central to the determination of the relevant market, which is required in the evaluation of mergers among firms operating in the same market.

Given the role of the demand elasticity in understanding market outcomes and addressing policy questions, several methods have been developed to estimate demand in the case of differentiated products. These new methodologies solve the dimensionality problem inherent in earlier approaches, where a firm's output was usually estimated as a function of its price and the price of other products. The latter not only requires instruments for the identification of the price coefficients, but quickly becomes infeasible as the number of products grows: with N products, there are N own-price elasticities plus N^2 cross-price elasticity coefficients to estimate. Two main approaches have been developed in response. One method defines consumer preferences directly over the products they purchase but simplifies estimation by separating products in a market into subgroups. This is the case of the multi-stage budgeting approach developed by Hausman, Leonard, and Zona (1992; 1994). Here, the researcher separates the demand estimation problem into three stages. At the highest level, consumers are viewed as choosing how much of their budget they wish to allocate to a general type of product (e.g., cars). At the next stage, the consumer decides how much of her budget she will divide among different categories of the product (e.g., fuel-efficient cars, sports cars, luxury cars). At the final stage, the consumer allocates the budget for a category among the products in that category (e.g., car models within each category). The methodology does restrict some cross-price elasticities, but it permits flexible cross-price elasticities for products within a particular product category. For example, one can estimate a flexible demand system for luxury cars. Changes in prices in other categories (e.g., fuel-efficient) will still affect the demand for luxury cars, but only indirectly through their effect on overall luxury car spending. This method is not particularly well suited for applications where the number of products

and/or firms is large, and perhaps explains why there are no existing applications in banking, where usually many banks exist in a given market.

Another method, which has been applied to study competition in banking, is that of discrete choice models, which define consumer preferences over the characteristics of these products. The seminal work by Berry, Levinsohn, and Pakes (1995) (known as 'BLP') provides the framework to structurally estimate demand and supply using aggregate price and quantity data, and has become the reference for this generation of models. Their machinery introduced more reasonable demand elasticities and corrected the endogeneity problems present in earlier methods. While one might immediately think of instrumental variables in such estimation, the methodology proposed by the authors does not require per se new data, but imposes instead moment conditions as a way to obtain the demand parameters. Clearly, there is a cost to imposing such specification structure, but the benefit is in that no new, and perhaps hard to obtain, data are required. Several developments followed, and the methodology has been applied to industries that range from breakfast cereal to cable TV. For instance, Nevo (2001) estimates a demand system for ready-to-eat breakfast cereal in order to determine the effects of mergers in the industry on market power and market outcomes.

In more detail, these studies estimate demand by first specifying the consumer's utility function, which reflects preferences over product characteristics. Letting the K-vector of product attributes be x_j, the conditional indirect utility of consumer i from choosing product j at price p_j is

$$U(\zeta_i,p_j,x_j,\xi_j;\theta_D) \equiv \delta_j + \sum_k \sigma_k x_{jk} \nu_{ik} + \epsilon_{ij} \equiv x_j\beta - p_j\alpha + \xi_j + \sum_k \sigma_k x_{jk} \nu_{ik} + \epsilon_{ij}$$

where ξ_j represents product characteristics unobserved by the econometrician (such as quality), ν_{ik} are consumer characteristics, and ϵ_{ij} is an i.i.d. random disturbance. Note that δ_j only depends on product characteristics, while the following term introduces interactions between consumer characteristics and product characteristics, allowing individuals to have different marginal utilities from product characteristics. This is a flexible specification known as the random coefficients model. The interaction terms allow for reasonable substitution patterns, as consumers, faced with an increase in a given product's price, will substitute toward similar products. Given a density $f(\zeta)$ for consumer characteristics ζ, one can derive a firm's market share to be:

$$s_j(p,x,\xi;\theta_D) = \int_{\zeta \in Aj} f(\tilde{\zeta}) d\tilde{\zeta}$$

where $A_j = \{\zeta : U(\zeta, p_j, x_j, \xi_j; \theta_D) \geq U(\zeta, p_r, x_r, \xi_r, ; \theta_D)$ for $r = 0, 1, \ldots, J\}$, the set of values of ζ in the population that induce choice of good j. Letting S_j represent the observed market share for firm j, with s_j representing the market share predicted by

the model, and stacking all products into a vector, the following holds exactly at the true values of δ: $S = s(\delta)$. This system of equations can be solved either analytically, such as in the logit case (where no interactions between product and consumer characteristics are allowed), or numerically, such as is necessary in the random coefficients model. The details of the estimating procedure can be found in Berry, Levinsohn, and Pakes (1995); Berry (1994); Nevo (2000); and Nevo (2001).

For the banking industry, Dick (2008) was the first to structurally estimate demand for deposits as a way to determine the impact on consumer welfare from the significant changes in banking markets that followed deregulation of branching networks throughout the 1990s in the US. Her findings suggest that depositors' welfare was generally unaffected, or even increased, despite the significant structural changes throughout the period, especially as banks increased their branch networks. Along a similar vein, Adams, Brevoort, and Kiser (2007) estimate deposit demand for banks as well as thrifts in order to determine whether they are close substitutes, an important question for antitrust regulation. Their findings suggest that banks operate in separate markets from thrifts, thereby providing guidance to regulators as to whether it is appropriate to cluster these deposits in the calculation of concentration measures for merger analysis.

Using structural estimation, recent papers have also recognized the importance of branch networks and other significant fixed costs in affecting competitive conduct in the banking industry. Indeed, deviations from marginal cost pricing are not always linked in a clear manner to market power and welfare losses. In industries where fixed costs are large, such as banking, we really want to know whether prices persistently exceed average costs. In her exploration of ATM networks, Ishii (2005) estimates a structural model of deposit demand and bank behavior in order to determine the effects of surcharges—fees charged to unaffiliated customers—on demand, ATM investment, and competition. The analysis is useful in understanding the trade-off between competition and investment incentives in the industry: surcharges might lead firms to invest in ATMs, but it might also reduce competition and lead to overinvestment. Along a similar vein, Knittel and Stango (2008) estimate a deposit demand to determine the effects of ATM-fee induced incompatibility, finding a similar trade-off between competition and ATM deployment. Most recently, Degryse, Ferrari, and Verboven (forthcoming) develop and estimate a model of ATM investment and consumer demand for cash, and find that banks underinvest in ATMs under a system of no cash withdrawal fees.

Sunk costs and entry models

The presence of fixed costs can affect the degree of barriers to entry in the industry. The work of Sutton (1991) provides a theoretical framework to explain how these barriers might arise endogenously, leading to concentrated market structures. Central to the theory is the interplay of exogenous (set-up costs) and endogenous sunk costs (advertising or quality investments that are fixed with respect to output

and chosen by the firm). When endogenous sunk costs are large relative to the set-up costs of the firm, the theory predicts that markets will remain concentrated and quality investments per firm will increase as market size grows.

Most recently, banking economists have analyzed the role of endogenous fixed-cost investments on equilibrium banking industry structure, with findings suggesting that quality investments matter in the structure of the industry, and may raise barriers to entry. In particular, Dick (2007) finds that the leading banks in a market make larger investments in branch networks and that these investments grow with market population. An implication of this finding is that when profit opportunities arise in a market, such as those created by an inflow of new customers, large banks are likely to open new branches as a way to take up the additional demand and prevent further entry. Along similar lines, Cohen and Mazzeo (2007a) find that branch networks are larger whenever market conditions are attractive for multi-market bank entry, suggesting that incumbent firms that anticipate entry by these banks add branches to their network to deter entry. Consistent with these findings on entry barriers, Berger and Dick (2007) find that the order of entry into local markets matters, such that early bank entrants enjoy permanently larger market shares, even accounting for survivorship bias.

Structural entry models, by exploiting the equilibrium entry conditions as they relate to fixed costs, provide an approach to assess the extent to which market power is exercised. This methodology is related to the work first proposed by Bresnahan and Reiss (1987). Their approach addresses the question of how abruptly prices decline from monopoly levels as additional firms are added to the relevant market. Since accurate data on prices and marginal costs are not typically available, their approach exploits the fact that the relationship between price and the number of firms can be inferred by examining how much larger a market must be to accommodate an additional firm. If pricing becomes more competitive as more firms enter the market, then firms will enjoy lower variable profits as a result. This means that the size of the market per firm when an additional firm is in the market (referrred to as the entry threshold) must increase in order to generate enough variable profits to cover fixed costs (assuming fixed costs per firm remain constant with the number of firms). Thus, by estimating entry thresholds, which use mostly cross-sectional variations on market size and number of firms, one may infer the effect of entry on prices and variable profits, and therefore the degree of competition in the market.

Cetorelli (2002), the first to apply this approach to banking, uses data on over 2,000 rural counties and finds that US local banking markets approach competitive levels quickly, with the biggest increase in market size occurring to accommodate the third or fourth firm in the market. Most recently, Cohen and Mazzeo (2007b) estimate a structural model of entry to study the effect of market segmentation on competition, focusing on the differences among multi-market banks, single-market banks, and thrift institutions. Their findings suggest that competition among depository institutions of the same type is greater than competition among different

types. Thus, a merger between a multi-market bank and a single-market bank, for instance, should have less of an effect on market competition than a merger between two multi-market banks.

Structural models and antitrust practice

Estimates from structural models are particularly relevant to antitrust policy. Coupled with supply side assumptions, one can use these estimates to predict markups under various competition models, which can then be compared to observed markups. Importantly, the estimates can be used to predict price changes that would result from hypothesized mergers. Antitrust activity has indeed centered on merger analysis, and as such, the possibility of carrying appropriate counterfactual exercises could be quite important in determining whether a merger is appropriate. As discussed in more detail below, antitrust authorities have relied on HHI, a simple measure of market concentration, in order to determine whether a proposed merger raises anticompetitive concerns. In particular, regulators have used the 'HHI rule', such that if the merger increases concentration by a 'significantly large amount', the proposed merger has to undergo further scrutiny. The nature of the latter varies dramatically, but it usually involves gathering additional data and statistical analysis of diverse rigor. With a structural model, one can obtain the pre- and post-merger markups, given estimates of own- and cross-price elasticities. Pre-merger markups for each firm are obtained from the markup equation, derived from estimates of the demand elasticities, while post-merger markups can be obtained by assuming that the merged firm will maximize profits by 'internalizing' the externality imposed by the other firm offering a substitute product (how close of a substitute depends on the degree of differentiation between banks) when it itself tried to raise its price. Clearly, a realistic estimate of demand elasticity is important in obtaining meaningful information about the potential effects of the merger, and structural demand models are capable of offering the most accurate parameter estimates.

There are, however, a number of challenges that a widespread application of these methodologies to bank merger analysis must face. Unlike cars and breakfast cereal, defining what a banking product is can be a challenge, given their multiplicity and complexity and the possibility of synergies between banking products at a given bank. Banking transactions are dynamic in nature, and there are usually non-negligible switching costs in going from one bank to the next. In addition, defining the relevant geographic market and the relevant consumer is far from straightforward and has been the source of debate in the banking literature. For instance, while some consumers may care to transact with a single branch from their bank, others may want to use their bank where they live and where they work.

Another challenge is the time and resources required by such analyses. The evaluation of prospective mergers usually occurs under significant time constraints, and therefore estimating a structural model might not be a viable alternative in many cases. Nevertheless, many of the techniques are by now

becoming standard and should over time become easily available as they are added to statistical software, especially for the more complex models providing greater flexibility and robustness.

Another issue concerns the need for any antitrust policy to be transparent to firms whose actions may subject them to antitrust enforcement. Transparency is desirable because it reduces the likelihood that regulatory and firm resources are unnecessarily devoted to mergers applications that are later denied. Adoption of the new methodologies to assess merger applications would undoubtedly introduce more uncertainty as to the ultimate regulatory decision, but this problem too could diminish over time as the techniques become less costly to use.

A more fundamental issue regarding these methodologies concerns the sensitivity of results to the fairly detailed assumptions regarding the nature of the demand system to be estimated, its stability over time, and the nature of competitive conduct both before and after the merger. As DiGiacomo (2007) shows, results can be quite sensitive to assumptions regarding the nature of the demand system. While no study has sought to compare the results of such simulations with observed merger-induced price changes in the banking industry, at least two studies have undertaken such comparisons in other industries. Peters (2006), who conducted such an analysis for five airline mergers, concludes that these methods, when based on standard assumptions, do not generally provide an accurate forecast of post merger prices. He suggests that deviations from standard assumptions of firm conduct play an important role in accounting for the differences between the predicted and observed price changes. Weinberg (2007), who conducted such an analysis for two mergers (one involving passenger car motor oil and the other breakfast syrup), reaches similar conclusions. He also reports that the assumption of stability of the demand system before and after these mergers could be rejected.

Despite these challenges, structural models may in the future prove to be quite useful additions to the tools available to antitrust authorities, particularly with further research and refinements. In this regard, it is important to keep in mind that the real question is not whether the structural model provides perfect estimates, but rather whether one can improve upon current practice by using these new methodologies, especially as the cost of adopting them decreases over time and refinements are introduced.

Antitrust policy

Antitrust policy, as currently practiced, has a much longer history in the US than in other nations. Whereas US antitrust laws were first introduced at the end of the

nineteenth century, such laws (usually referred to as 'competition' laws outside the US) did not appear in most other countries until the 1960s or later. Although there are some important differences across countries in antitrust laws and policies, overall, the similarities far outweigh the differences. For this reason, most of our discussion will focus on antitrust policy in the US; we end with a brief discussion of antitrust policy as it currently exists in the European Union.

Antitrust policy in the US

In the banking industry, antitrust regulation has mostly centered on horizontal mergers (or mergers among firms competing in the same market). The general aim has been to ensure that the industry remains competitive, as opposed to a direct regulatory approach. US antitrust policy is largely governed by three major statutes: the Sherman Antitrust Act of 1890, the Clayton Act of 1914, and the Federal Trade Commission Act of 1914. The Sherman Act prohibits (i) contracts, combinations, and conspiracies that restrain trade or commerce; and (ii) actions that have the effect or intent of creating or maintaining market power. The Clayton Act outlaws specific practices, such as tying and exclusive dealing agreements, that adversely affect competition, as well as mergers that would substantially lessen competition. The Federal Trade Commission Act establishes the Federal Trade Commission as an antitrust enforcement agency and outlaws unfair methods of competition. Although antitrust policy has many dimensions, this discussion will focus on antitrust policy toward horizontal mergers given its importance to actual antitrust practice.

Evaluation of horizontal mergers

In the US, the Federal Trade Commission (FTC) and the Department of Justice (DOJ) are both involved in administering the antitrust laws. The FTC and the DOJ have issued *Merger Guidelines* that outline the method they apply in evaluating horizontal mergers. This method involves (i) defining the relevant geographic and product markets; (ii) calculating the structural effects of the merger in each of these markets using the HHI, and using the resulting measures to determine whether the merger 'potentially raises significant competitive concerns'; and (iii) if the merger does potentially raise competitive concerns, considering other market factors such as entry conditions, as well as possible efficiency gains associated with the merger. The guidelines are intended to reduce uncertainty surrounding the enforcement of antitrust policy with respect to horizontal mergers.

In the case of the banking industry, a bank merger or acquisition is subject to antitrust review by both the relevant federal bank regulator and the DOJ. The bank regulators' responsibility for assessing the competitive effects of proposed mergers stems from the Bank Merger Act of 1960 and the Bank Holding Company Act of 1956. These two acts, as amended by Congress in 1966, require bank regulators to apply the standards of the Clayton Act in assessing the competitive effects of proposed bank mergers and acquisitions. Among the bank regulators, the Federal

Reserve Board has been the most actively involved in antitrust enforcement. Because of the extremely large number of bank mergers occurring each year, the bank regulators and the DOJ have developed streamlined approaches for defining the relevant geographic and product markets.

Market definition

In general, market definition is often the most difficult step in the process. The regulators begin by asking whether a hypothetical monopolist would raise prices from pre-merger levels. According to the *Guidelines*:

> A market is defined as a product or group of products and a geographic area in which it is produced or sold such that a hypothetical profit-maximizing firm, not subject to price regulation, that was the only present and future producer or seller of those products in that area likely would impose at least a 'small but significant and non-transitory' increase in price, assuming the terms of sale of all other products are held constant. A relevant market is a group of products and a geographic area that is no bigger than necessary to satisfy this test.[4]

In practice, a 'small but significant and non-transitory' increase in price (SSNIP) is typically interpreted as a 5 percent increase lasting for at least one year. Implementing the so-called 'SSNIP test' requires econometric estimation of a demand system for the set of products being considered as potentially comprising a relevant market. The demand estimates can then be used, together with information about the merging firms' cost functions and the current prices of all of the products within the set, to determine whether a monopolist of that set of products could profitably raise its price by 5 percent. If the answer is no, then a broader set of products is considered by introducing the next closest substitute; if the answer is yes, then the set of products is deemed to comprise a market. Although this approach sounds fairly straightforward, in practice it can be quite complicated, especially when the number of products included in the set of products being considered becomes large.

For the banking industry, the determination of the product and geographic market for banking services is complicated by the fact that banks typically provide a wide variety of services to many different types of customers. The scope of the geographic market may vary considerably across different services and customer types. For example, the market for loans to large national or multinational corporations is almost certainly national or international in scope—these customers can, and often do, obtain their banking services from providers located anywhere around the world. In contrast, the market for providing retail banking services, such as transactions accounts and small loans to households or small businesses, is

[4] See section 1.0 <http://www.usdoj.gov/atr/hmerger/11251.htm>.

probably much smaller in geographic area—these customers will typically seek and obtain these types of services from providers within very close proximity. US antitrust authorities tend to focus their competitive analyses on retail banking services. The rationale for this focus is that a proposed bank merger that does not raise competitive concerns in the provision of retail services is not likely to raise such concerns in the provision of other products for which the geographic market is larger, encompassing more actual or potential competitors.

The geographic markets for these retail services are defined as economically integrated local areas, which frequently, though not necessarily, correspond to a metropolitan area, a single rural county, or two or more contiguous rural counties. Recently, some researchers have argued that, in an era of Internet banking and of large, geographically diversified banking organizations, the local market for retail banking services has become obsolete (see Radecki, 1998). However, evidence from a variety of large surveys still provide support for the existence of local retail banking markets (see Amel and Starr-McCluer, 2002; Kwast, Starr-McCluer, and Wolken, 1997; and Elliehausen and Wolken, 1990).

In terms of the product market, the antitrust authorities and the courts have traditionally defined the relevant product market for banking antitrust analysis to consist of the cluster of products and services provided by a commercial bank. The courts have explicitly rejected the idea that each product line (e.g., transactions accounts, savings accounts, mortgage loans) comprises a separate product market. Total deposits within the relevant geographic market are used as a proxy for a bank's capacity to provide the cluster of commercial banking products within that geographic area. Reliance on total deposits to measure the overall scale of a bank's activity in a local area and its ability to compete in the various product lines offered by commercial banks is justified on grounds that (i) deposits make up the bulk of a bank's liabilities, and (ii) funds are fungible across different types of assets. Furthermore, in the US, deposits are the only product for which data are collected at the bank branch level, so they are the only product for which the antitrust authorities have access to local-level data for all competitors. The survey evidence on household and small business behavior also seems to suggest that these consumers cluster their purchases, at least for deposits services, within a single depository institution (Amel and Starr-McCluer, 2002; Elliehausen and Wolken, 1990; and Kwast, Starr-McCluer, and Wolken, 1997).

Market structure and the HHI rule

Once a market has been defined, market shares and the structural effects of the proposed merger can be computed. With the HHI calculated as the sum of squared market shares (expressed in percentages, not decimals), the US *Merger Guidelines* categorize a market as unconcentrated (HHI below 1,000), moderately concentrated (HHI between 1,000 and 1,800), or highly concentrated (HHI above 1,800).

Mergers are generally considered unlikely to have any adverse effects on competition unless they would produce on a pro forma basis an increase in the HHI of more than 100 points in a market that would be moderately concentrated after the merger, or an increase of more than 50 points in a market that would be highly concentrated after the merger. If these thresholds are exceeded, the proposed transaction receives further review to assess its likely competitive effects.

For the banking industry, things are slightly different, perhaps due to the fact that banking services are assumed to have relatively more substitutes. A proposed bank merger will typically receive further scrutiny only if it would increase the HHI in one or more local banking markets by more than 200 points to a post-merger level above 1,800. Given the geographic and product market definitions discussed above, the group of competitors considered in constructing structural measures for proposed bank mergers and acquisitions typically includes all commercial banks and thrifts with a branch presence in the local market area. Non-bank providers of financial services (e.g., credit unions, finance companies, etc.) are explicitly excluded from the structural calculations because they do not offer the full range of commercial banking products. However, their presence is indirectly accounted for by applying less stringent structural guidelines to banking than to other industries.

Mitigating factors

If a proposed bank merger exceeds the 200/1,800 threshold in a local banking market, the antitrust authorities look more closely at the competitive conditions in that market. One crucial element is the likelihood of future entry. If mergers raise market power, one might expect new entry after the merger, in which case, provided no cost synergies are realized, the merger might not be profitable (see Berger, et al., 2004a, for instance, who find evidence that mergers result in 'de novo' bank entry). However, the prevalence of sunk costs in the banking industry might raise barriers to entry and provide the incentive for firms to merge. Potential entry is taken into account by considering both recent past entry and market characteristics (e.g., population growth) that affect the attractiveness for future entry. Consideration is also given to the presence of credit unions, particularly those that are open to most or all residents of the market and have easily accessible branches. If the proposed transaction raises significant competitive concerns when all of the relevant information is taken into account, the applicant often proposes a divestiture that alleviates those concerns. As a result, few bank merger applications have been denied by the US antitrust authorities in recent years.

Other issues regarding antitrust analysis

One might note at this point the omission of any consideration of the potential efficiency gains of the merger. The latter are believed to be difficult to assess

in general, but are particularly so for the antitrust authority, which has less information than the firm in this respect. Moreover, banks may have an incentive to exaggerate them. Thus, with no cost-side considerations, the natural premise is that prices are likely to rise following a substantial merger, as firms internalize their aggressive behavior when competing with each other.

Another important omission is the lack of consideration of the other important strategic decisions of the firm regarding non-price attributes that might well change with the merger, such as the branch network. Even if we were able to predict correctly price changes following a horizontal merger, the analysis might not account adequately for other aspects of competition related to non-price variables, such as the branch and the ATM networks and the development of new products.

Emergency procedures

As of this writing, the financial crisis that began in August 2007 has resulted in the acquisition of several large and weakened depository institutions by other depository institutions. Examples are the acquisition of Countrywide Financial by Bank of America, the acquisition of Wachovia by Wells Fargo, and the acquisition of Washington Mutual by JPMorgan Chase. In this crisis, large banking institutions have also acquired troubled non-bank institutions, such as the acquisition of Merrill Lynch by Bank of America and that of Bear Stearns by JPMorgan Chase.

While the latter type of acquisition rarely poses antitrust issues, the acquisition of one depository institution by another certainly may. A formal declaration by bank regulatory authorities that a firm is failing can trigger a weaker set of criteria to be applied to the competitive analysis described above. In the cases noted above, however, no such declarations were made, and although procedures in some cases were expedited, the competitive analyses and the criteria applied to them were not altered as a result of the crisis.

Antitrust policy in the European Union

Antitrust review of horizontal mergers in the European Union began in the early 1990s, although enforcement in individual member countries predated the unified policy.[5] Current European Union enforcement policy is based on the 2004 reforms that prohibit mergers that 'significantly impede' effective competition. European Union merger policy applies to all European Union firms that exceed certain size thresholds. For companies below these size thresholds, the individual countries undertake merger review. The European Commission is responsible for merger review and has established guidelines that are quite similar to those applied in the

[5] Some of the information describing the general approach to antitrust policy is taken from Whinston (2007).

US. Like the US guidelines, European Union guidelines include HHI thresholds beyond which the Commission *may* have concerns. Unlike the US guidelines, the European Union guidelines also include a market dominance test. Thus, a transaction that would lead to a post-merger market share greater than 40 percent is likely to receive antitrust scrutiny, even if it does not exceed the HHI guidelines. Antitrust policy in other countries (e.g., Canada, Australia, New Zealand) has also come to resemble that in the US in recent years, though significant cross-country differences in both processes and standards do exist.

Similar to the US regime, the relevant market must be defined in both product and geographic dimensions. Products are presumed to be sold in the same relevant market when competition from one product constrains the market behavior of sellers of another product. Similarly, the relevant geographic market is defined as the smallest geographic area within which competition is not substantially influenced by activities outside the defined area.

Given the definition of the markets for which an analysis of horizontal mergers is to be conducted, the question arises as to how that analysis might differ between the European Union and the US. Bergman, et al. (2007) examine this issue by modeling the enforcement decision under both regimes. They accomplish this by conducting probit estimations of the relationship between the decision to deny a proposed horizontal merger and various explanatory variables, such as the post-merger HHI, the market shares of the firms involved, an index indicating the ease of entry in the market, and other characteristics potentially relevant to the enforcement decision. In general, they find that characteristics suggested as important by economic theory play an important role in both regimes, with observed enforcement decisions affected by similar characteristics. While they do find some differences, they conclude that overall, it is not possible to characterize one regime as universally more aggressive than the other.

In the case of the banking industry, the European Commission on Competition oversees antitrust enforcement of financial services affecting the single market, thereby pertaining mostly to cross-border events such as mergers between banks in different European countries. For mergers and other antitrust activity within a country's banking system, it is usually the national antitrust authority that has the responsibility to enforce competition policy. Even in cross-border events, the national competition authorities might also intervene if in disagreement with the Commission. This has occurred a number of times in the past. Such opposition may have created legal uncertainty, which may affect the degree of economic dynamism in the sector. While the evolution of the banking industry in Europe has been similar to that of the US, concentration has increased mostly as a result of domestic consolidation, as opposed to large cross-border integration.

In recent times, the European Commission on Competition has stated explicitly its intention to take on a more proactive role in competition policy, in the manner of the US, to not only deal with current merger cases, but also identify potential

barriers to competition in the financial services industry. Moreover, it has started to pay significant attention to the retail sector and to carry out official inquiries in the area. The integration of retail markets has lagged that seen in the wholesale and corporate banking markets in Europe. This may stem from the fact that retail markets tend to have greater barriers to entry for new banking firms and high switching costs for account holders.

Antitrust regulation, bank competition, and fragility

The premise in antitrust analysis is that more competition is good for welfare, especially given its focus on consumers. Could banking, however, be special in this sense? Is more competition always good? These questions have become particularly pertinent following the financial crisis that began in 2007. Past banking deregulation and the ensuing increase in market contestability have played a role in the implementation of improved credit scoring technologies and the extension of credit to riskier and previously excluded borrowers (Dick and Lehnert, forthcoming). Is it possible that too much competition is problematic in banking, where externalities on risk taking may lead to systemic crises? Are banks with more market power more resilient to shocks? From a theoretical perspective, competition has an ambiguous effect on bank risk taking and financial fragility.[6] This is an important issue that should be addressed by future research. In defining the course forward, however, the design of new banking regulation must address bank risk taking and systemic risk, without compromising competition leading to innovation and access to better and more banking services by consumers.

Concluding remarks

Analyses of competition among banks have followed fairly closely the continuing innovations and developments that have dramatically changed the field of industrial organization over time. Starting with the simple concentration-profits study and progressing, more recently, to structural models and merger simulations, economists have applied nearly all these innovations and developments to the banking industry.

[6] For instance, in Keeley (1990), more competition leads banks to take on more risk as a bank's charter value, in the context of deposit insurance, decreases. By contrast, in Petersen and Rajan (1995) more bank market power increases the incentives of banks to invest in the relationship necessary to serve riskier customers.

Traditional antitrust analysis in the banking industry has focused on measures of market structure and how they would change as a result of a proposed merger, as well as on other factors. Recent developments, particularly the estimation of structural models and merger simulations based on them, offer a promising alternative to this traditional type of analysis. As discussed, however, effective use of these newer methodologies in antitrust analysis face numerous challenges. Continued efforts in applied theoretical research will be key to our understanding of the nature of competition in the banking industry and to the development of improved antitrust analysis.

References

Adams, R., Brevoort, K., and Kiser, E. (2007). 'Who Competes with Whom? The Case of Depository Institutions', *Journal of Industrial Economics*, 55: 141–67.

Amel, D. F. and Starr-McCluer, M. (2002). 'Market Definition in Banking: Recent Evidence', *The Antitrust Bulletin*, 47: 63–87.

Berg, S. A. and Kim, M. (1998). 'Banks as Multioutput Oligopolies: An Empirical Evaluation of The Retail and Corporate Banking Markets', *Journal of Money, Credit, and Banking*, 30: 135–53.

Berger, A. N. (1995). 'The Profit–Structure relationship in Banking: Tests of Market Power and Efficient Structure Hypotheses', *Journal of Money, Credit, and Banking*, 27: 404–31.

—— (2007). 'Obstacles to a Global Banking System: "Old Europe" versus "New Europe"', *Journal of Banking & Finance*, 31: 1955–73.

—— Bonime, S. D., Goldberg, L. G., and White, L. J. (2004a). 'The Dynamics of Market Entry: The Effects of Mergers and Acquisitions on Entry in the Banking Industry', *Journal of Business*, 77: 797–834.

—— Demirgüç-Kunt, A., Levine, R., and Haubrich, J. (2004b). 'Bank Concentration and Competition: An Evolution in the Making', *Journal of Money, Credit and Banking*, 36: 433–51.

—— and Dick, A. A. (2007). 'Entry into Banking Markets and the Early-Mover Advantage', *Journal of Money, Credit, and Banking*, 39: 775–807.

—— and Hannan, T. H. (1989). 'The Price-Concentration Relationship in Banking', *Review of Economics and Statistics*, 71: 291–9.

—— —— (1997). 'Using Efficiency Measures to Distinguish among Alternative Explanations of the Structure–Performance Relationship in Banking', *Managerial Finance*, 23: 6–31.

Bergman, M. A., Coate, M. B., Jakobsson, M., and Ulrick, S. W. (2007). 'Comparing Merger Policies: the European Union versus the United States', *Potmac Paper in Law and Economics*, 07–01.

Berry, S. (1994). 'Estimating Discrete Choice Models of Product Differentiation', *RAND Journal of Economics*, 25: 242–62.

—— Levinsohn, J., and Pakes, A. (1995). 'Automobile Prices in Equilibrium', *Econometrica*, 63: 895–947.

Bikker, J. A., and Groeneveld, J. M. (2000). 'Competition and Concentration in the EU Banking Industry', *Kredit und Kapital*, 33: 62–98.

Bresnahan, T. F. (1982). 'The Oligopoly Solution Concept Is Identified', *Economics Letters*, 10: 87–92.

—— (1989). 'Empirical Studies of Industries with Market power', in R. Schmalensee and R. Willig (eds.), *Handbook of Industrial Organization*, 1st edn, ch. 17, Amsterdam: Elsevier, 1011–57

—— and Reiss, P. C. (1990). 'Entry in Monopoly Markets', *Review of Economic Studies*, 57: 531–53.

Cetorelli, N. (2002). 'Entry and Competition in Highly Concentrated Banking Markets', *Economic Perspectives*, Federal Reserve Bank of Chicago, 18–27.

Claessens, S. and Laeven, L. (2004). 'What Drives Bank Competition? Some International Evidence', *Journal of Money, Credit, and Banking*, 36: 563–92.

Cohen, A. M., and Mazzeo, M. J. (2007a). 'Market Structure and Competition among Retail Depository Institutions', *Review of Economics and Statistics*, 89: 60–74.

—— —— (2007b). 'Investment Strategies and Market Structure: An Empirical Analysis of Bank Branching Decisions', Working Paper, < http://www.kellogg.northwestern.edu/faculty/mazzeo/htm/branches_cohen_mazzeo1.pdf>.

De Bandt, O., and Davis, E. P. (2000). 'Competition, Contestability and Market Structure in European Banking Sectors on the Eve of EMU', *Journal of Banking & Finance*, 24: 1045–66.

Degryse, H., Ferrari, S., and Verboven, F. (forthcoming). 'Investment and Usage of New Technologies: Evidence from a Shared ATM Network', *American Economic Review.*

—— and Ongena, S. (2008). 'Competition and Regulation in the Banking Sector: A Review of the Empirical Evidence on the Sources of Bank Rents', in Boot, A., Thakor, A. (eds) *Handbook of Corporate Finance, Financial Intermediation and Banking*, Amsterdam: Elsevier.

Demsetz, Harold (1973). 'Industry Structure, Market Rivalry, and Public Policy', *Journal of Law & Economics*, 16: 1–10.

Dick, A. A. (2007) 'Market Size, Service Quality and Competition in Banking', *Journal of Money, Credit, and Banking*, 39: 49–81.

—— (2008). 'Demand Estimation and Consumer Welfare in the Banking Industry', *Journal of Banking & Finance*, 32: 1661–76.

—— and Lehnert, A. (forthcoming). 'Personal Bankruptcy and Credit Market Competition', *Journal of Finance.*

DiGiacomo, M. (2007). 'What Can We Learn from Merger Simulations: An Application of Logit Models to the Yogurt Industry', Working Paper, University of Turin.

Elliehausen, G. E. and Wolken, J. D. (1990). 'Banking Markets and the Use of Financial Services by Small and Medium-Sized Businesses', *Federal Reserve Bulletin*, 801.

Evans, W. N., Froeb, L. M., and Werden, G. J (1993). 'Endogeneity in the Concentration–Price Relationship: Causes, Consequences, and Cures', *The Journal of Industrial Economics*, 41: 431–8.

Focarelli, D. and Panetta, F. (2003). 'Are Mergers Beneficial to Consumers? Evidence from the Market for Bank Deposits', *American Economic Review*, 93: 1152–72.

Goddard, J. and Wilson. J. O. S. (forthcoming). 'Competition in Banking: A Disequilibrium Approach', *Journal of Banking & Finance*, forthcoming.

Hannan, T. H. (1991). 'Bank Commercial Loan Markets and the Role of Market Structure', *Journal of Banking & Finance*, 15: 133–49.

—— and Prager, R. A. (1998). 'The Relaxation of Entry Barriers in the Banking Industry: An Empirical Investigation', *Journal of Financial Services Research*, 14: 171–88.

Hausman, J., Leonard, G., and Zona, J. D. (1992). 'A Proposed Method for Analyzing Competition among Differentiated Products', *Antitrust Law Journal*, 60.

——————(1994), 'Competitive Analysis with Differentiated Products', *Annales d'Économie et de Statistique*, 34: 159–80.

Ishii, J. (2005). ' Interconnection Pricing, Compatibility, and Investment in Network Industries: ATM Networks in the Banking Industry', Working Paper.

Jayaratne, J. and Strahan, P. E. (1998). 'Entry Restrictions, Industry Evolution and Dynamic Efficiency', *Journal of Law & Economics*, 41: 239–74.

Keeley, M. C. (1990). 'Deposit Insurance, Risk, and Market Power in Banking', *American Economic Review*, 80: 1183–200.

Kim, E. H., and Singal V. (1993). 'Mergers and Market Power: Evidence from the Airline Industry', *American Economic Review*, 83: 549–69.

Knittel, C. R. and Stango, V. R. (2008). 'Incompatibility, Product Attributes and Consumer Welfare: Evidence from ATMs', *Advances in Economic Analysis and Policy*, 8: 1.

Kwast, M. L., Starr-McCluer, M., and Wolken, J. D. (1997). 'Market Definition and the Analysis of Antitrust in Banking', *The Antitrust Bulletin*, 973.

Lau, L. (1982). 'On Identifying the Degree of Competitiveness from Industry Price and Output Data', *Economics Letters*, 10: 93–9.

Molyneux, P., Lloyd-Williams, D. M., and Thornton, J. (1994). 'Competitive Conditions in European Banking', *Journal of Banking & Finance*, 18: 445–59.

——Thornton, J., and Lloyd-Williams, D. M. (1996). 'Competition and Market Contestability in Japanese Commercial Banking" *Journal of Economics & Business*, 48: 33–45.

Nathan, A. and Neave, E. H. (1989). 'Competition and Contestability in Canada's Financial System: Empirical Results', *Canadian Journal of Economics*, 22: 576–94.

Nevo, A. (2000). 'Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry', *RAND Journal of Economics*, 31: 395–424.

—— (2001). 'Measuring Market Power in the Ready-to-Eat Cereal Industry'. *Econometrica*, 69: 307–42.

Panzar, J. C. and Rosse, J. N. (1987). 'Testing for Monopoly Equilibrium', *The Journal of Industrial Economics*, 35: 443–56.

Peters, C. (2006). 'Evaluating the Performance of Merger Simulation: Evidence from the U.S. Airline Industry', *Journal of Law & Economics*, 49: 627–49.

Petersen, M. A. and Rajan, R. G. (1995). 'The Effect of Credit Market Competition on Lending Relationships', *Quarterly Journal of Economics*, 110: 407–43.

Prager, R. A. and Hannan, T. H. (1998). 'Do Substantial Horizontal Mergers Generate Significant Price Effects? Evidence from the Banking Industry', *The Journal of Industrial Economics*, 46: 433–52.

Radecki, L. J.(1998). 'The Expanding Geographic Reach of Retail Banking Markets', *Federal Reserve Bank of New York Economic Policy Review*, 15–33.

Shaffer, S. (1982) 'A Nonstructural Test for Competition in Financial Markets', Federal Reserve Bank of Chicago, *Proceedings of a Conference on Bank Structure and Competition*, 225–43.

—— (1989). 'Competition in the U.S. Banking Industry', *Economics Letters*, 29: 321–23.

—— (1993). 'A Test of Competition in Canadian Banking', *Journal of Money, Credit and Banking*, 25: 49–61.

—— (2001). 'Banking Conduct before the European Single Banking License: A Cross-Country Comparison', *North American Journal of Economics and Finance*, 12: 79–104.

Sutton, J. (1991). *Sunk Cost and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*. Cambridge, MA: MIT Press.

Weinberg, M. (2007). 'An Evaluation of Merger Simulations', Working Paper, University of Georgia.

Weiss, L. W. (1989). 'A Review of Concentration-Price Studies in Banking', in L. W. Weiss (ed.), *Concentration and Price*, Cambridge, MA: MIT Press, 219–65.

Whinston, M. D. (2007). 'Antitrust Policy toward Horizontal Mergers', in M. Armstrong and R. Porter (eds.), *Handbook of Industrial Organization*, 3, Armsterdam: Elsevier.

CHAPTER 17

WORLD TRADE ORGANIZATION COMMITMENTS vs. REPORTED PRACTICES ON FOREIGN BANK ENTRY AND REGULATION

A CROSS-COUNTRY ANALYSIS

JAMES R. BARTH
JUAN A. MARCHETTI
DANIEL E. NOLLE
WANVIMOL SAWANGNGOENYUANG[1]

Introduction

The World Trade Organization's (WTO) General Agreement on Trade in Services (GATS) is the first and only multilateral trade agreement liberalizing international

[1] The views expressed in this chapter are the authors' alone and are not to be taken to represent those of the WTO or its Members, nor of the Office of the Comptroller of the Currency or the United States Treasury Department. The authors wish to thank Xuan Shen for excellent research assistance.

commerce in services. Under the auspices of GATS, WTO members negotiated in 1997 a permanent agreement on financial services. As of July 2007, 151 economies were covered by this new element of the international financial framework. In the area of financial services, the principal aim of the GATS is to encourage greater openness among WTO member countries to the provision of financial services from foreign entities. A particularly significant area of coverage for the financial services agreement is member countries' banking sectors.

Despite the emergence of an increasingly broad and deep literature on comparative banking systems, one issue that has received scant attention is the degree to which countries' commitments under the GATS financial services agreement match regulatory practices 'on the ground' in member countries. This study is the first to address this issue across a wide range of countries. It does so by using country-by-country data on banking regulation from Barth, Caprio, and Levine (2006), and adding to that a new and comprehensive cross-country dataset on the WTO financial services sector commitments.

The chapter is organized as follows. The second section provides a brief explanation of the GATS and the nature of WTO commitments. The third section describes the two companion parts of our unique data set: our 'WTO commitments' data consists of information on specific entry, permissible activities, and operations requirements applying to the banking sector; and our 'reported practices' data is a comparable set of cross-country information based on the World Bank's 2003 survey of banking supervisory authorities. We present an aggregate profile of the countries included in each set separately, highlighting in particular similarities and differences between developed and developing countries. The fourth section turns to a systematic comparison of the two data sets. We first identify significant discrepancies between countries' WTO and 'reported practices' postures, and then construct an index that allows us to gauge the overall degree of openness of countries to foreign bank entry, as reflected in each set of data. The 'WTO commitments' and 'reported practices' variants of this 'market openness' index are described in detail and, for each index variant separately, countries are ranked according to the degree of overall openness to banking entry and range of permissible activities. Subsequently, we compare country-by-country values for the WTO commitments and reported practices index values in order to address the question of whether countries' WTO commitments for the banking sector match their reported banking system practices. Finally, the fourth section develops and applies our 'degree of discrimination' index as a measure of the extent to which there is a regulations-related disadvantage under which foreign banks operate, post-entry, relative to domestic banks. The final section summarizes our findings, concluding with an observation about the ongoing relevance of this information for informing discussions about better coordination of cross-border banking supervision in the wake of the global financial crisis of 2008.

The General Agreement on Trade in Services under the World Trade Organization: What is it and how does it work?

The main objective of the GATS is to facilitate progressively higher levels of liberalization of trade in services, including financial services, between member countries. This objective is addressed through members' participation in periodic rounds of multilateral negotiations focusing on the elimination of barriers to trade among WTO members. The GATS applies to all governmental measures affecting trade in services. In principle, the GATS covers measures implemented at both the federal level and the subfederal (provincial or state) level, as well as regulatory measures under the auspices of delegated non- or quasi-governmental bodies such as securities and futures exchanges. Financial services are defined in the GATS as any service of a financial nature offered by a financial service supplier, including traditional banking activities (e.g., deposit taking and lending), all insurance and insurance-related services (e.g., direct insurance, reinsurance, insurance intermediation, and auxiliary insurance services), and other financial services (e.g., financial leasing, asset management, trading in securities, and financial advice).

Marchetti (2003) explains that the GATS has what he characterizes as 'three different layers' of obligations. The first layer consists of the general obligations that bind all WTO members regardless of whether they have agreed to undertake market access commitments for a particular sector. The most important of these obligations is the most-favored nation principle (MFN), which makes it mandatory for every WTO member to treat services and service suppliers of any other WTO member no less favorably than it treats similar services and service suppliers of any other country—that is, the MFN principle imposes the obligation not to discriminate among foreign services and service suppliers. Other general obligations include the publication of all measures of general application affecting the operation of the GATS, and the availability of legal remedies.

The second layer of obligations consists of the specific commitments by members on market access and national treatment to services and service suppliers of other WTO members. WTO members are free to choose which financial services will be subject to market access and national treatment disciplines. Members are permitted to qualify or impose limitations on their market access and national treatment commitments. For example, market access limitations may include limits on the number of suppliers allowed to enter the host market, and national treatment limitations may include measures such as higher income tax rates for (all) foreign suppliers. Importantly, WTO members may choose to retain 'full discretion' with

respect to the treatment of foreign firms, either by excluding a specific sector from the commitmetns on market access and national treatment, or by including the sector but leaving a specific mode of supply unbound (i.e., the level and degree of access and national treatment is not specified). Note that this does not necessarily (or even usually) signal a restrictive stance toward foreign access. Members are free to provide greater access and more favorable treatment to foreign firms than outlined in specific commitments.

The existence of specific commitments on market access and national treatment triggers a third layer of obligations, which apply only to sectors listed in a WTO member's schedule of commitments. These include the notification of new measures that have a significant impact on trade in the sector concerned; the reasonable, objective, and impartial administration of measures of general application; and the avoidance of restrictions on international payments and transfers for current international transactions and, eventually, on capital transactions.

Like any other trade agreement, the GATS contains exception provisions which allow WTO members to depart from their obligations or commitments under the agreement in very specific circumstances. One of those exception-type provisions is the so-called 'prudential carve-out', which allows WTO members to take measures for prudential reasons, including for the protection of investors, depositors, and policy holders, and for preserving the integrity and stability of the financial system. Members do not need to inscribe such exceptions for prudential reasons in their national schedules of commitments. However, such measures are not to be used as a means of avoiding a country's commitments or obligations under the GATS. WTO members are also allowed to introduce restrictions of a temporary nature in the event of serious balance of payments and external financial difficulties, subject to consultations with other members.

Empirical measures of the degree of openness in banking: World Trade Organization commitments data and World Bank reported practices data

A relatively small body of research has explored methods to quantify the degree of openness to foreign entry in services industries. Barth, Marchetti, Nolle, and Sawangngoenyuang (2006) provide a useful survey of that literature, which serves as background for the present study. The current study focuses on banking services in particular, and offers two major advancements over previous research. First, our

dataset represents a substantial addition to the information base for researchers and policymakers. In particular, we have compiled comprehensive information on 123 countries' banking sector GATS commitments under the World Trade Organization (our 'WTO commitments' data), and have paired that, on a country-by-country basis, with equally extensive World Bank data (our 'reported practices' data) on the regulatory regimes countries have in place. The second contribution is that our dataset allows us to examine the degree to which reported regulatory practices are consistent with the WTO commitments countries have made under the GATS.

WTO commitments data

Table 17.1 lists and briefly defines seventeen specific banking entry or permissible activities issues under the WTO that are included in the WTO commitments part

Table 17.1. Glossary for WTO commitments—entry into banking and permissible banking activities

Entry or Permissible Activity: Overall Concept	Relevant WTO Commitment	
	WTO Code	Specific Substance of WTO Component
Entry and licensing of banks	103	Commitment vs. Discretion ('full discretion' = no explicit WTO commitment made)
	104	Use of discretionary licensing or application of 'economic needs tests'
	105	National treatment for foreign banks
	106	Total number of foreign banks allowed to operate in the country
	107	New entry after a specified number/level of foreign bank penetration
Forms of entry	1.12.1	Entry via acquisition of an existing bank
	1.12.2	Entry via establishment of subsidiary
	1.12.3	Entry via branching
Minimum capital requirements	1.3.1	Minimum capital entry requirement similar for both foreign and domestic banks
Expansion of physical presence	105a	National treatment for expansion (once established in the country) via new offices, branches, ATMs
Composition of board of directors	105b	National treatment for requirements on composition of board of directors
Foreign banking concentration	109	Foreign bank share of total banking system assets
Permissible activities: securities	4.1.1	Banks engaging in securities underwriting
	4.1.2	Banks engaging in securities dealing and brokering
	4.1.3	Banks engaging in mutual funds activities
Permissible activities: insurance	110	Banks engaging in insurance activities

of our data set. These issues involve countries' commitments (or lack thereof) under the WTO on entry requirements (including whether the minimum-capital-entry requirement is similar for foreign and domestic entrants), methods of entry (i.e., via the acquisition of an existing bank in the host country, and/or the establishment of a subsidiary, and/or via branching); the expansion of banking presence after entry (specifically, whether 'National Treatment' is accorded to foreign entrants with respect to establishing additional offices, branches, and ATMs); and whether there are limitations on the share of the banking market which can be foreign-owned. Also covered are permissible or required activities and operations issues, including whether banks can engage in securities activities, and if so whether there are restrictions on specific types of securities activities, whether banks can engage in insurance activities; whether there are limitations on foreign ownership of the equity of individual banks; and whether 'National Treatment' is accorded to foreign banks with respect to a country's requirements on the composition of a bank's board of directors.

With these bank entry, permissible activities, and operations requirements in mind, Table 17.2 presents a summary profile of countries' postures under the WTO. The basic perspective in Table 17.2 is to consider the number and percentage of countries that have not committed to openness or 'National Treatment' under the WTO. It is important to bear in mind that, under the WTO, not making a commitment to openness does not necessarily equate to embracing restrictive policies; and, indeed, specifically in the case of the first-entry concept included in Table 17.2—WTO 103—it is quite possible that a country's policy under the WTO is to retain full discretion to impose some manner of banking sector restrictions, while at the same time the country adopts policies to encourage entry. On the other side of the issue, a country, under the WTO, that explicitly renounces discretion is, by that act, declaring that it will adopt liberal entry and activities policies. Nevertheless, in the first case, because a country retains the right under the WTO to impose restrictive policies, we regard that posture as less open than that of another country which declares it is not retaining such discretion.

As a strong generalization, Table 17.2 shows that for most entry, activities, and operations issues, developing countries as a group strike a more-restrictive posture under the WTO than do developed countries. Note that there are no WTO definitions of 'developed' and 'developing' countries: WTO members basically announce for themselves whether they are 'developed' or 'developing' countries for the purpose of the application of WTO agreements. We chose to use a more raditional classification scheme in our study: 'developing countries' are those which the World Bank categorizes as 'low-income' and 'lower-middle-income', and 'developed countries' include those which the World Bank includes in its 'upper-middle-income' and 'high-income' categories; some of the latter may be considered as 'developing' in the WTO context.

Table 17.2. WTO commitments on foreign bank entry and activities: Aggregate profile of member countries' restrictiveness

Entry or Permissible Activity	WTO Component	Nature of Restriction/Limitation/ Prohibition	Number of Member Countries			Percent of Member Countries[1]		
			Total	Developed	Developing	Total	Developed	Developing
Entry and Licensing of banks	WTO 103	Countries retaining full discretion (i.e., those not making any banking sector entry/liberalization commitments)	28	0	28	22.8	0.0	29.8
	WTO 104	Countries making commitments, but retaining a high degree of discretion (by, e.g., making entry subject to an economic needs test)	58	3	55	47.2	10.3	58.5
	WTO 105	Countries not committing to full national treatment for foreign banks	77	14	63	62.6	48.3	67.0
	WTO 106	Countries scheduling limitations on the number of foreign banks allowed to operate in the country	38	1	37	30.9	3.5	39.4
	WTO 107	Countries scheduling limitations or prohibitions on new entry after a specified number/level of foreign bank penetration	41	1	40	33.3	3.5	42.6
Forms of entry	WTO 1.12.1	Countries not making commitments to allow foreign bank entry via acquisition of an existing bank	33	1	32	26.8	3.5	34.0
	WTO 1.12.2	Countries not making commitments to allow foreign bank entry via establishment of subsidiary	44	2	42	35.8	6.9	44.7
	WTO 1.12.3	Countries not making commitments to allow foreign bank entry via branching	42	0	42	34.2	0.0	44.7

Minimum capital requirements	WTO 1.3.1	Countries with higher capital entry requirements for foreign vs. domestic banks	37	1	36	30.1	3.5	38.3
Expansion of physical presence	WTO 105a	Countries not commiting to national treatment for expansion (once established in the country) via new offices, branches, ATMs	58	8	50	47.2	27.6	53.2
Composition of board of directors	WTO105b	Countries not commiting to national treatment for requirements on composition of board of directors	60	7	53	48.8	24.1	56.4
Foreign equity limitations	WTO 108	Countries scheduling limitations on foreign ownership of a bank's equity	41	0	41	33.3	0.0	43.6
Foreign banking concentration	WTO 109	Countries scheduling limitations on foreign bank share of total banking system assets	34	0	34	27.6	0.0	36.2
Securities activities	WTO 4.1.1	Countries restricting or prohibiting banks from engaging in securities underwriting	73	8	65	59.4	27.6	69.2
	WTO 4.1.2	Countries restricting or prohibiting banks from engaging in securities dealing and brokering	70	7	63	56.9	24.1	67.0
	WTO 4.1.3	Countries restricting or prohibiting banks from engaging in mutual funds activities	81	9	72	65.9	31.0	76.6
Insurance activities	WTO 110	Countries restricting or prohibiting banks from engaging in insurance	110	26	84	89.4	89.7	89.4

[1] Percentages are calculated on the basis of the total number of countries in the group responding to the given item/question. In most cases, 123 total countries responded to the World Bank questionnaire about a given activity/restriction, including 29 developed countries and 94 developing countries.

World Bank 'reported practices' data

A growing body of research has used the World Bank's survey data on banking regulations and practices based on information provided by financial supervisory authorities around the world; see Barth, Caprio, and Levine (2006) for a comprehensive description of that data. Table 17.3 lists and briefly defines the World Bank 'reported practices' concepts we employ in this chapter, from the 2003 version of the survey, displayed in a manner similar to our Table 17.1 glossary of WTO commitment-related concepts. In some cases the overlap between the World Bank concepts and the WTO concepts is exact—for example, both data sets cover possible forms of foreign bank entry, including via acquisition of an existing bank, establishment of a separately capitalized subsidiary, on via branching. In other cases, the data sets complement each other, as, for example, in several of the 'entry and licensing' concepts. In addition, as the memorandum to Table 17.3 shows, the World Bank data contain detailed information on foreign banks' applications for entry, including the number of such applications each country received over the five-year period up to and including 2003, the number of applications of various kinds denied, and the reasons for denials.

Aggregating the World Bank data for both developed and developing countries reveals more similarities than differences in outcomes. For example, as with developed countries, all or almost all developing countries allow entry via acquisition, establishment of a subsidiary, and branching. In addition, the clear majority of both developing and developed countries apply similar minimum-capital entry requirements for foreign and domestic banks; indeed, a somewhat greater percentage of developing countries (73.4 percent) subscribe to this policy as compared to the percentage of developed countries (62.1 percent). For permissible securities and insurance activities for banks, developing countries as a group are somewhat less liberal than developed countries, but, except in the case of insurance underwriting, the clear majority of developing countries allow a rather wide range of non-commercial banking activities for banks.[2]

The last two items, World Bank components 3.8.1 and 3.8.2, dealing with, respectively, government ownership and foreign ownership of banking system assets, present interesting contrasts between developed and developing countries. In both cases a significant number of countries' supervisory authorities reported to the World Bank that they had government-owned banks in operation, and foreign banking presence. But differences between the two groups of countries are noteworthy. First, somewhat under half (42.9 percent) of developed countries have government-owned banks operating in their banking markets (where 'government ownership' is defined as the government having a 50 percent or higher equity

[2] Details for each component concept, across both developed and developing countries, are available from the authors.

Table 17.3. Glossary for World Bank data on reported practices: Entry into banking and permissible banking activities

Entry, Permissible Activity, or Banking System Characteristic: Overall Concept	Relevant Reported Practice or Banking Industry Characteristic	
	WB Code	Specific Substance of WB Component
Entry and licensing of banks	1.1.1	Is there more than one body or agency that grants licenses to banks?
	1.1.2	Is more than one license required? (e.g., separate licenses for commercial banking, securities operations, insurance, etc.)
Forms of entry	1.12.1	Entry via acquisition of an existing bank
	1.12.2	Entry via establishment of subsidiary
	1.12.3	Entry via branching
Ownership of banks	2.3	Can non-financial firms own shares in commercial banks?
	2.3.1	Can non-financial firms own voting shares in commercial banks?
	3.8.1	Fraction of banking system assets in banks 50% or more government-owned
	3.8.2	Fraction of banking system assets in banks 50% or more foreign-owned
Minimum capital requirements	1.3.1	Minimum capital entry requirement similar for both foreign and domestic banks
Deposit insurance system	8.1	Is there an explicit deposit insurance protection system?
Permissible activities: securities	4.1.1	Banks engaging in securities underwriting
	4.1.2	Banks engaging in securities dealing and brokering
	4.1.3	Banks engaging in mutual funds activities
Permissible activities: insurance	4.3.1	Underwriting
	4.3.2	Selling
Entry applications	1.10a	Number of commercial banking applications received from foreign countries in past 5 years
	1.10b	Number of such applications denied
Entry applications	1.10.1a	Number of applications from foreign entities to enter through acquisition of domestic bank
	1.10.1b	Number of such applications denied
	1.10.2a	Number of applications from foreign entities to enter through new, capitalized subsidiary
	1.10.2b	Number of such applications denied
	1.10.3a	Number of applications from foreign entities to enter through opening a branch
	1.10.3b	Number of such applications denied
	1.10.4a	Number of applications from foreign entities to enter through some other means
	1.10.4b	Number of such applications denied

(cont.)

Table 17.3. (*Continued*)

Entry, Permissible Activity, or Banking System Characteristic: Overall Concept	Relevant Reported Practice or Banking Industry Characteristic	
	WB code	Specific Substance of WB Component
	1.11.1	Primary reason for denial of application: Capital amount or quality
	1.11.2	Primary reason for denial of application: Banking skills of applicants
	1.11.3	Primary reason for denial of application: Reputation of applicants
	1.11.4	Primary reason for denial of application: Incomplete application
	1.11.5	Primary reason for denial of application: Other reason(s)

share in a bank). This contrasts with the nearly two-thirds (65.1 percent) of developing countries reporting government ownership of banking system assets. However, the median proportion of government ownership of banking is similar in both developed and developing groups, and in fact is slightly lower, at 16.8 percent, in developing countries than in developed countries (at 18.5 percent). Hence, although a larger percent of developing countries have government-owned banks as compared to developed countries, on average the share of the banking market in the hands of government owned banks is lower in developing countries, as compared to developed countries with government-owned banks.

A complex story also emerges from the data on foreign ownership of banking system assets. For both sets of countries, high proportions reported foreign ownership of some share of banking system assets, with a strong majority (86.4 percent) of developed countries and almost all (97.5 percent) of developing countries responding in the affirmative. However, the median value of banking system assets that are foreign-owned (i.e., in banks 50 percent or more of whose equity shares are owned by foreigners) is much higher—almost half (47.3 percent)—in developing countries, compared with just over 10 percent on average (i.e., median) for developed countries.

Complementarities in the data sets

Comparing countries' stances toward foreign bank entry and operations, as reflected in both the WTO commitments data and the World Bank reported

practices data is important for the following reason: knowledge of policy without knowledge of facts 'on the ground' does not give a complete picture—but, by the same token, knowing the facts about the level of foreign banking participation does not reveal, per se, a country's policy tilt. In particular, it is possible that, on the one hand, a country may in principle commit to a very open policy stance with respect to foreign banking but have, in practice, very little of it. On the other hand, a country may strike a seemingly restrictive policy stance toward foreign bank entry and operations but nevertheless enjoy a substantial foreign banking presence.

To investigate this issue, we grouped countries according to their position under WTO component 109: those countries which do not place limitations on foreign bank share, or which do not declare under WTO 109 that they retain the discretion to do so are considered to be 'without WTO limitations'; all other countries either have specific limitations or retained the discretion to impose them, and these are classified as 'with WTO limitations'. Using the World Bank data, we then stratified countries according to the proportion of banking system assets in the hands of foreign-owned banks (into those with 0–25 percent, 25–50 percent, 50–75 percent, and 75–100 percent foreign ownership of banks). We found that the percentage of countries with WTO limitations is actually higher in the upper two foreign bank penetration quartiles than in the two lower foreign bank penetration groups. In particular, half of the eighteen countries in the 50–75 percent foreign-ownership-of-banking-system-assets group, and more than one-third (34.8 percent) of the twenty-three countries with 75–100 percent foreign-ownership-of-banking-system-assets group have WTO limitations on foreign penetration of their banking systems (or have retained the possibility of imposing such limitations). Of the sixty-two countries with 50 percent or less foreign-owned bank shares, less than one-fifth had WTO limitations on foreign bank asset shares. These results suggest the degree of foreign bank entry does not correlate especially highly with a country's WTO stance.

We further investigated the WTO 109 limitations-cum-foreign-owned banking asset quartile groupings by looking at developed and developing countries separately. Developed and developing countries are very different in their stances on the adoption of WTO limitations on foreign bank penetration. No developed countries set limitations on foreign bank ownership shares under WTO 109, while more than one-third of developing countries (34 out of 94) have asserted limitations on (or retained the discretion to place limitations on) foreign banks' share of banking system assets. It is of course an empirical question whether high foreign ownership share 'causes' countries to place such limitations at the WTO. What is clear from this analysis is that the adoption at the WTO of limitations (or the assertion that a country retains the discretion to adopt limitations) on foreign bank penetration does not, of itself, signal that a country is closed to foreign banks.

In a similar vein, we might at first assume that 'open' countries receive more banking applications than relatively 'closed' countries. We might also suppose that application rejection rates for closed countries are higher than for open countries. Because of our unique ability to consult both WTO commitments data and the banking applications information in the World Bank data set, we can test these notions directly. We consider in particular four banking-entry components in the WTO data, and look at the number of banking applications, and application rejection rates in the World Bank data.

Based on our first line of inquiry, thirty-eight countries announced limitations under WTO 106 on the total number of foreign banks they allow to operate within their borders, and eighty-five countries did not, under WTO 106, place such limitations on foreign banks. Of the thirty-eight countries placing WTO 106 limitations on the number of foreign banks, fourteen nevertheless received applications, while fewer—thirteen—did not receive any applications for entry (and for eleven of the WTO 106 'Yes' countries there was no World Bank entry applications information—so we do not know if these countries received no applications, or for one reason or another chose not to answer, or were unable to answer, the survey question). For sixty of the eighty-five countries not placing limitations under WTO 106 there was also World Bank applications data. These countries together received 1,061 applications for foreign bank entry, and rejected 5.9 percent of them over the five-year period ending in 2003. But, by no means did the WTO 106 'Yes' countries reject all foreign entry applications. Indeed, they rejected just over 30 percent of the forty-six entry application they received. Thus, despite a much lower number of foreign bank entry applications, and a higher rejection rate, the countries imposing limitations on entry under WTO 106 nevertheless approved almost seventy of the applications they received.

We employ the same strategy to pair World Bank banking applications data with countries' policies under WTO 107—that is, whether new entry is prohibited after a certain number of foreign banks have entered. Forty-one countries declared 'Yes' under WTO 107 and eighty-two said 'No'. As with WTO 106, a greater percent of the 'Yes' countries nevertheless received foreign banking applications than did not. Also, as in the preceding discussion, far more applications were received by the 'No' group of countries, which also had a lower rejection rate for those applications. Nevertheless, the gap between rejection rates for the 'Yes' and the 'No' countries was narrower than in the case of WTO 106. Employing the same strategy, we found similar results for the pairing of WTO 108 (whether there are limitations on the share of a bank's equity that can be held by foreigners) with applications data. Thus the data show that it is true that 'open' countries receive more banking applications than relatively 'closed' countries, and application rejection rates for 'closed' countries are higher than for open countries. However, it is also clear that the existence of WTO limitations does not completely stifle foreign bank entry.

Do countries' World Trade Organization commitments match their reported banking practices?

We approach the issue of consistency and discrepancies between the two views of the world reflected in the data sets two ways. First, we focus on a comparison of forms of entry components and permissible banking activities components, sets of components where the two data sets' information is conveyed nearly identically. At first blush, the data sets seem to match up at least roughly consistently across these elements, but further examination reveals substantial inconsistencies between them. In light of this, our second analytic approach is to construct overall gauges of openness, across multiple components in each data set, in order to get a more comprehensive answer to the question posed in this section's title.

Entry requirements and permissible banking activities: Do WTO commitments and reported practices data give the same picture?

As the 'glossaries' presented in Tables 17.1 and 17.3 show, both data sets cover three foreign bank entry strategies in an identical fashion. They also cover banking involvement in securities and insurance activities in a similar way. Table 17.4 compares the two sets of data. For each component of entry strategy or permissible banking activity the table shows, in its left-hand half, the number of countries in each data set answering 'Yes' (indicating more openness—that is, that a given form of entry or banking activity is allowed). The left-hand portion of the table also shows the number of countries answering 'no'—that is, that there are limitations with respect to the given element. Clearly, in no cases do the WTO commitments data match the reported practices data. Indeed, for all components except the minimum capital entry requirements element the discrepancies appear to be quite wide.

However, from the left-hand side of Table 17.4 we do not know, for example, how many of the eighty-six countries answering 'Yes' about minimum capital entry requirements under the WTO are included among the eighty-seven countries answering 'Yes' according to the World Bank data. In order to address this uncertainty, the right-hand side of Table 17.4 summarizes the country-by-country comparisons we undertook. In the case of any given element, if discrepancies exist, they could take one of two forms. First, it is possible that a country positioned itself in an open posture under the WTO but in practice is more restrictive. In this case,

Table 17.4. Entry requirements and permissible banking activities: WTO commitments and reported practices compared

Entry requirements and permissible banking activities: Common coverage in WTO and WB databases	WTO and WB ('reported practices') responses				Where WTO and WB data are different, are reported practices more restrictive or less restrictive than WTO commitments?			
	Number of 'yes' countries		Number of 'no' countries		Reported practices more restrictive (WTO = 'yes' **but** WB = 'no')		Reported practices less restrictive (WTO = 'no' *but* WB = 'yes')	
	WTO	WB	WTO	WB	Number of Countries	Percent More Restrictive[1]	Number of Countries	Percent Less Restrictive[2]
Foreign banks can enter via acquisition	90	123	33	0	0	0.0	33	100.0
Foreign banks can enter via establishment of subsidiary	79	122	44	1	1	1.3	44	100.0
Foreign banks can enter via branching	81	112	42	11	5	6.2	36	85.7
Minimum capital entry requirement *similar* for domestic and foreign banks	86	87	37	36	25	29.1	26	70.3
Permissible banking activities:								
Securities activities: underwriting	50	111	73	12	1	2.0	62	84.9
Securities activities: dealing and brokering	53	99	70	24	3	5.7	49	70.0
Securities activities: mutual funds activities	42	89	81	34	12	28.6	59	72.8
Insurance: underwriting and/or selling[3]	13	37	110	86	3	23.1	35	31.8

[1] Number of countries with WTO 'Yes' *but* WB 'No' countries as a percent of the total number of WTO 'Yes' countries.
[2] Number of countries with WTO 'No' *but* WB 'Yes' countries as a percent of the total number of WTO 'No' countries.
[3] Insurance underwriting and selling are not separately designated in the WTO data; for the WB data, countries which allow either underwriting or selling, or both, are included.

it would be a 'Yes' country in the WTO data but a 'No' country in the World Bank data. The other possible discrepancy would be if a country took a restrictive stance under the WTO but in practice was open (that is, was a WTO 'No' but a World Bank 'Yes'). A review of the right-hand side of Table 17.4 shows the following:

- Can foreign banks enter via acquisition of an existing domestic bank, and can they enter via the establishment of a subsidiary of the home country parent bank (first two rows)? Only one country was more restrictive in practice than under the WTO (i.e., WTO = 'Yes' but World Bank response = 'No'). On the other hand, for these two forms of foreign bank entry, thirty-three and forty-four countries, respectively, were less restrictive in practice than their WTO posture would indicate. That amounted to 100 percent of the countries that set limitations on these form of entry under the WTO.
- Can foreign banks enter via branching? Five countries were more restrictive in practice than they declared themselves to be under WTO, but thirty-six countries were less restrictive in practice about foreign branch entry. Looked at another way, 85.7 percent of the countries which said under the WTO that they restricted or prohibited foreign entry via branching did in fact allow it, according to their supervisory authorities' responses to the World Bank survey.
- Are minimum capital entry requirements similar for foreign and domestic banks? Twenty-five countries said under WTO 1.3.1 that they did not impose different capital entry requirements on foreign banks when in practice such requirements were different. Hence, almost 30 percent of the countries that committed to similar treatment in this respect under the WTO were more restrictive in practice for this crucial entry requirement. On the other hand, twenty-six of the countries—over 70 percent—that said under the WTO that they had (or retained the discretion to impose) different capital entry requirements on foreign banks did not in fact do so, and were therefore more open in practice than their WTO posture would indicate.
- Can banks engage in securities activities (underwriting, dealing and brokering, or mutual funds activities)? For securities underwriting and securities dealing and brokering, a few countries are more restrictive in practice than their WTO stances indicate. For mutual funds activities, more than one-fourth of the countries committing to an open policy on this issue in fact imposed restrictions or prohibitions (twelve out of forty-two countries, or 28.6 percent). It is the other side of the coin that stands out about permissible banking activities, however: between 70 and 85 percent of the countries committing to limitations on various kinds of securities activities for banks under the WTO are not, in practice, restrictive.
- Can banks engage in insurance activities? Three countries committed to openness for banks engaging in insurance activities but were in fact restrictive. On the other hand, thirty-five countries took a more open stance toward insurance activities for banks than they asserted under the WTO.

The substantial discrepancies summarized above merit further examination. The next section does this by constructing and then employing a more comprehensive yardstick.

A comprehensive comparison of WTO commitments and reported practices

In this section, we first tackle the challenge of developing a meaningful overall gauge of banking 'openness' by constructing an index which weights the relative importance of many different aspects of openness and then sums them up. Our 'market openness index' is described in Table 17.5. Because both the WTO commitments data and the World Bank reported practices data have identical or similar coverage of a range of components, we are able to construct comparable variants of the market openness index from each set of data. Both the WTO commitments variant and the reported practices variant include multiple aspects of licensing of banks, forms of entry by foreign banks, and securities activities. In addition, both the WTO and World Bank data cover limitations on foreign ownership of a bank's equity, minimum-capital-entry requirements, and insurance activities for banks. Only the WTO data explicitly cover the expansion of physical presence for foreign banks once they have entered the host country's market, and requirements about the composition of a bank's board of directors; for these two aspects of market openness, we use the values assigned to the WTO commitments data in both variants of the index.

Following Barth, Marchetti, Nolle, and Sawangngoenyuang (2006), we assign values from 0 to 1 to the specific information on each component for each country. A '0' value indicates that the WTO stance or reported practice is completely open (i.e., the given country does not restrict the activity or entry conduit or condition in question). So, for example if, under the WTO, a country commits to placing no (non-prudential) limitations on the share of banking system assets that can be owned by foreign banks, the value assigned for the specific component (WTO 109) for that country is 0. Similarly, if the same response also holds for the country for World Bank item 3.8.2, that variable in the reported practices variant of the index also is assigned a value of 0. At the other end of the assigned-values spectrum, in a case where there is a hard-and-fast limit on the share of banking system assets that foreign banks can own, that component receives a value of 1. In between are countries with partial limitations, restrictions, or conditions, and these are assigned values from 0.25 (almost but not completely open) to 0.75 (heavily restricted but not completely prohibited), in increments of 0.25. For each component then, the lower the assigned value, the more open the country with respect to that particular aspect. For index elements, such as the 'forms of entry' element,

Table 17.5. Index of market openness: WTO commitments and World Bank reported practices

Index element	Index weight	Alternative constructs[1]	
		'Commitments' variant: WTO components included[2]	'Reported practices' variant: WB components included[3]
Licensing of banks	20	WTO 103, WTO 104, WTO 106, WTO 107	WB 1.10a, WB 1.10b, WB 1.11.1, WB 1.11.2, WB 1.11.3
Foreign equity limitations	15	WTO 108	WB 1.12.1
Foreign banking concentration: limitations on foreign bank share of banking system assets	15	WTO 109	WB 3.8.2
Forms of entry	10	WTO 1.12.1, WTO 1.12.2, WTO 1.12.3	WB 1.12.1, WB 1.12.2, WB 1.12.3
Minimum capital requirements	10	WTO 1.3.1	WB1.3.1
Securities activities	10	WTO 4.1.1, WTO 4.1.2, WTO 4.1.3	WB 4.1.1, WB 4.1.2, WB 4.1.3
Insurance activities	10	WTO110	WB 4.3.1, WB 4.3.2
Expansion of physical presence	5	WTO 105a	No comparable WB component: WTO value used
Composition of board of directors	5	WTO 105b	No comparable WB component: WTO value used
Highest possible index value (higher values indicate less openness/greater restrictiveness)	(Sum of weights = 100)	100	100

[1] Calculation of index values: for each index element, WTO and WB components were evaluated according to the degree of openness reflected in individual country responses, as recorded in the respective databases. Values ranged from 0.0 (completely open, permitted without restriction) to 1.0 (very restricted, prohibited). These values, ranging from 0.0 to 1.0, were multiplied by the weights assigned to each index element (e.g., for the 'licensing of banks' index element, values across countries ranged from a low of 0 [indicating no impediments to obtaining a banking license in order to enter a country's banking market] to 20 [indicating very substantial impediments to obtain a banking license in the country]). Weighted values for each index element for a country were then summed to obtain an overall 'Market Openness' index value for that country. Conceptually, overall index values range from 0 (no artificial impediments to/restrictions on market openness) to 100 (very closed/highly restrictive banking market). Details of index calculations available on request from the authors.

[2] See Table 17.1 for definitions of WTO components.

[3] See Table 17.3 for definitions of WB components.

where there are multiple components under consideration, the values for each component are weighted equally within the element and then summed up for the index element to an overall value of no more than 1. Finally, each index element is weighted according to our judgment of its overall relative importance in characterizing market openness. For example, the value assigned to the foreign equity limitations component is multiplied by 15, while the value assigned to the requirements about a bank's board of directors is multiplied by 5. The weighted component values are summed up, with the maximum index value being 100 (i.e., '1s' assigned to every component, in every element), signifying a banking market very restrictive toward foreign entry.

Table 17.6 lists country-by-country market openness index results. There were sufficient WTO data to calculate a WTO commitments variant of the index for all 123 countries; sixty-five of these countries also reported enough information for the index elements to allow us to calculate the reported practices variant of the market openness index as well.

One of our main goals is to ascertain in a comprehensive manner for each country the degree to which WTO commitments about the country's openness to foreign banks is in fact reflected in what the country's supervisory authorities say is taking place 'on the ground'. Hence, for the sixty-five countries for which we could calculate both variants of the market openness index, we ranked countries by the 'degree of discrepancy' between the two variants. Specifically, we subtracted the value of the WTO commitments variant from the value of the reported practices variant to quantify the 'degree of discrepancy'. The more positive the value of this difference between the index variants, the more restrictive/closed are reported practices compared to a country's WTO declarations about market openness across the range of component aspects. A negative value for the 'degree of discrepancy' indicates that reported practices are in fact less restrictive/more open than a country's WTO commitments would indicate; and, across countries, higher negative values indicate relatively greater openness in practice, relative to WTO stances, as compared to other countries with lower negative values (or positive values).

Table 17.6 groups the sixty-five countries for which both variants were calculated into one of three groups: (i) those in which the 'degree of discrepancy' is positive—that is, where reported practices are more restrictive than WTO commitments; (ii) countries where the values of the variants of the index are equal, and therefore where the 'degree of discrepancy' is 0; and (iii) countries where the 'degree of discrepancy' is negative, and therefore where reported practices are less restrictive/more open than the WTO stance.

There are thirty-four countries in the first group—that is, where reported practices are more restrictive than WTO commitments suggest. For about fifteen of these countries (44 percent), the 'degree of discrepancy' is less than a value of 10. Another seven or so (about 20 percent) have values close to 10, an arbitrary cut-off for a 'low' 'degree of discrepancy'. This evaluation still leaves twelve countries—over one-third

Table 17.6. Measuring market openness: WTO commitments and World Bank reported practices compared

Country	Market openness index[1]		Degree of discrepancy[2]
	'Reported practices' variant	'Commitments' variant	
Countries with reported practices more restrictive than WTO commitments			
Guyana	61.3	25.0	36.3
Finland	46.3	15.0	31.3
Ecuador	48.8	20.0	28.8
Israel	36.3	10.0	26.3
Greece	31.3	10.0	21.3
Slovenia	21.3	0.0	21.3
Bulgaria	23.8	5.0	18.8
Jordan	33.8	15.0	18.8
South Korea	37.5	20.0	17.5
Czech Republic	20.0	3.3	16.7
Saudi Arabia	26.3	10.0	16.3
Egypt	46.3	30.8	15.5
Costa Rica	35.4	23.3	12.1
Germany	21.3	10.0	11.3
Italy	21.3	10.0	11.3
Liechtenstein	31.3	20.0	11.3
South Aftrica	16.3	5.0	11.3
Spain	16.3	5.0	11.3
Lesotho	15.0	5.0	10.0
Hungary	15.0	8.3	6.7
Australia	36.3	30.0	6.3
Honduras	36.3	30.0	6.3
Netherlands	16.3	10.0	6.3
Portugal	21.3	15.0	6.3
Macau, China	25.0	20.0	5.0
Armenia	28.8	25.0	3.0
Cyprus	21.3	18.3	3.0
Republic of Moldova	22.5	20.0	2.5
Romania	22.5	20.0	2.5
Venezuela	32.5	30.0	2.5
Slovak Republic	15.0	13.3	1.7
Japan	16.3	15.0	1.3
Norway	26.3	25.0	1.3
Turkey	21.3	20.0	1.3
Countries with reported practices equal to WTO commitments			
Chile	12.5	12.5	0.0
Lithuania	15.0	15.0	0.0
Countries with reported practices more open than WTO commitments			
Guinea	10.0	100.0	−90.0
Botswana	18.3	100.0	−81.7
Rwanda	30.0	100.0	−70.0

(cont.)

Table 17.6. (*Continued*)

Country	Market openness index[1]		Degree of discrepancy[2]
	'Reported practices' variant	'Commitments' variant	
Fiji	35.0	100.0	−65.0
Gambia	40.0	100.0	−60.0
Belize	40.0	100.0	−60.0
Guatemala	41.3	100.0	−58.7
Trinidad and Tobago	49.6	100.0	−50.4
Tunisia	26.3	75.0	−48.7
Namibia	58.8	100.0	−41.2
El Salvador	38.8	75.8	−37.0
Malta	8.8	40.0	−31.2
Aruba	5.0	25.0	−20.0
India	41.3	59.1	−17.8
Poland	8.8	25.0	−16.2
Ghana	28.8	45.0	−16.2
Latvia	13.8	25.0	−11.2
New Zealand	10.0	20.0	−10.0
Luxembourg	0.0	10.0	−10.0
United States	21.3	30.0	−8.7
Argentina	12.5	20.0	−7.5
Brazil	27.5	34.1	−6.6
Panama	13.8	20.0	−6.2
Bahrain	33.8	40.0	−6.2
Kyrgyz Republic	26.3	30.0	−3.7
Uruguay	22.5	25.0	−2.5
Peru	17.5	20.0	−2.5
Kenya	37.5	40.0	−2.5
Albania	22.5	25.0	−2.5

[1] Lower index values indicate greater openness (or less restrictiveness).

[2] Countries ranked from greatest positive to greatest negative value of degree of discrepancy, which is calculated as reported practices index value minus commitments index value. Positive values for the degree of discrepancy indicate that reported practices are more restrictive (or less open) than WTO commitments; negative values indicate that reported practices are more open (or less restrictive) than WTO commitments. For 56 countries there was insufficient data to calculate the reported practices variant.

of the group—with a positive difference between the reported-practices variant and the commitments variant large enough to conclude that a substantial proportion of countries have, in effect, overstated their market openness under WTO.

For the third group of countries—those with a negative value for the 'degree of discrepancy'—a greater proportion (66 percent) have substantial negative values, again using an arbitrary cut-off of (negative) −10.0. Countries ranging from Guinea (−90.0) to Luxembourg (−10.0) all show a sufficiently wide gap between reported practices and WTO commitments by this gauge for us to conclude that there are a

substantial proportion of countries whose reported practices are, in fact, more open than their WTO stance would lead one to believe. Interestingly, only two countries—Chile and Lithuania—show equal index variant values, indicating that the degree of market openness in which they engage in practice accurately reflects their WTO posture.

We can also summarize the results of the comparison of the two variants of the market openness index results for developed and developing countries. For the sixty-five countries for which we were able to calculate both the reported practices and the WTO commitments variants of the index, sixteen are developed countries and forty-nine are developing countries. For both groups, the mean values (23.7 for developed, 27.2 for developing) and standard deviations (11.6 and 13.1 respectively) of the reported practices variants are similar, indicating that, on average, both groups of countries are similarly 'open/restrcitive'. Indeed, the mean values for the two groups are not statistically different. The story is very different when we look at mean values for the commitment variant. In this case, developing countries as a group have an average commitment-market openness index value more than double that for developed countries (38.8 vs. 15.8), which is statistically higher, indicating much less open markets on average for developing countries. The 'degree of discrepancy' is, however, indispensable for completing the summary. Even though developed countries are more open, as judged by the group's mean value for the commitments variant, they are in practice somewhat more restrictive compared to their WTO commitments, with a mean value for the 'degree of discrepancy' of 7.8, which is statistically significant. By contrast, although developing countries are more restrictive relative to developed countries, as per the group's mean value for the commitments variant, they are as a group significantly more open (less restrictive) in practice, with a mean value for the 'degree of discrepancy' of −11.6, which is statistically significant.

Table 17.7 introduces one additional, related perspective to this issue. Because several of the market openness index components apply not just to foreign banks, but to domestic banks as well, we can calculate a 'domestic banks' gauge of restrictiveness. That is, using the WTO and World Bank components listed in Table 17.7, we can calculate measures of how restrictive countries' postures are with respect to a blend of licensing requirements, securities activities, and insurance activities that apply to any bank, foreign or domestic, that operates or wishes to operate in a country. As Table 17.7 shows, using exactly the same values-assignment and index-elements-weighting schemes as for the overall index of market openness, we construct a gauge of how restrictive countries are, where the maximum domestic banks index value is forty instead of 100.

It is not, however, the domestic banks index variants by themselves on which we focus, but rather, as illustrated in Table 17.8, the combined consideration of the market openness index and the domestic banks index together by calculating the *difference* between the overall market openness index and the abbreviated domestic

Table 17.7. Index of restrictions on domestic banks' powers, activities, and requirements: WTO commitments and World Bank reported practices

Index element	Index weight	Alternative constructs[1] (excludes elements applying only to foreign banks)	
		'Domestic banks—WTO commitments' variant[2] (WTO commitments data)	'Domestic banks—reported practices' variant[3] (World Bank 'reported practices' data)
Licensing of banks	20	WTO 103, WTO 104, WTO 106, WTO 107	WB 1.10a, WB 1.10b, WB 1.11.1, WB 1.11.2, WB 1.11.3
Securities activities	10	WTO 4.1.1, WTO 4.1.2, WTO 4.1.3	WB 4.1.1, WB 4.1.2, WB 4.1.3
Insurance activities	10	WTO110	WB 4.3.1, WB 4.3.2
Highest possible index value (higher values indicate less openness/greater restrictiveness)	(Sum of weights = 40)	40	40

[1] Calculation of index values: for each index element, WTO and WB components were evaluated according to the degree of openness reflected in individual country responses, as recorded in the respective databases, in a manner identical to the construction of the 'market openness' index (see Table 17.4), except that elements applying only to foreign bank entry were excluded. Values ranged from 0.0 (completely open, permitted without restriction) to 1.0 (very restricted, prohibited). These values, ranging from 0.0 to 1.0, were multiplied by the weights assigned to each index element (e.g., for the 'licensing of banks' index element, values across countries ranged from a low of 0 [indicating no impediments to obtaining a banking license in order to operate in a country's banking market] to 20 [indicating very substantial impediments to obtain a banking license in the country]). Weighted values for each index element for a country were then summed to obtain an overall 'domestic banks' index value for that country. Conceptually, overall index values range from 0 (no artificial impediments to/restrictions on market openness) to 40 (very closed/highly restrictive banking market). Details of index calculations available on request from the authors.

[2] See Table 17.1 for definitions of WTO components.

[3] See Table 17.3 for definitions of WB components.

banks index to arrive at a measure of the 'degree of discrimination' between foreign and domestic banks. For this construct, the higher the value, the greater the regulations-related disadvantage under which foreign banks operate. The degree of discrimination measure can be calculated for both the WTO commitments data, and the reported practices data. Our inspection of the country-by-country results showed that countries do not rank in the same order in terms of highest to lowest degree of discrimination.[3] Hence, in Table 17.8, we focus on the difference between the degree of discrimination values using the WTO commitments data and that using the reported practices data to ascertain which countries discriminate against foreign banks vs. domestic banks 'on the ground', as compared to what they have

[3] Country-specific results are available from the authors.

Table 17.8. Are foreign banks treated differently from domestic banks? Reported practices compared to WTO commitments

Country	Degree of Discrimination under WTO Commitments	Degree of Discrimination in Reported Practice	Difference between Reported Practices and WTO Commitments with respect to the Degree of Discrimination[1] (Positive value indicates there is a greater degree of dsicrimination against foreign-owned banks relative to domestic banks in Reported Practice than under a country's WTO Commitments)
Australia	5.0	26.3	21.3
Ecuador	0.0	21.3	21.3
Finland	5.0	26.3	21.3
Greece	0.0	21.3	21.3
Guyana	5.0	26.3	21.3
Honduras	5.0	26.3	21.3
Israel	0.0	21.3	21.3
Norway	5.0	26.3	21.3
Saudi Arabia	0.0	21.3	21.3
Egypt	8.3	26.3	18.0
South Korea	5.0	22.5	17.5
Costa Rica	3.3	17.9	14.6
Jordan	5.0	18.8	13.8
Germany	0.0	11.3	11.3
Italy	0.0	11.3	11.3
Japan	5.0	16.3	11.3
Liechtenstein	5.0	16.3	11.3
Netherlands	0.0	11.3	11.3
Portugal	0.0	11.3	11.3
Slovenia	0.0	11.3	11.3
South Africa	5.0	16.3	11.3
Spain	0.0	11.3	11.3
Turkey	5.0	16.3	11.3
United States	10.0	21.3	11.3
Macau, China	5.0	15.0	10.0
New Zealand	0.0	10.0	10.0
Cyprus	8.3	16.3	8.0
Argentina	0.0	7.5	7.5
Chile	0.0	7.5	7.5
Peru	0.0	7.5	7.5
Romania	0.0	7.5	7.5
Uruguay	0.0	7.5	7.5
Venezuela	5.0	12.5	7.5
Czech Republic	3.3	10.0	6.7
Slovak Republic	3.3	10.0	6.7
Armenia	5.0	8.8	3.8
Bulgaria	5.0	8.8	3.8
Latvia	5.0	8.8	3.8
Panama	0.0	3.8	3.8
Poland	5.0	8.8	3.8

(cont.)

Table 17.8. (*Continued*)

Country	Degree of Discrimination under WTO Commitments	Degree of Discrimination in Reported Practice	Difference between Reported Practices and WTO Commitments with respect to the Degree of Discrimination[1] (Positive value indicates there is a greater degree of dsicrimination against foreign-owned banks relative to domestic banks in Reported Practice than under a country's WTO Commitments)
Kyrgyz Republic	10.0	11.3	1.3
Brazil	11.6	12.5	0.9
Aruba	5.0	5.0	0.0
Lesotho	0.0	0.0	0.0
Lithuania	5.0	5.0	0.0
Luxembourg	0.0	0.0	0.0
Bahrain	25.0	23.8	−1.3
Kenya	20.0	17.5	−2.5
Moldova Republic	10.0	7.5	−2.5
Hungary	8.3	5.0	−3.3
Ghana	20.0	13.8	−6.3
Malta	15.0	8.8	−6.3
Albania	15.0	7.5	−7.5
India	46.6	31.3	−15.4
Trinidad and Tobago	60.0	34.6	−25.5
Tunisia	50.0	21.3	−28.8
El Salvador	53.3	21.3	−32.1
Rwanda	60.0	25.0	−35.0
Namibia	60.0	23.8	−36.3
Guatemala	60.0	21.3	−38.8
Belize	60.0	20.0	−40.0
Gambia	60.0	20.0	−40.0
Botswana	60.0	13.3	−46.7
Fiji	60.0	10.0	−50.0
Guinea	60.0	10.0	−50.0
All Countries (Total number with information = 65)			
Mean	14.7	15.0	0.3 (2.4)
Standard deviation	21.2	7.7	19.6
Developed countries (Total number with information = 16)			
Mean	2.7	16.5	13.8*** (1.6)
Standard deviation	3.4	7.5	6.2
Developing countries (Total number with information = 49)			
Mean	18.6	14.5	−4.1 (2.9)
Standard deviation	23.1	7.8	20.5
Differences: Developing−developed			
Mean	15.9*** (3.4)	−2.0 (2.2)	−17.9*** (3.3)

Notes: [1]'Difference in discrimination' is calculated as the reported practices degree of discrimination minus the WTO commitments degree of discrimination.

*** denotes significance at the 90, 95, and 99% confidence levels, respectively. Values in parentheses are standard errors.

declared under the WTO. Before considering the results displayed in Table 17.8, we summarize two intermediate steps.

First, based on the calculation of the degree of discrimination between foreign and domestic banks using the WTO commitments data, twenty-eight of the 123 countries have the maximum value of 60—that is, compared to all other countries, the banking requirements and limitations to which these countries committed under the WTO yields the greatest degree of foreign vs. domestic banks. At the other end of the spectrum, thirty-two countries have a value of 0 for the degree of discrimination, meaning that, under the WTO, these countries have committed to treating foreign and domestic banks equally, regulations-wise. The remaining half (sixty-three) of the countries fall in between, some rather clearly committing to a greater regulatory burden for foreign banks, and some to almost no difference. Also, there are substantial differences between developed and developing countries. The mean value of the degree of discrimination for the twenty-nine developed countries as a group is, at a value of 3.4, far lower than the 25.5 value for developing countries, and the difference in means is statistically significant. Hence, under the auspices of the WTO, developed countries have committed to much less unequal treatment, regulations-wise, of foreign banks relative to domestic banks than have developing countries.

We also performed a parallel set of calculations of the degree of discrimination measure using the reported practice information from the World Bank survey data. With this data, we had enough observations with the complete set of index components to investigate a set of sixty-five countries. A number of the high-value WTO commitments countries dropped out, but, nevertheless, we retained a representative set of countries around the world, and found wide variations across countries in the degree of discrimination against foreign vs. domestic banks. There were far fewer '0'-values countries, and the maximum value of the degree of discrimination for these sixty-five countries was 34.6. Interestingly, the mean values of the degree of discrimination for the sixteen developed countries and the forty-nine developing countries were very close, unlike with the WTO commitments data (and, indeed, the mean value for developing countries is actually lower than for developed countries), but the difference is not statistically significant. Thus, it appears that while developing countries commit under the WTO to greater discrimination against foreign banks operating within their borders compared to domestic banks, in practice they discriminate less than their commitments suggest—and less than do developed countries.

To nail this conclusion down, Table 17.8 compares only those sixty-five countries for which we could calculate the degree of discrimination using both the WTO commitments and the reported practices data. In this case, we focus on the *difference* between the two measures of the degree of foreign vs. domestic bank discrimination. Specifically, we subtract the degree of discrimination under the WTO commitments data from the degree of discrimination using the reported

practices data. A positive value for this difference indicates that there is in fact a greater degree of discrimination against foreign vs. domestic banks in practice than a country committed to under the WTO. As Table 17.8 shows, for forty-two of the sixty-five countries it is the case that they discriminate more in practice than they said they would under the WTO. On the other hand, nineteen of the sixty-five countries discriminated less against foreign banks than they declared they would under the WTO (i.e., those with negative values); four countries' reported practices in this respect matched their WTO stance (i.e., those with a value of 0).

The summary part at the bottom of Table 17.8 reinforces our tentative developed vs. developing countries conclusion above. For this set of sixty-five directly comparable countries, we continue to find that, under the WTO, developing countries commit to a significantly greater degree of discrimination against foreign vs. domestic banks (see the respective mean values in the 'WTO commitments' column on the left-hand half of Table 17.8); while, under reported practices, developing countries discriminate less on average than do developed countries (see the 'reported practices' column on the right-hand side of Table 17.8). Furthermore, the means in the 'differences' column on the far-right-hand side of Table 17.8 show that developed countries in practice discriminate more against foreign banks than they indicate under the WTO (a positive mean value of 13.8, which is statistically significant), while developing countries discriminate less in practice (a negative value of −4.1, but it is not statistically significant).

The global financial crisis that began in 2007 in the subprime mortgage market in the US and, through 2008, enveloped all sectors of the financial market across countries around the world generated calls for greater cross-border banking supervision and regulation. The Group of Twenty countries became the focus of much of that discussion, and within that group of countries, the G-7 countries led the debate. A key prerequisite to any modification or reform of international banking supervision and regulation will be to have a clear picture of the landscape in which policy changes will take place. In this vein, we note that we are able to present in Table 17.8 index values for four of the G-7 countries (i.e., Germany, Italy, Japan, and the US), and that the the average index values of these four as a group are similar to (and not statistically significantly different from) the mean values for all developed countries in our data set. In particular, for these four G-7 countries, their mean value for the degree of discrimination index under WTO commitments is 3.7; their mean value for the degree of discrimination index in reported practices is 15.0; and their mean value for the difference between these two measures is 11.3. Hence, for these four significant members of the G-7, the observations about all developed countries also hold.

Although the analysis in this chapter shows clearly that WTO commitments frequently do not reflect what takes place in practice in many countries, there is one more set of information about the international provision of banking services available only in the WTO data that is worthwhile considering. Physical presence

within a country, as investigated in this study, is only one mode of services provision covered under the WTO (see Barth, Marchetti, Nolle, and Sawangngoenyuang, 2006 for an explanation of the three other modes of service provision). With this in mind, we find that there are a number of countries that allow the provision of banking services in a cross-border manner without the establishment of a physical presence. For most of the financial services, there are a few developed countries that allow cross-border (non-physical presence) provision. But most of the countries doing so, for any given financial service, are developing countries. Furthermore, many of these developing countries are those either at the lowest end of the income spectrum, or formerly centrally planned economies—both of which are among the countries least likely to have sufficiently advanced banking systems to be comfortable with allowing direct, physically present foreign competition. In light of this information, to the extent we can do so when describing the openness of developing compared to developed countries, we should bear in mind that cross-border provision of financial services mitigates to some extent otherwise high restrictions on physical bank entry.

Summary and conclusions

Our analysis is the first to offer a wide-ranging, cross-country picture of WTO commitments on banking services as compared to parallel information that country supervisory authorities have provided to the World Bank about banking regulations and entry requirements. We anticipate that the specific country-by-country data we include in the chapter will be of direct use to researchers and policymakers. Heretofore, research has taken into account reported banking regulations, banking industry structure, banking practices and corporate governance, the underlying legal system, and other variables affecting bank performance, economic development, and other key developments. This chapter adds a new empirical dimension to that mix, the policy stances that countries take under the WTO. Clearly, this information is becoming increasingly important from a policy perspective, and therefore researchers will wish to account for it in systematic analyses of banking and finance around the world.

In the process of assembling and working with our data we also have been able to investigate the important issue of the extent to which countries' commitments under the WTO are reflected in their reported banking practices. Using our new data set, and in particular comparing gauges of openness as measured, alternatively, by WTO commitments data and World Bank reported practices data, we found a number of significant discrepancies. Further, we were also able to compare results for developed and developing countries.

Our main conclusions are as follows. First, a significant number of countries are more restrictive in practice than they have obliged themselves to be under the WTO, but an even greater number are more open in practice than they committed to under the WTO. Second, developed countries as a group take a more open stance under the WTO than do developing countries, and in practice developed countries are also more open as a group than are developing countries. Third, nevertheless, developing countries are as a group less restrictive in practice than their WTO commitments oblige them to be, while developed countries are on average slightly more restrictive in practice than they say they will be under the WTO. Fourth, developed countries commit under the WTO to less unequal treatment for foreign as compared to domestic banks operating within their borders, while developing countries commit to a substantially greater degree of discrimination, regulation-wise, for foreign banks vs. domestic banks. However, fifth, in practice, the degree of discrimination against foreign banks and in favor of domestic banks is about equal in the two sets of countries, so that, sixth, developing countries disadvantage foreign banks less, and developed countries disadvantage foreign banks more, than they commit to under the WTO. As a policy matter, then, one could say it is more important to judge developed countries by what they do as compared to what they say under the WTO if one wishes to understand market openness in world banking. This conclusion will become increasingly important to bear in mind as the global community of policymakers grapples with reform of the international banking system.

References

Alexander, K. (2002). 'The World Trade Organization and Financial Stability: The Need to Resolve the Tension between Liberalization and Prudential Regulation', University of Cambridge Economic and Social Research Council (ESRC) Centre for Business Research, Working Paper 5.

Arner, D., Z. Zhou, M. Busheri, B. F. C. Hsu., J. Lou, and W. Wang (2004). 'Financial Regulation and the WTO: Liberalization and Restructuring in China Two Years Post-Accession', East Asian International Economic Law and Policy Programme (EAIEL) Policy Paper 1, University of Hong Kong.

Barth, J. R., G. Caprio, and R. Levine (2006). *Rethinking Bank Regulation: Till Angel Govern*. New York, NY: Cambridge University Press.

—— J. A. Marchetti, D. E. Nolle, and W. Sawangngoenyuang (2006). 'Foreign Banking: Do Countries' WTO Commitments Match Actual Practices?', Staff Working Paper ERSD-2006-11, World Trade Organization, October.

—— Z. Zhou, D. W. Arner, B. F. C. Hsu, and W. Wang. (2007). *Financial Restructuring and Reform in Post-WTO China*. The Netherlands: Kluwer Law International.

Claessens, S. and T. Glaessner (1998). 'Internationalization of Financial Services in Asia', World Bank Policy Research Working Paper 1911, April.

Dobson, W. and P. Jacquet (1998). *Financial Services Liberalization in the WTO*. Washington, DC: Institute for International Economics.

Hardin, A. and L. Holmes (1997). 'Services Trade and Foreign Direct Investment', Staff Research Paper, Industry Commission. Canberra: Australian Government Publishing Services.

Hoekman, B. (1995). 'Tentative First Steps: An Assessment of the Uruguay Round Agreement on Services', World Bank Policy Research Working Paper 1455, May.

Key, S. (2003). *The Doha Round and Financial Services Negotiations*. Washington, DC: AEI Press.

Marchetti, J. A. (2003). 'What Should Financial Regulators Know About the GATS?', mimeo. Paper presented at the Central Banking seminar, Cambridge University, 7–11.

Mattoo, A. (1998). 'Financial Services and WTO: Liberalization in the Developing and Transition Economies', manuscript. Geneva: World Bank Organization.

——(2000). 'Financial Services and the WTO: Liberalization Commitments of the Developing and Transitional Economies, *World Economy*, 23, March.

McGuire, G. (1998). 'Australia's Restrictions on Trade in Financial Services', Productivity Commission Staff Research Paper, AusInfo. Canberra, November.

——and M. Schuele (2000). 'Restrictiveness of International Trade in Banking Services', in Christopher Findlay and Tony Warren (eds.), *Impediments to Trade in Services: Measurement and Policy Implications*. New York: Routledge.

This page intentionally left blank

PART III

BANK PERFORMANCE

This page intentionally left blank

CHAPTER 18

EFFICIENCY IN BANKING

THEORY, PRACTICE, AND EVIDENCE[1]

JOSEPH P. HUGHES
LORETTA J. MESTER

Introduction

What do commercial banks do? What are the key components of banking technology? What determines whether banks operate efficiently? The literature on financial intermediation suggests that commercial banks, by screening and monitoring borrowers, can solve potential moral hazard and adverse selection problems caused by the imperfect information between borrowers and lenders. From the information obtained from checking account transactions and other sources, banks assess and manage risk, write contracts, monitor contractual performance, and, when required, resolve non-performance problems (Bhattacharya and Thakor, 1993 review the modern theory of financial intermediation).

[1] The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Philadelphia or of the Federal Reserve System.

Banks' ability to ameliorate informational asymmetries between borrowers and lenders and their ability to manage risks are the essence of bank production. These abilities are integral components of bank output and influence the managerial incentives to produce financial services prudently and efficiently. That banks' liabilities are demandable debt gives banks an incentive advantage over other intermediaries. The relatively high level of debt in a bank's capital structure disciplines managers' risk taking and their diligence in producing financial services by exposing the bank to an increased risk of insolvency. The demandable feature of the debt, to the extent that it is not fully insured, further heightens performance pressure and safety concerns by increasing liquidity risk. These incentives tend to make banks good monitors of their borrowers. Hence, the banking relationship can improve the financial performance of bank customers and increase access to credit for firms too informationally opaque to borrow in public debt and equity markets. The uniqueness of bank production, in contrast to the production of other types of lenders, is derived from the special characteristics of banks' capital structure: the funding of informationally opaque assets with demand deposits.[2] (For a discussion of the optimal capital structure of commercial banks, see Calomiris and Kahn, 1991 and Flannery, 1994.)

But banks' ability to perform efficiently—to obtain accurate information concerning its customers' financial prospects and to write effective contracts and to enforce them—depends in part on the property rights and legal, regulatory, and contracting environments in which they operate. Such an environment includes accounting practices, chartering rules, government regulations, and the market conditions (e.g., market power) under which banks operate. Differences in these features across political jurisdictions can lead to differences in the efficiency of banks across jurisdictions.[3] The operating environment can also influence the external and internal mechanisms that discipline bank managers. Internal discipline might be induced or reduced by organizational form, ownership and capital structure, governing boards, and managerial compensation. External discipline might be induced or reduced by government regulation and the safety net, capital

[2] Berlin and Mester (1999) find empirical evidence of an explicit link between banks' liability structure and their distinctive lending behavior. As discussed in Mester (2007), relationship lending is associated with lower loan rates, less stringent collateral requirements, a lower likelihood of credit rationing, contractual flexibility, and reduced costs of financial distress for borrowing firms. Banks' access to core deposits, which are rate inelastic, enable banks to insulate borrowers with whom they have durable relationships from exogenous credit shocks. Mester, Nakamura, and Renault (2007) also find empirical evidence of a synergy between the liability and asset sides of a commercial bank's balance sheet, showing that information on the cash flows into and out of a borrower's transactions account can help an intermediary monitor the changing value of collateral that a small business commercial borrower has posted.

[3] Demirgüç-Kunt, Kane, and Laeven (2007) use a sample of 180 countries to study the external and internal political features that influence the adoption and design of deposit insurance, which, in turn, affect the efficiency of the domestic banking system.

market discipline (takeovers, cost of funds, stakeholders' ability to sell stock (stock price), managerial labor market competition, outside blockholders (equity and debt), and product market competition).[4]

Banking technology and performance

The empirical measurement of banking technology and performance

There are two broad approaches to measuring technology and explaining performance: non-structural and structural. Using a variety of financial ratios that capture various aspects of performance, the non-structural approach compares performance among banks and considers the relationship of performance to investment strategies and other factors such as characteristics of governance. For example, the non-structural approach might investigate technology by asking how performance ratios are correlated with such investment strategies as growing by asset acquisitions and diversifying or focusing the bank's product mix. It looks for evidence of agency problems in correlations of performance ratios and variables characterizing the quality of banks' governance. While informal and formal theories may motivate some of these investigations, no general theory of performance provides a unifying framework for these studies.

The structural approach is choice-theoretic and, as such, relies on a theoretical model of the banking firm and a concept of optimization. The older literature applies the traditional microeconomic theory of production to banking firms. The newer literature views the bank as a financial intermediary that produces informationally intensive financial services and diversifies risks, and combines the theory of financial intermediation with the microeconomics of bank production. This helps guide the choice of outputs and inputs in the bank's production structure. For example, as discussed in Mester (2008), the standard application of efficiency analysis to banking does not allow bank production decisions to affect bank risk. This rules out the possibility that scale- and scope-related improvements in

[4] La Porta, Lopez-de-Silanes, and Shleifer (2002) examine banking systems in 92 countries and find that government ownership is correlated with poorer countries and countries with less developed financial systems, poorer protection of investors' rights, more government intervention, and poorer performance of institutions. They also find that government ownership is associated with higher cost ratios and wider interest rate margins. Aghion, Alesina, and Trebbi (2007) provide evidence that democracy has a positive impact on productivity growth in more advanced sectors of the economy, possibly by fostering entry and competition.

diversification could lower the cost of borrowed funds and induce banks to alter their risk exposure. Also, much of the earlier literature does not account for the bank's role in producing information about its borrowers in its underwriting decisions when specifying the bank's outputs and inputs. An exception is Mester (1992), which directly accounted for banks' monitoring and screening role by measuring bank output treating loans purchased and originated loans as separate outputs entailing different types of screening, and treating loans held on balance sheet and loans sold as separate outputs entailing different types of monitoring.

Banks make choices about their capital structure and the amount of risk to assume, which should be taken into account when modeling bank production. Part of the input prices and part of the output prices a bank faces are not exogenous—the bank makes strategic decisions regarding asset quality and capital structure, which affect the risk premium in its output and input prices. These decisions also relate to how one should view bank performance. In the standard efficiency literature, the bank is assumed to choose a production plan that minimizes costs given its output mix and input prices or that maximizes profits given the prices of its inputs and outputs. In newer research (e.g., Hughes, 1999; Hughes, et al., 1999; Hughes, et al., 2000; and Hughes, Mester, and Moon, 2001) bank managers are modeled as maximizing their utility, which is a function of market value and risk. To the extent that production decisions affect bank risk, they also affect the discount rate applied to evaluating the present value of costs and profit streams. Production decisions that increase expected profit but also increase the discount rate applied to that profit may not increase the bank's market value. In addition, managers may trade off expected return and risk, so that production choices that maximize managers' utility depend not only on the expected profits they generate but also on the variability of the profit stream they generate. Banks with high levels of agency problems between owners and managers might choose utility-maximizing production plans, but these need not be value-maximizing plans if the risk–return trade-offs being made are not efficient.

How one gauges performance in structural models, then, depends on whether one views the bank as minimizing cost, maximizing profits, or maximizing managerial utility. In the last case, one would want to gauge the trade-offs between risk and expected return being made in banks with minimum agency problems between owners and managers—that is, banks with strong corporate controls (see Hughes, Mester, and Moon, 2001). In both the structural and non-structural approaches, the performance metric and the specification of the performance equation reflect implicitly or explicitly an underlying theory of managerial behavior.

As a general specification of the structural and non-structural approaches, let y_i represent the measure of the i^{th} bank's performance. Let $\mathbf{z_i}$ be a vector of variables that capture key components of the i^{th} bank's technology (e.g., output levels and input prices) and $\boldsymbol{\tau}_i$ be a vector of variables affecting the technology (e.g., the ratio

of non-performing to total loans). Jensen and Meckling (1979) add a vector, $\boldsymbol{\theta}_i$, of characteristics of the property rights system, contracting, and regulatory environment in which the i^{th} firm operates (e.g., whether the country has a deposit insurance scheme and the degree of investor protection) and a vector, $\boldsymbol{\phi}_i$, of characteristics of the organizational form and the governance and control environment of the i^{th} firm (e.g., whether the bank is organized as a mutual or stock-owned firm, the degree of product market concentration, and the number of outside directors on its board). When the sample of banks used in the estimation includes financial institutions located in environments with different property rights and contracting environments or with different governance and control structures, estimating this model permits one to investigate how these differences are correlated with differences in bank performance.

Allowing for random error, the performance equation to be estimated takes the form,

$$y_i = f(\mathbf{z}_i, \boldsymbol{\tau}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i|\boldsymbol{\beta}) + \epsilon_i. \quad (1)$$

The specification of the vectors $\mathbf{z}_i$ and $\boldsymbol{\tau}_i$ differs between the structural and non-structural approaches.

The structural approach to bank efficiency measurement: cost minimization, profit maximization, and managerial utility maximization

The *structural approach* usually relies on the economics of cost minimization or profit maximization, where the performance equation denotes a cost function or a profit function. Occasionally, the structural performance equation denotes a production function. While estimating a production function might tell us if the firm is *technically efficient*—that is, if managers organize production so that the firm maximizes the amount of output produced with a given amount of inputs (so that the firm is operating on its production frontier), we are more interested in *economic efficiency*—that is, whether the firm is correctly responding to relative prices in choosing its inputs and outputs, which subsumes technical efficiency.

In the newer literature, the optimization problem is managerial utility maximization, where the manager trades off risk and expected return. The vector $\mathbf{z}$ includes input prices and output prices in a profit function. In the cost function and the non-standard profit function (Humphrey and Pulley, 1997), the vector contains input prices and output levels. In all of these cases, $\boldsymbol{\tau}$ might include controls like non-performing loans to total loans or off-balance sheet assets to total assets.

These functions can also differ by the definition of cost they use: accounting (cash flow) cost excludes the cost of equity capital, while economic cost includes it. The theoretically proper specification of accounting cost is addressed below. The challenge of specifying economic cost is estimating the cost of equity capital. McAllister and McManus (1993) arbitrarily pick the required return and assume it is uniform across banks. Clark (1996) and Fiordelisi (2007) use the capital asset pricing model to estimate it. Fiordelisi (2007) describes the resulting profit function as 'economic value added'.

The structural performance equation can be fitted to the data as an average relationship, which assumes that all banks are equally efficient at minimizing cost or maximizing profit, subject to random error, ϵ_i, which is assumed to be normally distributed. Alternatively, the structural performance equation can be estimated as a *stochastic frontier* to capture best-practice and to gauge inefficiency, the difference between the best-practice performance and achieved performance. Berger and Mester (1997) review the estimation methods. Note that best-practice performance is sometimes called 'potential performance'. However, this is somewhat of an abuse of terms since the best-practice performance does not necessarily represent the best-possible practice, but merely the best practice observed among banks in the sample (see Berger and Mester, 1997, and Mester, 2008).

In the stochastic frontier, the error term, ϵ_i, consists of two components; one is a two-sided random error that represents noise (ν_i), and one is a one-sided error representing inefficiency (μ_i). The stochastic frontier approach disentangles the inefficiency and random error components by making explicit assumptions about their distributions. The inefficiency component measures each bank's extra cost or shortfall of profit relative to the frontier—the best practice performance observed in the sample.[5] Let y_i denote either the cost or profit of firm *i*. The stochastic frontier gives the highest or lowest potential value of y_i given $\mathbf{z_i}$, $\boldsymbol{\tau}_\mathbf{i}$, $\boldsymbol{\phi}_\mathbf{i}$, and $\boldsymbol{\theta}_\mathbf{i}$,

$$y_i = F(\mathbf{z_i}, \boldsymbol{\tau}_\mathbf{i}, \boldsymbol{\phi}_\mathbf{i}, \boldsymbol{\theta}_\mathbf{i}|\boldsymbol{\beta}) + \epsilon_i, \tag{2}$$

where $\epsilon_i \equiv \mu_i + \nu_i$ is a composite error term comprising ν_i, which is normally distributed with zero mean, and μ_i, which is usually assumed to be half-normally distributed and negative when the frontier is fitted as an upper envelope in the case of a profit function and positive when the frontier is fitted as a lower envelope as in the case of a cost function. $\boldsymbol{\beta}$ are parameters of the deterministic kernel, $F(\mathbf{z_i}, \boldsymbol{\tau}_\mathbf{i}, \boldsymbol{\phi}_\mathbf{i}, \boldsymbol{\theta}_\mathbf{i} \mid \boldsymbol{\beta})$, of the stochastic frontier. The i^{th}

[5] Leibenstein (1966) called such inefficiency, which can result from poor managerial incentives or the failure of the labor market to allocate managers efficiently and to weed out incompetent managers, 'X-inefficiency'. Jensen and Meckling (1976) called such inefficiency 'agency costs' and provided a theoretical model of managerial utility maximization to explain how, when incentives between managers and outside stakeholders are misaligned, managers may trade off the market value of their firm to enjoy more of their own private benefits, such as consuming perquisites, shirking, discriminating prejudicially, taking too much or too little risk to enhance their control.

bank's inefficiency is usually estimated by the mean of the conditional distribution of μ_i given ϵ_i, i.e., $E(\mu_i|\epsilon_i)$. The difference between best-practice and achieved performance gauges managerial inefficiency in terms of either excessive cost—*cost inefficiency*—or lost profit—*profit inefficiency*. Expressing the shortfall and excess as ratios of their frontier (best-practice) values yields profit and cost inefficiency ratios. While the fitted stochastic frontier identifies best-practice performance of the banks in the sample, it cannot explain the behavior of inefficient banks. A number of papers have surveyed investigations of bank performance using these concepts: for example, Berger and Humphrey (1997); Berger and Mester (1997); and Berger (2007).

As discussed in Mester (2008), since inefficiency is derived from the regression residual, selection of the characteristics of the banks and the environmental variables to include in the frontier estimation is particularly important. These variables define the peer group that determines best-practice performance against which a particular bank's performance is judged. If something extraneous to the production process is included in the specification, this might lead to too narrow a peer group and an overstatement of a bank's level of efficiency. Moreover, the variables included determine which type of inefficiency gets penalized. If bank location—for example, urban vs. rural, is included in the frontier, then an urban bank's performance would be judged against other urban banks but not against rural banks, and a rural bank's performance would be judged against other rural banks. If it turned out that rural banks are more efficient than urban banks, all else equal, the inefficient choice of location would not be penalized. An alternative to including the variable in the frontier regression is to measure efficiency based on a frontier in which it is omitted and then to see how it correlates with efficiency. Several papers have looked at the correlations of efficiency measures and exogenous factors, including Mester (1993); Mester (1996); Mester (1997); and Berger and Mester (1997). Mester (1997) shows that estimates of bank cost efficiency can be biased if bank heterogeneity is ignored. See also Bos, et al. (2005) on the issue of whether certain differences in the economic environment belong in the definition of the frontier.

Either the average cost function or cost frontier can be used to measure *scale economies*, which refer to how the bank's scale of operations (its size) is related to cost and give a measure of whether the bank is operating at an optimal scale. A bank is operating with scale economies if a 1 percent increase in scale leads to a less than 1 percent increase in cost; it is operating with scale diseconomies if a 1 percent increase in scale leads to a greater than 1 percent increase in costs; it is operating with constant returns to scale if a 1 percent increase in scale leads to a 1 percent increase in cost. *Scope economies* refer to whether the bank is producing the optimal combination of products to minimize cost (or maximize profits). In particular, a bank is operating with scope economies if the cost of producing the bank's product

bundle is less than the cost of separating the bundle into specialized firms. The bank is operating with scope diseconomies if specialized banks could produce the product mix more cheaply.

Typically in the literature, the cost and profit functions or frontiers are measured without considering the bank's capital structure or bank's choice of risk. This is a serious omission since both are important parts of banking technology. Banks' production technologies embody their ability to diversify and offset a variety of risks, and the production decisions managers make reflect their incentives to take on risks as well as to diversify them. Modern banking theory emphasizes managers' contrasting incentives for risk taking. On the one hand, increased risk taking exploits the risk taking subsidy of explicit and implicit, mispriced deposit insurance, while, on the other hand, reduced risk taking protects a bank from costly episodes of financial distress involving liquidity crises, regulatory intervention, and even forfeiture of the bank's valuable charter. For most banks, valuable investment opportunities make trading profitability for reduced risk a value-maximizing strategy. Reducing risk can involve not just producing assets with lower expected profit, but also incurring higher costs to manage risks.

When market-priced risk varies across production plans, the discount rate on profit will also vary across firms so that the production plan that maximizes expected profit may not maximize the discounted value of expected profit. Modeling the behavior of value-maximizing managers requires a more general objective function than profit maximization. Hughes, et al. (1996; 1999; 2000) incorporate risk into managers' choice of production plans by defining managerial utility as a function of profit and the production plan (i.e., the choice of inputs and outputs). Technology defines all feasible production plans. The utility function ranks feasible production plans according to the utility the managers derive from each production plan. Each production plan is linked to a subjective probability distribution of profit by managers' beliefs about the probability distribution of future economic states and how these states interact with feasible production plans to determine profit. Thus, managerial utility expressed as a function of profit and the production plan is equivalent to utility expressed as a function of subjective, conditional probability distributions of profit. Hence, it allows managers to rank production plans not just by their expected profit, the first moment of their distribution, but also by higher moments that capture the risk of production plans.

This managerial utility function is also sufficiently general that it can also account for rankings of production plans that reflect agency problems. To the extent that managers are able to pursue personal objectives that sacrifice firm value, such as empire building and risk avoidance, maximizing utility need not be the same as maximizing value, and the utility function can represent such rankings. Thus, unlike the standard maximum profit function and minimum cost function, this utility framework is able to explain *inefficient* as well as efficient managerial decisions.

To specify the performance equation (1), Hughes, et al. (1996; 1999; 2000) adapt the 'almost ideal demand system' to derive a *utility-maximizing* profit equation and its associated input demand equations. This profit function does not necessarily maximize profit, since it follows from managers' assessment of risk and risk's effect on asset value and perhaps their job security. The profit function also might not represent value-maximizing output production plans or risk-expected return choices, to the extent that there are agency costs and managers are able to pursue non-value-maximizing objectives. Profit maximization (cost minimization) can be tested by noting that the standard translog profit (cost) function and share equations are nested within the model and can be recovered by imposing the parameter restrictions implied by profit maximization (cost minimization) on the coefficients of this adapted system. Hughes, et al. (1996; 1999; 2000) test these restrictions and reject the hypothesis of profit maximization (and cost minimization) in their applications.

Since the utility-maximizing profit function explains inefficient as well as efficient production, it cannot be fitted as a frontier. To gauge inefficiency, Hughes, et al. (1996) and Hughes, Mester, and Moon (2001) estimate a best-practice risk–return frontier and measure inefficiency relative to it. The estimated utility-maximizing profit function yields a measure of expected profit for each bank in the sample, and, when divided by equity capital, the expected profit is transformed into expected return on equity, $E(\pi_i / k_i)$. Each bank's expected (or, predicted) return is a function of its production plan and other explanatory variables. When the estimation of the profit function allows for heteroscedasticity, the standard error of the predicted return (profit), σ_i, a measure of econometric prediction risk, is also a function of the production plan and other explanatory variables and varies across banks in the sample.[6] The estimation of a stochastic frontier similar to (2) gives the highest expected return at any particular risk exposure:

$$E(\pi_i/k_i) = \alpha_0 + \alpha_1 \sigma_i + \alpha_2 \sigma_i^2 - \mu_i + \nu_i, \tag{3}$$

where ν_i is a two-sided error term representing noise, and μ_i is a one-sided error term representing inefficiency. A bank's *return inefficiency* is the difference between its potential return and its noise-adjusted expected return, gauged among

[6] Note that the estimated profit (or return) function resembles a multi-factor model where the factors are the explanatory variables in the profit function. The regression coefficients can be interpreted as marginal returns to the explanatory variables, and the standard error of the predicted return, a function of the variance-covariance matrix of the estimated marginal returns, resembles the variance of a portfolio return. Hughes (1999) and Hughes, Mester, and Moon (2001) report that the regression of *ln* (market value of equity) on *ln* $(E(\pi_i / k_i))$ and $ln(\sigma_i)$ for 190 publicly traded bank holding companies has an R-squared of 0.96, which implies that the production-based measures of expected return and risk explain a large part of a bank's market value.

its peers with the same level of return risk. (Note, however, that if a bank's managers are taking too much or too little risk relative to the value-maximizing amount, this inappropriate level of risk is not taken into account by this measure of inefficiency.)

Koetter (2006) uses the model of managerial utility maximization and the associated measure of risk–return efficiency developed in Hughes, et al. (1996; 1999; 2000) to investigate the efficiency of universal banks in Germany between 1993 and 2004. He compares the measure of return efficiency with cost and profit efficiency estimated by standard formulations and finds evidence that *efficient* banks using a low-risk investment strategy score poorly in terms of standard profit efficiency measures, since they also expect lower profit.

Hughes, Mester, and Moon (2001) take this a step further by recognizing that the utility-maximizing choices of bank managers need not be value maximizing to the extent that there are agency problems within the firm and managers are able to pursue their own, non-value-maximizing objectives. To identify the value-maximizing banks among the set of all banks, they select the quarter of banks in the sample that have the highest predicted return efficiency. These banks are the mostly likely group to be maximizing value or, at least, producing with the smallest agency costs. One can use this set of efficient banks to gauge characteristics of the value-maximizing production technology. For example, mean scale economies across this set of banks would indicate whether there were scale economies as banks expand output along a path that maximizes value. In contrast, mean scale economies across *all* banks would indicate whether there were scale economies as banks expand output along a path that maximizes managers' utility, but this can differ from the value-maximizing expansion path to the extent that managers are able to pursue their own objectives and these objectives differ from those of outside owners.

While the model of managerial utility maximization yields a structural utility-maximizing profit function that includes as special cases the standard maximum profit function and a value-maximizing profit function, it is, nevertheless, based on accounting measures of performance. An alternative model developed by Hughes and Moon (2003) gauges performance using the market value of assets. They develop a *utility-maximizing Q-ratio function* derived from a model where managers allocate the potential (frontier) market value of their firm's assets between their consumption of agency goods (market value inefficiency) and the production of market value, which, given their ownership stake, determines their wealth. The utility function is defined over wealth and the value of agency goods and is conditioned on capital structure, outside blockholder ownership, stock options held by insiders, and other managerial incentive variables. The authors derive a utility-maximizing demand function for market value and for agency goods (inefficiency). Hence, their *Q-ratio* equation is *structural* and, consequently, enjoys the properties of a well-behaved consumer demand

function. The authors use these properties to analyze the relationship between value (or inefficiency) and the proportion of the firm owned by insiders, which is their opportunity cost of consuming agency goods.

The non-structural approach to bank efficiency measurement

The *non-structural approach* to bank performance measurement usually focuses on achieved performance and measures y_i, in equation (1) by a variety of financial ratios—for example, return-on-asset, return-on-equity, or the ratio of fixed costs to total costs. However, some applications have used measures of performance that are based on the market value of the firm (which inherently incorporates market-priced risk)—for example, Tobin's Q ratio (which is the ratio of the market value of assets to the book value of assets), the Sharpe ratio (which measures the ratio of the firm's expected excess return over the risk-free return to the volatility of this excess return—as measured by the standard deviation of the excess return), or an event study's cumulative abnormal return (CAR) (the cumulative error terms of a model predicting banks' market return around a particular event). Other applications have measured performance by an inefficiency ratio obtained by estimating either a non-structural or structural performance equation as a frontier. The non-structural approach then explores the relationship of performance to various bank and environmental characteristics, including the bank's investment strategy, location, governance structure, and corporate control environment. For example, the non-structural approach might investigate technology by asking how performance ratios are correlated with asset acquisitions, the bank's product mix, whether the bank is organized as a mutual or stock-owned firm, and the ratio of outside to inside directors on its board. While informal and formal theories may motivate some of these investigations, no general theory of performance provides a unifying framework.

Using the frontier methods in a non-structural approach, Hughes, et al. (1997) proposed a proxy for Jensen and Meckling's agency cost: a frontier of the market value of assets fitted as a potentially non-linear function of the book-value investment in assets and the book value of assets squared. This frontier gives the highest potential value observed in the sample for any given investment in assets. For any bank, the difference between its highest potential value and its noise-adjusted achieved value represents its lost market value—a proxy for agency cost (*X*-inefficiency). Several studies have used either this systematic lost market value or the resulting noise-adjusted Q ratio to measure performance: Baele, DeJonghe, and Vennet (2007); Hughes, et al. (2003); DeJonghe and Vennet

(2005); Hughes and Moon (2003); Hughes, et al. (1999); and Hughes, Mester, and Moon (2001).

Habib and Ljungqvist (2005) specified an alternative market value frontier as a function of a variety of managerial decision variables, including size, financial leverage, capital expenditures, and advertising expenditures. Thus, the peer grouping on which the frontier is estimated is considerably narrower than the wide grouping based on investment in assets, and inefficient choices of these conditioning values are not accounted for in the measurement of agency costs.

Specifying outputs and inputs in structural models of production

In estimating the standard cost or profit function or the managerial utility maximization model, one must specify the outputs and inputs of bank production. The intermediation approach focuses on the bank's production of intermediation services and the total cost of production, including both interest and operating expenses. Outputs are typically measured by the dollar volume of the bank's assets in various categories. (As mentioned above, an exception is Mester, 1992, which, to account for the bank's screening and monitoring activities, measured outputs as loans previously purchased, which require only monitoring, loans currently originated for the bank's own portfolio, loans currently purchased, and loans currently sold.) Inputs are typically specified as labor, physical capital, deposits and other borrowed funds, and, in some studies, equity capital. While the intermediation approach treats deposits as inputs, there has been some discussion in the literature about whether deposits should be treated as an output since banks provide transactions services for depositors. Hughes and Mester (1993) formulated an empirical test for determining whether deposits act as an input or output. Consider variable cost, VC, which is the cost of non-deposit inputs and is a function of the prices of non-deposit inputs, $\mathbf{w}$, output levels, $\mathbf{q}$, other variables affecting the technology, $\boldsymbol{\tau}$, and the level of deposits, x. If deposits are an input, then $\partial VC/\partial x < 0$: increasing the use of some input should decrease the expenditures on other inputs. If deposits are an output, then $\partial VC/\partial x > 0$: output can be increased only if expenditures on inputs are increased. Hughes and Mester's empirical results indicate insured and uninsured deposits are inputs at banks in all size categories.

Specifying capital structure in performance equations

As discussed above, typically, cost and profit functions are measured without considering the bank's capital structure. However, the newer literature

recognizes the importance of bank managers' choice of risk and capital structure on bank performance. Some of the first structural models to include equity capital as an input are Hancock (1985); Hancock (1986); McAllister and McManus (1993); Hughes and Mester (1993); Clark (1996); and Berger and Mester (1997).

As discussed in Hughes and Mester (1993); Hughes (1999); Mester (2008); and Berger and Mester (1997), a bank's insolvency risk depends not only on the riskiness of its portfolio but on the amount of financial capital it has to absorb losses. Insolvency risk affects bank costs and profits via the risk premium the bank has to pay for uninsured debt, through the intensity of risk management activities the bank undertakes, and through the discount rate applied to future profits. A bank's capital level also directly affects costs by providing an alternative to deposits as a funding source for loans.

Most studies use the cash flow (accounting) concept of cost, which includes the interest paid on debt (deposits) but not the required return on equity, as opposed to economic cost, which includes the cost of equity. Failure to include equity capital among the inputs can bias efficiency measurement. If a bank were to substitute debt for some of its financial equity capital, its accounting (cash flow) costs could rise, making the less-capitalized bank appear to be more costly than a well-capitalized bank. To solve this problem, the level of equity capital can be included as a quasi-fixed input in the cost function. The resulting cost function captures the relationship of cash flow cost to the level of equity capital, and the (negative) derivative of cost with respect to equity capital—the amount by which cash flow cost is reduced if equity capital is increased—gives the shadow price of equity. The shadow price of equity will equal the market price when the amount of equity minimizes cost or maximizes profit. Even when the level of equity does not conform to these objectives, the shadow price nevertheless provides a measure of its opportunity cost. Hughes, Mester, and Moon (2001) find that the mean shadow price of equity for small banks is significantly smaller than that of larger banks. This suggests that smaller banks over-utilize equity relative to its cost-minimizing value, perhaps to protect charter value. On the other hand, larger banks appear to under-utilize equity relative to its cost-minimizing value, perhaps to exploit a deposit subsidy and the subsidy due to the 'too-big-to-fail' doctrine.

Specifying output quality in the performance equation

In measuring efficiency, one should control for differences in output quality to avoid labeling unmeasured differences in product quality as differences in efficiency. Controls for loan quality—for example, non-performing loans to total loans by loan category or loan losses, are sometimes included in the cost or profit

frontier as controls (see Mester, 2008, for further discussion). As discussed by Berger and Mester (1997), whether it is appropriate to include non-performing loans or loan losses in the cost or profit function depends on the extent to which these variables are exogenous. They would be exogenous if caused by economic shocks (bad luck), but could be endogenous to the extent that management is inefficient or has made a conscious decision to cut short-run expenses by cutting back on loan origination and monitoring resources. Berger and Mester (1997) attempt to solve this problem by using the ratio of non-performing loans to total loans in the bank's state as a control variable. This state average would be nearly entirely exogenous to any one bank, but can control for negative shocks that affect bank output quality.

The variable, non-performing loans, can also play a role as a quasi-fixed 'input' whose quantity rather than price is included in the performance equation. As such, its 'cost' is excluded from the performance metric, either cost or profit. Its price is the expected loan-loss rate. Hence, when the cost of non-performing loans—that is, loan losses, is excluded from the performance measure, a case can be made for including the level of non-performing loans, and when the performance measure is net of loan losses, the logic suggests that the loss rate be included in the specification of the performance equation.

Applications of the structural approach

Performance in relation to organizational form, governance, regulation, and market discipline

An increasing number of papers using structural models are exploring the importance of governance and ownership structure to the performance of banks. The structural model is first used to obtain a frontier-based measure of inefficiency. Then inefficiency is regressed on a set of explanatory variables.

Using confidential regulatory data on small, closely held commercial banks, DeYoung, Spong, and Sullivan (2001) use a stochastic frontier to measure banks' profit efficiency. They find banks that hire a manager from outside the group of controlling shareholders perform better than those with owner-managers; however, this result depends on motivating the hired managers with sufficient holdings of stock. They calculate an optimal level of managerial ownership that minimizes profit inefficiency. Higher levels of insider holdings lead to entrenchment and lower profitability.

Berger and Hannan (1998) consider the relationship of bank cost efficiency, estimated by a stochastic frontier, to product market discipline, gauged by a Herfindahl-Hirschman Index (HHI) of market power. They find that the reduced discipline of concentrated markets is associated with a loss of cost efficiency far more significant than any welfare loss due to monopoly pricing.

DeYoung, Hughes, and Moon (2001) use the model of managerial utility maximization developed by Hughes, et al. (1996; 2000) to estimate expected return and return risk. Using these values, they estimate a stochastic risk–return frontier as in equation (3) to obtain each bank's return inefficiency. They consider how banks' supervisory CAMELS ratings[7] are related to their size, their risk–return choice, and their return inefficiency. They find that the risk–return choices of efficient banks are not related to their supervisory rating, while higher-risk choices of inefficient banks are penalized with poorer ratings. Moreover, the risk–return choices of large inefficient banks are held to a stricter standard than smaller banks and large efficient banks.

Two studies by Mester (1991; 1993) investigate differences in scale and scope measures for stock-owned and mutual savings and loans by estimating average cost functions. She finds evidence of agency problems at mutual savings and loans, as evidenced by diseconomies of scope, prior to the industry's deregulation, and evidence that these agency costs were lessened after the deregulation in the mid-1980s.

Using data for the period 1989–96, Altunbas, Evans, and Molyneux (2001) estimate separate and common frontiers for three organizational forms in German banking: private commercial, public (government-owned) savings, and mutual cooperative banks. They argue that the same technology of intermediation is available to all so that the choice of technology is a management decision whose efficiency should be compared among all types of forms. The private sector appears to be less profit and cost efficient than the other two sectors. These results are especially clear in the case of the common frontier, but they are also obtained from the estimation of separate frontiers.

Uncovering evidence of scale economies by accounting for risk and capital structure

Berger and Mester (1997) use data on the almost 6,000 US commercial banks that were in continuous existence over the six-year period 1990–5. They estimate scale economies, cost X-efficiency, and profit X-efficiency for banks in different size categories based on their preferred model that incorporates asset quality, financial capital, and off-balance sheet assets and based on several alternative specifications.

[7] The components of a bank's condition that are assessed: (C)apital adequacy, (A)sset quality, (M)anagement, (E)arnings, (L)iquidity, and (S)ensitivity to market risk.

In the preferred model, which includes financial capital, they find significant cost scale economies for banks in each size class: the typical bank would have to be two to three times larger in order to maximize cost scale efficiency for its product mix and input prices.

Hughes and Mester (1998) use 1989 and 1990 data on US banks with assets over $1 billion and estimate cost function conditioned on the level of financial capital. They find that banks do not hold the cost-minimizing level of capital and that the level of capitalization increases less than proportionately with assets. They find significant scale economies across banks of all size in the sample.

Hughes, Mester, and Moon (2001) apply a model of managerial utility to data on US bank holding companies to consider how incorporating capital structure and endogenous risk taking into the production model affects the ability of the empirical investigator to detect scale economies. For example, better diversification may lead to a lower cost of risk and an incentive to increase risk taking for greater profitability. The increased risk taking may be costly. If larger banks are better diversified and more risky than smaller banks, this source of scale economies may be hard to detect without accounting for endogenous risk taking: the increase in cost due to the increased risk taking can lead to the conclusion that there are no economies of scale. The authors provide evidence that better diversification is associated with larger scale economies, and increased risk taking and inefficiency are related to smaller scale economies.

Bossone and Lee (2004) use the Hughes and Mester (1998) and Hughes, Mester, and Moon (2001) methodologies to study the relationship between productive efficiency and the size of the financial system. Using data on 875 commercial banks from seventy-five countries, they estimate a cost function and measure scale economies allowing for banks' endogenous choice of risk and financial capital. Consistent with the results from Hughes and Mester (1998) and Hughes, Mester, and Moon (2001), they find significant scale economies that are increasing with the size of the financial system. They also find that small banks in larger financial systems are more cost efficient than those in small financial systems. They interpret their findings as evidence of what they call 'systemic scale economies'.

Berger and Mester (2003) investigate cost and profit productivity, where productivity is measured as a combination of technological change (i.e., changes in the best-practice frontier) and changes in inefficiency, holding constant the exogenous environmental variables. (This discussion is taken largely from Mester, 2008.) They find that during 1991–7, cost productivity in the banking industry worsened while profit productivity improved substantially and concluded this was because revenue-based productivity changes are not accounted for in measuring cost productivity. Banks have been offering wider varieties of financial services and have been providing additional convenience, which may have raised costs but also raised revenues by more than the cost increases. They also found that banks involved in merger activity might be responsible for their main findings.

The merging banks had greater cost productivity deterioration and profit productivity improvements than other banks. Merging banks may have also improved their profit performance, on average, by shifting their portfolios into investments with higher risk and higher expected return to take advantage of the diversification gains from mergers, as suggested by the work of Hughes, et al. (1996) and Hughes, Mester, and Moon (2001).

Applications of the non-structural approach

Measuring the value of investment opportunities ('charter value')

The value of a bank's investment opportunities is often measured by Tobin's Q ratio; however, in the presence of agency cost the Q ratio captures only the ability of the incumbent managers to exploit these opportunities. Ideally, the value of investment opportunities should be gauged independently of the ability and actions of the current management. Hughes, et al. (1997) and Hughes, et al. (2003) propose a measure based on fitting a stochastic frontier to the market value of assets as a function of the book value of assets and variables characterizing the market conditions faced by banks. These conditions include an HHI of market power and the macroeconomic growth rate. The fitted frontier gives the highest potential value of a bank's assets in the markets in which it operates. Thus, this potential value is conditional on the location of the bank and represents the value the bank would fetch in a competitive auction. Hughes, et al. (1997) define this value as the bank's 'charter value'—its value in a competitive auction.

Measuring the performance of business and capital strategies

Several papers have used the non-structural performance equation to examine the relationship between bank value and bank capital structure. Hughes, et al. (1997) regress performance measured by Tobin's Q ratio and market value inefficiency on a number of variables characterizing bank production. Calomiris and Nissim (2007) regress the ratio of the market value of equity to its book value on a similar list of variables. De Jonghe and Vennet (2005) apply the market value frontier of Hughes, et al. (1997) to derive a noise-adjusted measure of Tobin's Q ratio, which

they use to evaluate how leverage and market power are related to value. All three studies find evidence that banks follow dichotomous strategies for enhancing value as predicted by Marcus (1984): a lower risk, lower leverage strategy, and a higher risk, higher leverage strategy.

Relationship of ownership structure to bank value

Jensen and Meckling (1976) defined agency cost as the difference in value of a firm owned entirely by its manager (so that there are no agency problems) and one where the manager does not own all of the firm. Since firms with no agency costs should outperform those with agency problems, some studies have sought evidence of agency costs by looking for a correlation between firm value measured by Tobin's Q ratio and variables characterizing potential agency problems, such as the proportion of the firm owned by managers and the proportion owned by outside blockholders.

In an influential study, Morck, Shleifer, and Vishny (1988) hypothesized that managerial ownership creates two contrasting incentives: a higher ownership stake, first, better aligns the interests of managers and outside owners and, second, enhances managers' control over the firm and makes it harder for managers to be ousted when they are not efficient. Measuring performance by Tobin's Q ratio, these authors provide evidence that the so-called 'alignment-of-interests' effect dominates the entrenchment effect at lower levels of managerial ownership, while the entrenchment effect dominates over a range of higher levels.

Studies that attempt to measure the *net* effect of the alignment and entrenchment effects on firm valuation cannot identify these effects individually—only their sum in the form of the sign of a regression coefficient or a derivative of a regression equation. Adams and Santos (2006) cleverly isolate the entrenchment effect by considering how the proportion of a bank's common stock controlled but not owned by the bank's own trust department is statistically related to the bank's economic performance. The voting rights exercised by management through the trust department enhance management's control over the bank but do not align their interests with outside shareholders', since the beneficiaries of the trusts, not the managers, receive the dividends and capital gains and losses.

Caprio, Laeven, and Levine (2003) study the effect of ownership, shareholder protection laws, and supervisory and regulatory policies on the valuations of banks around the world. The authors construct a database of 244 banks—in each of forty-four countries. They measure performance by Tobin's Q ratio and by the ratio of the market value of equity to the book value of equity. They find evidence that banks in countries with better protection of minority shareholders are more highly valued; bank regulations and supervision have no significant effect on bank value; the degree of cash flow rights of the largest owner has a significant positive effect on

bank value; and an increase in ownership concentration has a larger positive effect on valuation when the legal protection of minority shareholders is weak.

Relationship of mergers and takeovers to bank value

Brook, Hendershott, and Lee (1998) examine the stock market reaction to the passage of the Interstate Banking and Branching Efficiency Act (IBBEA) of 1994. They find significantly positive abnormal returns that are negatively related to a bank's prior performance. Apparently, the increased probability of a takeover following the passage of IBBEA improves the value of underperforming banks more than better performing banks. This increase in value is offset among banks whose managers show evidence of entrenchment, such as higher insider ownership, lower outside blockholder ownership, and less independent boards.

If the threat of a takeover disciplines managers and improves profitability, differences in takeover restrictions across states imply differences in the threat of a takeover. Schranz (1993) finds that banks in states with a more active takeover market are more profitable than banks in states restricting takeover activity.

Hughes, et al. (2003) examine US bank holding companies and find evidence of managerial entrenchment among banks with higher levels of insider ownership, more valuable growth opportunities, poorer financial performance, and smaller asset size. When managers are not entrenched, asset acquisitions and sales are associated with reduced market value inefficiency. When managers are entrenched, sales are associated with smaller reductions in inefficiency, while acquisitions are associated with greater inefficiency.

DeLong (2001) studies 280 domestic US bank mergers from 1988 to 1995. Gauging performance by the CARs of the mergers, she finds that mergers that focus activity and geography increase shareholder value, while diversifying mergers do not.

Conclusions

Great strides have been made in the theory of bank technology in terms of explaining banks' comparative advantage in producing informationally intensive assets and financial services and in diversifying or offsetting a variety of risks. Great strides have also been made in explaining sub-par managerial performance in terms of agency theory and in applying these theories to analyze the particular

environment of banking. In recent years, the empirical modeling of bank technology and the measurement of bank performance have begun to incorporate these theoretical developments and yield interesting insights that reflect the unique nature and role of banking in modern economies.

This new literature recognizes that the choice of risk influences banks' production decisions, including the mix of assets, asset quality, off-balance sheet hedging activities, capital structure, debt maturity, and resources allocated to risk management, and so, in turn, affects banks' cost and profitability. Measures of bank performance should take account of this endogeneity. The estimation of structural models that incorporate managerial preferences for expected return and risk have uncovered significant scale economies in banking, a finding that differs from the earlier literature but accords with the consolidation of the banking industry that has been occurring worldwide.

Performance studies based on structural models of managerial utility maximization, as well as those based on non-structural models of bank production, have incorporated variables designed to capture incentive conflicts between managers and outside stakeholders. These studies have shown that factors associated with enhanced market discipline are also associated with improved bank performance.

References

Adams, R. B. and Santos, J. A. C. (2006). 'Identifying the Effect of Managerial Control on Firm Performance', *Journal of Accounting and Economics*, 41: 55–85.

Aghion, P., Alesina, A., and Trebbi, F. (2007). 'Democracy, Technology, and Growth', Working Paper (May), Department of Economics, Harvard University.

Altunbas, Y., Evans, L., and Molyneux, P. (2001). 'Bank Ownership and Efficiency', *Journal of Money, Credit, and Banking*, 33: 926–54.

Baele, L., De Jonghe, O., and Vander Vennet, R. (2007). 'Does the Stock Market Value Bank Diversification?', *Journal of Banking & Finance*, 31: 1999–2023.

Berlin, M. and Mester, L. J. (1999). 'Deposits and Relationship Lending', *Review of Financial Studies*, 12: 579–607.

Berger, A. N. (2007). 'International Comparisons of Banking Efficiency', *Financial Markets, Institutions & Instruments*, 16: 119–44.

——and Hannan, T. H. (1998). 'The Efficiency Cost of Market Power in the Banking Industry: A Test of the "Quiet Life" and Related Hypotheses', *Review of Economics and Statistics*, 80: 454–65.

——and Humphrey, D. B. (1997). 'Efficiency of Financial Institutions: International Survey and Directions for Future Research', *European Journal of Operational Research*, 98: 175–212.

——and Mester, L. J. (1997). 'Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions', *Journal of Banking & Finance*, 21: 895–947.

——— ——— (2003). 'Explaining the Dramatic Changes in Performance of US Banks: Technological Change, Deregulation, and Dynamic Changes in Competition', *Journal of Financial Intermediation*, 12: 57–95.

Bhattacharya, S. and Thakor, A. (1993). 'Contemporary Banking Theory', *Journal of Financial Intermediation*, 3: 2–50.

Bos, J. W. B., Heid F., Koetter, M., Kolari, J. W., and Kool, C. J. M. (2005). 'Inefficient or Just Different? Effects of Heterogeneity on Bank Efficiency Scores', Deutsche Bundesbank Discussion Paper 2.

Bossone, B. and Lee, J.-K. (2004). 'In Finance, Size Matters: The "Systemic Scale Economies" Hypothesis', IMF Staff Papers, 51: 1.

Brook, Y., Hendershott, R., and Lee, D. (1998). 'The Gains from Takeover Deregulation: Evidence from the End of Interstate Banking Restrictions', *Journal of Finance*, 53: 2185–204.

Calomiris, C. W. and Kahn, C. M. (1991). 'The Role of Demandable Debt in Structuring Optimal Banking Arrangements', *American Economic Review*, 70: 312–26.

——— and Nissim, D. (2007). 'Activity-Based Valuation of Bank Holding Companies', Working Paper 12918, National Bureau of Economic Research.

Caprio, G., Laeven, L., and Levine, R. (2003). 'Governance and Bank Valuation', Working Paper 10158, National Bureau of Economic Research.

Clark, J. (1996). 'Economic Cost, Scale Efficiency and Competitive Viability in Banking', *Journal of Money, Credit, and Banking*, 28: 342–64.

DeJonghe, O. and Vander Vennet, R. (2005). 'Competition versus Agency Costs: An Analysis of Charter Values in European Banking', Working Paper, Ghent University.

DeLong, G. L. (2001). 'Stockholder Gains from Focusing versus Diversifying Bank Mergers', *Journal of Financial Economics*, 59: 221–52.

Demirgüç-Kunt, A., Kane, E. J. and Laeven, L. (2007). 'Determinants of Deposit-Insurance Adoption and Design', National Bureau of Economic Research Working Paper 12862 (January).

DeYoung, R. E., Hughes, J. P., and Moon, C.-G. (2001). 'Efficient Risk-Taking and Regulatory Covenant Enforcement in a Deregulated Banking Industry', *Journal of Economics & Business*, 53: 255–82.

——— Spong, K., and Sullivan, R. J. (2001). 'Who's Minding the Store? Motivating and Monitoring Hired Managers at Small, Closely Held Commercial Banks', *Journal of Banking & Finance*, 25: 1209–43.

Fiordelisi, F. (2007). 'Shareholder Value Efficiency in European Banking', *Journal of Banking & Finance*, 31: 2151–71.

Flannery, M. J. (1994). 'Debt Maturity and the Deadweight Cost of Leverage: Optimally Financing Banking Firms', *American Economic Review*, 84: 320–31.

Habib, M. A. and Ljungqvist, A. (2005). 'Firm Value and Managerial Incentives: A Stochastic Frontier Approach', *Journal of Business*, 78: 2053–93.

Hancock, D. (1985). 'The Financial Firm: Production with Monetary and Nonmonetary Goods', *Journal of Political Economy*, 93: 859–80.

——— (1986). A Model of the Financial Firm with Imperfect Asset and Deposit Liabilities, *Journal of Banking & Finance* 10, 37–54.

Hughes, J. P. (1999). Incorporating Risk into the Analysis of Production, Presidential Address to the Atlantic Economic Society, *Atlantic Economic Journal* 27, 1–23.

HUGHES, J. P., LANG, W. W., MESTER, L. J., and MOON, C.-G. (1996). 'Efficient Banking under Interstate Branching', *Journal of Money, Credit, and Banking*, 28: 1045–71.

——————(1999). 'The Dollars and Sense of Bank Consolidation', *Journal of Banking & Finance*, 23: 291–324.

——————(2000). 'Recovering Risky Technologies Using the Almost Ideal Demand System: An Application to U.S. Banking', *Journal of Financial Services Research*, 18: 5–27.

——————and PAGANO, M. (2003). 'Do Bankers Sacrifice Value to Build Empires? Managerial Incentives, Industry Consolidation, and Financial Performance', *Journal of Banking & Finance*, 27: 417–47.

————MOON C.-G., and PAGANO, M. (1997). 'Measuring the Efficiency of Capital Allocation in Commercial Banking', Working Paper 98-2, Federal Reserve Bank of Philadelphia (revised as Working Paper 2004-1, Rutgers University Economics Department).

——and MESTER, L. J. (1993). 'A Quality and Risk-Adjusted Cost Function for Banks: Evidence on the 'Too-Big-to-Fail' Doctrine', *Journal of Productivity Analysis*, 4: 293–315.

————(1998). 'Bank Capitalization and Cost: Evidence of Scale Economies in Risk Management and Signaling', *Review of Economics and Statistics*, 80: 314–25.

————and MOON C.-G. (2001). 'Are Scale Economies in Banking Elusive or Illusive? Evidence Obtained by Incorporating Capital Structure and Risk-Taking into Models of Bank Production', *Journal of Banking & Finance*, 25: 2169–208.

——and MOON C.-G. (2003). 'Estimating Managers' Utility-Maximizing Demand for Agency Goods', Working Paper 2003-24, Department of Economics, Rutgers University.

HUMPHREY, D. B. and PULLEY, L. B. (1997). 'Banks' Responses to Deregulation: Profits, Technology, and Efficiency', *Journal of Money, Credit, and Banking*, 29: 73–93.

JENSEN, M. C. and MECKLING, W. H. (1976). 'Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure', *Journal of Financial Economics*, 5: 305–60.

————(1979). 'Rights and Production Functions: An Application to Labor-Managed Firms and Codetermination', *Journal of Business*, 52: 469–506.

KOETTER, M. (2006). 'The Stability of Efficiency Rankings When Risk-Preferences and Objectives are Different', Deutsche Bundesbank Discussion Paper 08/2006, Series 2, Banking and Financial Studies.

LA PORTA, R., LOPEZ-DE-SILANES, F., and SHLEIFER, A. (2002). 'Government Ownership of Banks', *Journal of Finance*, 57: 265–301.

LEIBENSTEIN, H. (1966). 'Allocative Efficiency vs. "X-Efficiency"', *American Economic Review*, 56: 392–415.

MARCUS, A. J. (1984). 'Deregulation and Bank Financial Policy', *Journal of Banking & Finance*, 8: 557–65.

MCALLISTER, P. H. and MCMANUS, D. (1993). 'Resolving the Scale Efficiency Puzzle in Banking', *Journal of Banking & Finance*, 17: 389–406.

MESTER, L. J. (1991). 'Agency Costs among Savings and Loans', *Journal of Financial Intermediation*, 1: 257–78.

——(1992). 'Traditional and Nontraditional Banking: An Information-Theoretic Approach', *Journal of Banking & Finance*, 16: 545–66.

——(1993). 'Efficiency in the Savings and Loan Industry', *Journal of Banking & Finance*, 17: 267–86.

——(1996). 'A Study of Bank Efficiency Taking into Account Risk Preferences', *Journal of Banking & Finance*, 20: 1025–45.
——(1997). 'Measuring Efficiency at U.S. Banks: Accounting for Heterogeneity is Important', *European Journal of Operational Research*, 98: 230–42.
——(2007). 'Some Thoughts on the Evolution of the Banking System and the Process of Financial Intermediation', *Federal Reserve Bank of Atlanta Economic Review*, I–II: 67–75.
——(2008). 'Optimal Industrial Structure in Banking', in *Handbook of Financial Intermediation*, Boot, A. and Thakor, A. (eds.), Amsterdam: North-Holland/Elsevier.
——Nakamura, L. I., and Renault, M. (2007). 'Transactions Accounts and Loan Monitoring', *Review of Financial Studies*, 20: 529–56.
Morck, R., Shleifer, A., and Vishny, R. W. (1988). 'Management Ownership and Market Valuation: An Empirical Analysis', *Journal of Financial Economics*, 20: 293–315.
Schranz, M. S. (1993). 'Takeovers Improve Firm Performance: Evidence from the Banking Industry', *Journal of Political Economy*, 101: 299–326.

CHAPTER 19

TECHNOLOGICAL CHANGE, FINANCIAL INNOVATION, AND DIFFUSION IN BANKING

W. SCOTT FRAME
LAWRENCE J. WHITE[1]

Introduction

The commercial banking business has changed dramatically over the past twenty-five years, due in large part to technological change.[2] Advances in telecommunications,

[1] The views expressed in this chapter are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Atlanta, the Federal Reserve System, or their staffs. The authors thank Pam Frisbee for research assistance. Helpful comments were provided by Allen Berger, Jose Negrin, and seminar participants at the Banco de Mexico and Autonomous University of Mexico.

[2] Restrictions on commercial banks' ability to diversify geographically and across product space were also significantly relaxed during this time, especially in the United States. This trend has

information technology, and financial theory and practice have jointly transformed many of the relationship-focused intermediaries of yesteryear into data-intensive risk management operations of today. Consistent with this, we now find many commercial banks embedded as part of global financial institutions that engage in a wide variety of financial activities.

To be more specific, technological changes relating to telecommunications and data processing have spurred financial innovations that have altered bank products and services and production processes. For example, the ability to use applied statistics cost effectively (via software and computing power) has markedly altered the process of financial intermediation. Retail loan applications are now routinely evaluated using credit scoring tools, rather than using human judgment. Such an approach makes underwriting much more transparent to third parties and hence facilitates secondary markets for retail credits (e.g., mortgages and credit card receivables) via securitization.[3] Statistically based risk measurement tools are also used to measure and manage other types of credit risks—as well as interest rate risks—on an ongoing basis across entire portfolios. Indeed, tools like 'value-at-risk' are even used to determine the appropriate allocation of risk-based capital for actively managed (trading) portfolios.

This chapter will describe how technological change has spurred financial innovations that have driven the aforementioned changes in commercial banking over the past twenty-five years. In this respect, our survey is similar to that of Berger (2003).[4] However, our analysis distinguishes itself by reviewing the literature on a larger number of new banking technologies and synthesizing these studies in the context of the broader economics literature on innovation. In this way, the chapter is more like our own previous survey of empirical studies of financial innovation (Frame and White, 2004). We note that this survey is US-centric, owing to our own experiences, the fact that many financial innovations originate in the US, and that most studies of such innovations rely on US data. Before proceeding, it will be helpful to understand better what is meant by financial innovation.

significantly reinforced technological change in terms of driving the observed evolution of commercial banking over the past 25 years.

[3] There is also a secondary market for wholesale loans to large corporations via a loan syndication process. This market has also benefited from securitization through the market for 'collateralized loan obligations' or CLOs, which are a type of 'collateralized debt obligation' or CDO. CDOs are discussed further below.

[4] See also Berger, Kashyap, and Scalise (1995) for discussion of the role of technological and regulatory changes in transforming the US banking industry.

Background: The role of finance and financial innovation

As noted by Merton (1992: 12), the primary function of a financial system is to facilitate the allocation and deployment of economic resources, both spatially and across time, in an uncertain environment. This function encompasses a payments system with a medium of exchange; the transfer of resources from savers to borrowers; the gathering of savings for pure time transformation (i.e., consumption smoothing); and the reduction of risk through insurance and diversification.

The operation of a financial system involves real resource costs (labor, materials, and capital) employed by financial intermediaries (e.g., commercial banks) and by financial facilitators (e.g., mortgage brokers). Much of these resources are expended in the data collection and analyses in which financial market participants engage so as to deal with problems of asymmetric information. There are also uncertainties about future states of the world that generate risks, which for risk-averse individuals represent costs. In this environment, new financial products and services that can better satisfy financial system participants' demands should generally be welcomed by those participants.

Hence, we define a financial innovation as something new that reduces costs, reduces risks, or provides an improved product/service/instrument that better satisfies financial system participants' demands. Financial innovations can be grouped as new products (e.g., subprime mortgages) or services (e.g., Internet banking); new production processes (e.g., credit scoring); or new organizational forms (e.g., Internet-only banks).[5]

The centrality of finance in an economy and its importance for economic growth (e.g., Levine, 1997) naturally raises the importance of financial innovation—and its diffusion. Since finance is a facilitator of virtually all production activity and much consumption activity, improvements in the financial sector will have direct positive ramifications throughout an economy. Further, since better finance can encourage more saving and investment and can also encourage better (more productive) investment decisions, these indirect positive effects from financial innovation add further to its value for an economy. The importance of financial innovation has been discussed in a number of articles, most notably: Van Horne (1985); Miller (1986); Miller (1992); Merton (1992); Merton (1995); and Tufano (2003).

Given its importance, an understanding of the conditions that encourage innovation would appear to be worthwhile. After all, observed streams of innovations are clearly not uniform across all enterprises, across all industries, or across all time periods. The general innovation literature in economics has sought to uncover the

[5] Of course, if a new intermediate product or service is created and used by banks, it may then become part of a new financial production process.

environmental conditions that affect the stream of innovations—focusing on hypotheses concerning roughly five structural conditions: (1) the market power of enterprises; (2) the size of enterprises; (3) technological opportunity; (4) appropriability; and (5) product market demand conditions.[6] Of course, when environmental changes occur, we expect to observe an initial wave of financial innovations followed by a new equilibrium flow consistent with the new environmental conditions. Over the past twenty-five years, each of these environmental conditions (1–5) was markedly altered—resulting in substantial changes to the commercial banking industry.

Furthermore, as we noted in our earlier review article (Frame and White, 2004), there has been a surprising dearth of empirical studies that test hypotheses with respect to financial innovation in general. This is especially true for hypotheses that focus on the structural conditions that encourage innovation.[7] Instead, the comparatively few empirical studies that have been done tend to focus on the characteristics of users/adopters of innovations—sometimes on a cross-sectional basis and other times in the context of the diffusion of the innovation. In surveying the literature in preparation for this chapter, we find that more empirical studies have appeared, but the field is still relatively sparse, and the studies still focus largely on the characteristics of users/adopters. This finding represents a supplementary contribution of this chapter.

Financial innovation and banking: 1980–2005

In this section, we survey the literature pertaining to several specific financial innovations appearing over the past twenty-five years or so that were specifically driven by technological change. We have organized our discussion along the lines of the three major categories that we described above: new products and services; new production processes; and new organizational forms.

Products

Mortgage loans are one suite of products that have experienced a great deal of change over the past twenty-five years in the US. In 1980, long-term fully amortizing

[6] See Cohen and Levin (1989) and Cohen (1995) for comprehensive surveys of this literature. The first two hypotheses are associated with Schumpeter (1950).

[7] We previously identified only two papers that tested hypotheses concerning structural conditions that encourage financial innovation (Ben-Horim and Silber, 1977; and Lerner, 2002). Lerner (2006) has made a more recent contribution.

fixed-rate mortgages were the norm; and this product was offered primarily by thrift institutions. Moreover, these loans required substantial down payments and a good credit history; and the accumulated equity was relatively illiquid.

These characteristics have markedly evolved. The first big change occurred in the early 1980s, with the widespread introduction of various types of adjustable rate mortgages (ARMs), which had previously been banned by federal regulators.[8] The Tax Reform Act of 1986, which ended federal income tax deductions for non-mortgage consumer debt, spurred substantial growth in home equity lending.[9] One mortgage innovation more directly tied to technological change is subprime lending, which was originally predicated on the use of statistics for better risk measurement and risk-based pricing to compensate for these higher risks. However, the subprime mortgage crisis has uncovered significant shortcomings in the underlying statistical models.

Subprime mortgages

Subprime mortgage lending, broadly defined, relates to borrowers with poor credit histories (e.g., a 'FICO score'[10] below 620) and/or high leverage as measured by either debt/income (personal leverage) or loan-to-value (property leverage). This market grew rapidly in the US during the first decade of the twenty-first century—averaging about 20 percent of residential mortgage originations between 2004 and 2006. At the end of 2007, subprime mortgages outstanding stood at $940 billion, down from over $1.2 trillion outstanding the previous year (Inside Mortgage Finance, 2008).

Subprime mortgage lending acts to expand the pool of potential homeowners and helped to lead the US to a record homeownership rate in 2004 of 69.2 percent—even in the face of declining housing affordability in many areas of the country. On the other hand, subprime mortgages typically come with more onerous terms, such as higher interest rates and prepayment penalties. Hence, there is some concern that subprime lending can be 'predatory' in nature, especially since lower-income and/or minority households are much more likely to have subprime mortgages. The wave of US subprime mortgage defaults (and associated foreclosures) during 2007–8 has led to a very public discussion about the social benefits and costs of subprime lending and about the manner in which such loans are marketed and financed.

The significant spillover effects of the subprime mortgage crisis on global credit markets has also led to serious questions about the financial markets' dependence on applied statistics (including the choice of historical time frame for calibration) as the

[8] See Strunk and Case (1988: chap. 5); and White (1991: 65) for further discussion.

[9] Manchester and Poterba (1989) report that second mortgages accounted for 3.6% of residential mortgage debt outstanding at the beginning of the 1980s and quickly rose to 10.8% by the end of 1987.

[10] FICO is a registered trademark of Fair Isaac Corporation.

basis of risk measurement and management as well as the construction of increasingly complex structured finance products. As discussed below, these represent important examples of process financial innovations for commercial banks in recent years.

Prior to the crisis, some research sought to explain the existence and efficiency of the subprime mortgage market. Lax, et al. (2004) characterize subprime borrowers—finding that (relative to prime borrowers) they are more likely to have poor credit, lower-incomes, less education, and belong to minority groups. Chomsisengphet and Pennington-Cross (2006) provide several stylized facts about subprime mortgage loans over time—specifically, borrower credit quality, interest rates, downpayment requirements, and the presence of prepayment penalties. Crews-Cutts and Van Order (2005) explain various stylized facts pertaining to subprime loan pricing and performance in the context of financial contracting theory. Chinloy and Macdonald (2005) discuss how the subprime market helps to complete the credit supply schedule and therefore enhance social welfare, while Nichols, Pennington-Cross, and Yezer (2005) explain why prime and subprime mortgage markets are distinct and not continuous. Other papers look at the geographic distribution of subprime borrowers generally (Calem, Gillen, and Wachter, 2004) and the incidence of prepayment penalties particularly (Farris and Richardson, 2004). Finally, there are a number of papers that study how local predatory lending laws affect subprime mortgage credit supply (e.g., Elliehausen and Staten, 2004; Harvey and Nigro, 2003; Harvey and Nigro, 2004; Quercia, Stegman, and Davis, 2004; and Ho and Pennington-Cross, 2006a).

Another strand of research studied subprime loan termination by jointly estimating empirical models of prepayment and default (e.g., Alexander, et al., 2002; Pennington-Cross, 2003; Danis and Pennington-Cross, 2005a; Ho and Pennington-Cross, 2006b; Ho and Pennington-Cross, 2006c; and Pennington-Cross and Chomsisengphet, 2007). Related papers have sought to explain the length of time between delinquency and default (Danis and Pennington-Cross, 2005b); time in foreclosure (Pennington-Cross, 2006; and Capozza and Thomson, 2006); and loss given default (Capozza and Thomson, 2005).

Since the onset of the subprime mortgage crisis, research has attempted to identify various sources of the problem. Mayer, Pence, and Sherlund (2009) provide an overview of the attributes of subprime mortgages outstanding during this time and investigate why delinquencies and defaults increased so substantially.[11] These authors, as well as Gerardi, et al. (2008), point to a significant increase in borrower leverage during the mid-2000s, as measured by combined loan-to-value (CLTV) ratios, which was soon followed by falling house prices.

CLTV is important because economic theory predicts that borrowers with positive home equity will not default. That is, distressed borrowers with positive

[11] In related work, Mayer and Pence (2008) examine the geographic dispersion of subprime lending (states/cities/neighborhoods) for 2005.

equity could borrow against this equity or simply sell the home and pocket any net proceeds. Hence, negative equity (owing more than the home is worth) is a *necessary condition* for mortgage default. (See Foote, Gerardi, and Willen, 2008b for an overview of this issue.) As house prices declined in many parts of the US during 2007–8, an increasing number of homeowners found themselves with negative equity in their homes. Many borrowers facing negative income shocks—especially, financially fragile subprime mortgage borrowers—subsequently defaulted on their loans.

But how did such financially fragile borrowers obtain mortgage financing in the first place? Some research attention has been paid to the evolution of subprime mortgage underwriting standards. In particular, the focus has been on declining underwriting standards as measured by observable characteristics (e.g., Mayer, Pence, and Sherlund, 2009) or by increased forecast errors from empirical default models (Demyanyk and Van Hemert, 2008; and Rajan, Seru, and Vig, 2008).[12] Keys, et al. (2008) find that such unobserved negative characteristics are correlated with the use of securitization and attribute this to lax screening by subprime mortgage originators. The declining underwriting standards probably emanated from the sizeable rise in US house prices between 2001 and 2006, which probably masked much of the weakness.

Services

Recent service innovations primarily relate to enhanced account access and new methods of payment—each of which better meets consumer demands for convenience and ease. Automated teller machines (ATMs), which were introduced in the early 1970s and diffused rapidly through the 1980s, significantly enhanced retail bank account access and value by providing customers with around-the-clock access to funds. ATM cards were then largely replaced through the 1980s and 1990s by debit cards, which bundle ATM access with the ability to make payment from a bank account at the point of sale. Over the past decade, remote access has migrated from the telephone to the personal computer. Online banking, which allows customers to monitor accounts and originate payments using 'electronic bill payment', is now widely used. Stored-value, or prepaid, cards have also become ubiquitous.[13]

[12] Relatedly, Dell'Ariccia, Igan, and Laeven (2008) document a decline in the denial rate on subprime mortgage applications and find that this decline is correlated with geographic areas with higher house price appreciation and securitization rates.

[13] Other small-dollar payment options have emerged in recent years, such as smart cards and PayPal. However, we do not discuss these further due to their limited penetration and a dearth of research relating to 'electronic cash'.

Debit cards

Debit cards are essentially 'pay-now' instruments linked to a checking account whereby transactions can happen either instantaneously using online (PIN-based) methods or in the near future with offline (signature-based) methods. Consumers typically have the choice of using online or offline methods, and their selection often hinges on the respective benefits: Online debit allows the cardholder also to withdraw cash at the point of sale, and offline provides float. According to *ATM and Debit News* (2007), there were approximately 26.5 billion debit transactions in the US during 2006. This is up from 6.5 billion transactions in 1999—a fourfold increase.[14]

Much of the research pertaining to debit cards relates to identifying the most likely users of this payment instrument. Such demand-side explorations have been conducted individually as well as jointly across multiple payment options. Stavins (2001), for example, uses data from the 1998 Survey of Consumer Finance (SCF) and finds that debit usage is positively related to educational attainment, home-ownership status, marital status, business ownership, and being a white-collar worker; and is negatively related to age and net worth. Klee (2006) extends this analysis to consider the 1995, 1998, and 2001 SCFs and reports a secular increase in adoption driven by similar demographic factors.[15] Additional US evidence is provided by Mantel and McHugh (2001) using survey data from Vantis International; Hayashi and Klee (2003) using data from a 2001 survey conducted by Dove Consulting; as well as Borzekowski and Kiser (2008) and Borzekowski, Kiser, and Ahmed (2006) using 2004 data from the Michigan Surveys of Consumers.[16]

Some additional analysis by Hayashi and Klee (2003) studied the circumstances under which consumers are likely to use debit cards and found that these are more often used at grocery stores and gas stations than at restaurants. Related to this, the authors also find that debit card usage is positively related to the incidence of self-service transactions.

Online banking

As households and firms rapidly adopted Internet access during the late 1990s, commercial banks established an online presence. According to DeYoung (2005),

[14] It is worth noting, however, that debit cards were originally introduced as an innovation in the early 1980s but did not succeed at that time. Among the problems may have been the following: The likely potential adopters (younger, more educated, more affluent households) usually also had credit cards and would have been sensitive to the value of the float on a credit card at a time of relatively high interest rates. The quick payment attribute of a debit card was therefore not a 'value proposition' for this group.

[15] See also Anguelov, Hilgert, and Hogarth (2004) for the relevant statistics pertaining to these surveys. Also, using data across four Survey of Consumer Financed (SCFs), Zinman (2009) reports that, other things being equal, the choice of using debit cards is positively related to being a 'revolver' of credit card balances (as opposed to paying off such balances each month).

[16] International evidence is provided by Jonker (2005) for the Netherlands and by Loix, Pepermans, and Van Hove (2005) for Belgium.

the first bank websites were launched in 1995; and by 2002 nearly one-half of all US banks and thrifts operated transactional websites. As of 2007, bank call report data suggests that 77.0 percent of commercial banks offer transactional websites (and these banks control 96.8 percent of commercial bank deposits).

The primary line of research relating to online banking has been aimed at understanding the determinants of bank adoption and how the technology has affected bank performance.[17] In terms of online adoption, Furst, Lang, and Nolle (2002) find that US national banks (by the end of the third quarter of 1999) were more likely to offer transactional websites if they were: larger, younger, affiliated with a holding company, located in an urban area, and had higher fixed expenses and non-interest income.[18] Turning to online bank performance, DeYoung, Lang, and Nolle (2007) report that Internet adoption improved US community bank profitability—primarily through deposit-related charges. In a related study, Hernando and Nieto (2007) find that, over time, online banking was associated with lower costs and higher profitability for a sample of Spanish banks. Both papers conclude that the Internet channel is a complement to—rather than a substitute for—physical bank branches.[19]

Unlike the aforementioned studies, Mantel (2000) focuses on the demand-side of electronic/online bill payment—empirically analyzing the demographic characteristics of users. Among other things, the author finds that electronic bill payers tend to be: older, female, higher income, and homeowners.

Prepaid cards

As the name implies, prepaid cards are instruments whereby cardholders 'pay early' and set aside funds in advance for future purchases of goods and services. (By contrast, debit cards are 'pay-now', and credit cards are 'pay later'.) The monetary value of the prepaid card resides either on the card or at a remote database. According to Mercator Advisory Group, prepaid cards accounted for over $180 billion in transaction volume in 2006.

Prepaid cards can be generally delineated as either 'closed' systems (e.g., a retailer-specific gift card, such as Macy's or Best Buy) or 'open' systems (e.g., a payment network branded card, such as Visa or MasterCard). Closed-system prepaid cards have been effective as a cash substitute on university campuses, as well as for mass transit systems and retailers. Open-card systems, while less effective in this regard to date, may ultimately have greater promise owing to functionality that more resembles traditional debit and credit cards. For example, these prepaid cards

[17] See also Pennathur (2001) for a discussion of the various risks associated with online banking.

[18] Sullivan (2000) presents some statistics for banks in the Tenth Federal Reserve District that are generally consistent with this study.

[19] Additional evidence is offered by Ciciretti, Hasan, and Zazzara (2007), who also find that Italian banks offering Internet-related services had higher profitability (and stock returns) relative to their peers.

can be used to withdraw money from an ATM and to make purchases or pay bills in person, over the phone, or online. Cheney and Rhine (2006) discuss two types of open-system prepaid programs—payroll cards and general spending reloadable cards—each of which provides functions similar to deposit accounts. Payroll cards, which were first introduced in 2001, are particularly attractive for unbanked workers and their employers because of lower transactions costs (McGrath 2005). Such cards have also been used to deliver welfare benefits and disaster relief. Reloadable cards, which are typically offered through grocery stores and convenience stores, have most often been targeted to immigrants for remittances, to travelers, or to parents for teen purchases.

Some descriptive research relating to prepaid cards exists and is focused on certain public policy issues related to this payments medium. Furletti and Smith (2005) note the lack of state and federal consumer protections, but mention that card associations and bank-issuers have voluntarily extended some safeguards in practice, such as 'zero liability' and 'charge-back' provisions. Sienkiewicz (2007) discusses the potential for prepaid cards to be used in money laundering schemes. The author notes instances with offshore card issuance and the ability to access cash at ATMs as being the most vulnerable to illicit activity.

Production processes

The past twenty-five years have witnessed important changes in banks' production processes. The use of electronic transmission of bank-to-bank retail payments, which had modest beginnings in the 1970s, has exploded owing to greater retail acceptance, online banking, and check conversion. In terms of intermediation, there has been a steady movement toward a reliance on statistical models. For example, credit scoring has been increasingly used to substitute for manual underwriting—and has been extended even into relationship-oriented products like small business loans. Similar credit risk measurement models are also used when creating structured financial products through 'securitization'. Statistical modeling has also become central in the overall risk management processes at banks through portfolio stress testing and value-at-risk models—each of which is geared primarily to evaluating portfolio value in the face of significant changes in financial asset returns.

Automated clearing houses

An automated clearing house (ACH) is an electronic funds transfer network connecting banks—primarily used for recurring, small-dollar payments. While several ACH networks emerged in the 1970s, volumes grew only modestly through the 1980s, being used almost exclusively for direct payroll deposits. Over the past fifteen years, however, consolidation has occurred and volumes have soared.

According to the National Automated Clearing House Association, the number of ACH payments has increased from just under 2 billion in 1991 to 16 billion in 2006. (Over the same timeframe, the dollar value of ACH items transmitted rose from $6.9 trillion to $33.7 trillion.) These payments, in turn, are now made through only two ACH networks: The New York Clearing House's Electronic Payments Network and the Federal Reserve System's FedACH.

The modest literature on ACH networks has been aimed at understanding supply and demand conditions in support of FedACH pricing policies. Bauer and Hancock (1995) found that over the 1979–94 period the cost of processing an ACH item fell dramatically owing to scale economies, technological change, and lower input prices.[20] Stavins and Bauer (1999), on the other hand, estimated ACH demand elasticities by exploiting FedACH price changes over time—finding ACH demand to be highly inelastic. The two most recent papers studied network externalities for ACH. Gowrisankaran and Stavins (2004) find support for significant network externalities, which they ascribe to technological advancement, peer-group effects, economies of scale, and market power. Ackerberg and Gowrisankaran (2006) identify large fixed costs of bank adoption as the barrier to greater use of ACH transactions and thus to society's capturing the accompanying potential cost savings.

Small business credit scoring

Banks use a number of different lending technologies to lend to informationally opaque small businesses (see Berger and Udell, 2006 for a summary of these technologies). One new technology that was introduced in the 1990s and continues to evolve is small business credit scoring (SBCS). This technology involves analyzing consumer data about the owner of the firm and combining it with relatively limited data about the firm itself using statistical methods to predict future credit performance. Credit scores had long been pervasive in consumer credit markets (e.g., mortgages, credit cards, and automobile credits)—and resulted in widely available, low-cost, commoditized credits that are often packaged and sold into secondary markets.

The empirical literature studying SBCS has focused on the determinants of bank adoption and diffusion of this technology, as well as on how SBCS has affected credit availability. Two studies have statistically examined the determinants of the probability and timing of large banks' adoption of SBCS. Frame, Srinivasan, and Woosley (2001) and Akhavein, Frame, and White (2005) both find an important role for organizational structure in the adoption decision: banking organizations with fewer bank charters and more bank branches were more likely to adopt and also to adopt sooner. This suggests that large banks with a more 'centralized' structure were more likely to adopt SBCS. The use of the SBCS technology still

[20] Using a much smaller sample, Bauer and Ferrier (1996) also found support for the existence of ACH scale economies as well as significant allocative inefficiencies.

appears to be mostly limited to large banking organizations. However, one recent study suggests that small banks now often make use of the consumer credit score of the principal owner of the firm (Berger, Cowan, and Frame, 2009).

Several studies have focused on the relationship between SBCS adoption and credit availability. Three studies documented increases in the quantity of lending (Frame, Srinivasan, and Woosley, 2001; Frame, Padhi, and Woosley, 2004; and Berger, Frame, and Miller, 2005). One found evidence consistent with more lending to relatively opaque, risky borrowers (Berger, Frame, and Miller, 2005); another with increased lending within low-income as well as high-income areas (Frame, Padhi, and Woosley, 2004); and another with lending over greater distances (DeYoung, Glennon, and Nigro, 2008).[21]

Asset securitization

Asset securitization refers to the process by which non-traded assets are transformed into tradable 'asset-backed securities' (ABS) by repackaging cashflows.[22] Today, in the US, securitization is widely used by large originators of retail credit—specifically mortgages, credit cards, and automobile loans. As of year end 2007, federally sponsored mortgage pools and privately arranged ABS issues (including private-label mortgage-backed securities) totaled almost $9.0 trillion of the $49.9 trillion in US credit market debt outstanding. By contrast, as of year end 1990, these figures were $1.3 trillion and $13.8 trillion, respectively.[23]

A large number of books and articles have been devoted to the process of securitization and the analytics required to structure and value the resulting assets. As a result, we provide only a cursory review of the issues. Generally speaking, asset securitization involves several steps. The first is the sale of a pool of financial assets to a legally separate ('bankruptcy remote') trust against which liabilities (the ABS) are issued.[24] In this way, the original holder of the assets receives a cash payment, thereby liquefying its position. However, since the seller presumably has better information about the assets than does the buyer of the ABS (who thus faces the potential for 'adverse selection'), the buyer requires some form of 'credit enhancement' in the form of third-party guarantees, over-collateralization, or

[21] In cases in which SBCS is used in conjunction with other lending technologies, it is also shown to result in increased loan maturity (Berger, et al., 2005) and reduced collateral requirements (Berger, et al., 2006).

[22] The Government National Mortgage Association ('Ginnie Mae') was the first issuer of any kind of ABS—residential mortgage-backed securities (RMBS)—in 1970. The Federal Home Loan Mortgage Corporation ('Freddie Mac') was a 'fast second', with its RMBS appearing in 1971.

[23] Thomas (2001): Table 1 documents the tremendous growth of securitization by presenting the number and dollar value of privately arranged ABS transactions between 1983 and 1997 as reported in the Securities Data Company's New Issues Data Base.

[24] This discussion implicitly assumes a 'liquidating pool' of assets with fixed (but prepayable) terms to maturity. Some assets, like credit cards, are placed into 'revolving pools', which allow for the *ex post* addition of assets, since these loans have no fixed payment amount or term.

the creation of a priority of claims via 'tranching'.[25] While the first two forms of credit enhancement are straightforward, the last one requires some explanation.

Tranching involves the creation of two or more security types defined by their priority of claims.[26] The original seller often retains the most junior ('equity') security—the one with the lowest payment priority (and thus the first absorption of losses)—as a way of assuaging skeptical investors about the quality of the assets in the pool.[27] However, sophisticated investors—such as hedge funds—sometimes also hold such positions.[28]

Besides liquidity, securitization may be socially beneficial in so far as it allows for lower-cost financing of loans (through the separation of origination and funding); securitization may also hold private benefits for depository institutions seeking to manage their required capital positions. Thomas (2001) presents empirical evidence that the stockholders of certain ABS issuers benefit from securitization—that is, first-time issuers, large issuers, frequent issuers, lower-quality issuers, and bank-issuers.[29]

One recent innovation in the structured finance/securitization area is the introduction of collateralized debt obligations (CDOs). According to Longstaff and Rajan (2006) these instruments, which were first introduced in the mid-1990s, are now in excess of $1.5 trillion. Like ABS, CDOs are also liabilities issued by financial-institution-sponsored trusts, which essentially pool and restructure the priority of cash flows associated with other types of risky financial assets, including senior and mezzanine ABS, high-yield corporate bonds, and bank loans.[30] Lucas, Goodman, and Fabozzi (2006; 2007) provide in-depth discussions of CDO purposes, structures, and risks. Most of the emerging literature relating to CDOs is aimed at risk measurement and pricing.[31]

[25] Investors may also believe that deal sponsors are additionally providing some level of implicit recourse as a method to maintain their reputation in the market. Higgins and Mason (2004) and Gorton and Souleles (2005) provide empirical evidence consistent with this conjecture—higher-rated sponsors execute ABS deals at tighter spreads. See also Cantor and Hu (2007) for an analysis of differences between bank-sponsored and other ABS deals.

[26] The case of two securities (senior and junior) is generally sufficient to make the stylized points about securitization, but in practice much more granular structures are observed.

[27] This is consistent with important theoretical work in financial economics by Leland and Pyle (1977) and Myers and Majluf (1984) relating to capital structure more generally. See DeMarzo and Duffie (1999); and DeMarzo (2005) for similar discussion specific to asset-backed securities.

[28] Boot and Thakor (1993b) and Plantin (2004) provide theoretical explanations for the sale of tranched securities to investors of differing financial sophistication.

[29] Prior empirical work by Lockwood, Rutherford, and Herrera (1996) and Thomas (1999) had previously found conflicting evidence using subsamples of the data. The former paper focused on 1985–92 and the latter paper on 1991–6.

[30] There are also 'synthetic' CDOs, in which the CDO entity does not actually own the pool of assets but instead owns a credit default swap. In this way, the sponsoring institution transfers the economic risk but not the legal ownership of the underlying assets. See Goodman (2002) and Gibson (2004) for overviews and discussions of the motivation for and risks inherent in these structures.

[31] See, e.g., Duffie and Garleanu (2001); Hull and White (2004); Meneguzzo and Vecciato (2004); Yang, Hurd, and Zhang (2006); Longstaff and Rajan (2006); Kaniovski and Pflug (2007); and Glasserman and Suchintabandid (2007).

However, other work explores the relationship between CDOs and systemic risk (Krahnen and Wilde, 2006) and the relationship between banks' use of CDOs and their lending behavior (Goderis, et al., 2007).

The precipitous rise in subprime mortgage defaults—and expectations of future defaults—led to a significant decline in the value of subprime mortgage-backed securities and CDOs backed by such securities. This development, in turn, resulted in the freezing-up of secondary markets for subprime mortgages and mortgage securities. Subprime mortgages caught in the originate-to-distribute pipeline at that time were then returned to the originator's balance sheet. The material decline in asset values had serious consequences for leveraged investors with material exposures to subprime mortgage credit. This is how the subprime mortgage crisis evolved into a global financial crisis.

Some analysts have pointed to incentive conflicts inherent in the originate-to-distribute model of financial intermediation as a key reason for the magnification of the crisis—that is, how and why the surge in defaults by subprime mortgage borrowers, who are inherently more risky borrowers, had such a negative effect on the value of investment-grade subprime mortgage-backed securities. Ashcraft and Schuermann (2008) identify seven key informational frictions that arise in the originate-to-distribute model; discuss how market participants work to minimize such frictions; and speculate as to how this process broke down. The paper also provides an overview on how subprime mortgage securitization deals are structured and rated using detailed information from a securitized mortgage pool.

Risk management

Advances in information technology (both hardware and software) and financial theory spurred a revolution in bank risk management over the past two decades. Two popular approaches to measuring and managing financial risks are stress testing and value-at-risk (VAR). In either case, the idea is to identify the level of capital required for the bank to remain solvent in the face of unlikely adverse environments.

Stress testing involves constructing adverse scenarios for credit and/or interest rate conditions and then evaluating assets and liabilities—and thus solvency—under these stressed circumstances.[32] Fender and Gibson (2001) present a survey of stress testing in financial institutions. Berkowitz (1999–2000) and Kupiec (2000) both discuss certain shortcomings of stress testing for risk management, including whether the results of such tests will generally achieve equity capital allocations sufficient to stave off default under duress.

[32] Related stress testing procedures are also used by some central banks as a method of evaluating financial system resiliency in the face of shocks. See, e.g., Čihák (2007); Goodhart (2006); Elsinger, Lehar, and Summer (2006); and Majnoni, et al. (2001).

VAR relies on a probabilistic approach that evaluates the return distributions of assets. In this case, a bank would define a probability level of the return distribution (e.g., 99.9 percent) as an outer limit of exposure and then calculate the economic losses associated with that point on the distribution. Because of the focus on return distributions, VAR has been applied most widely to trading books, which are populated by readily marketable securities. Nevertheless, the principles involved have also been applied to credit portfolios. A large number of books and articles have been devoted to VAR—primarily centered on the appropriate characterization of return distributions for various assets and the use of VAR principles in the Basel II Capital Accord.

Organizational forms

New bank organizational forms have emerged in the US over the past few decades. Securities affiliates (so-called 'Section 20' subsidiaries or the creation of 'financial holding companies') for very large banks and Subchapter S status for very small banks, were the by-product of regulatory/legal evolution. Indeed, only one new organizational form, the Internet-only bank, arose from technological change. These institutions, which quickly emerged and disappeared, may represent an interesting laboratory for the study of 'failed' financial innovations. We believe that understanding such experimental failures may hold important insights for understanding the keys to successful financial innovations.

The dramatic increase in individuals' use of the Internet in the 1990s created the possibility of a new organizational form in banking: the Internet-only bank. According to Delgado, Hernando, and Nieto (2007), as of mid-year 2002, there were some thirty-five Internet-only banks operating in Europe and another twenty in the US. However, in Europe, virtually all of these banks were affiliated with existing institutions, while in the US they tended to be 'de novo' operations. This may explain why most/all of the US Internet-only banks have disappeared (through acquisition, liquidation, or closure) or established a physical presence to supplement their Internet base. This suggests that the dominant technology is one of 'clicks *and* mortar'.

DeYoung (2001; 2005) finds that, as compared with conventional 'de novo' banks, the Internet 'de novo' banks are less profitable owing to low business volumes (fewer deposits and lower non-interest income) and high labor expenditures. However, the author also reports that the financial performance gaps narrow quickly over time due to scale effects. Delgado, Hernando, and Nieto (2007) similarly find that European Internet banks demonstrate technology-based scale economies.

Conclusions

This chapter has reviewed the literature on technological change and financial innovation in banking since 1980. This quarter-century has been a period of substantial change in terms of bank services and production technologies, but much less so with respect to organizational form. As this survey indicates, although much has been learned about the characteristics of users and adopters of financial innovations and the attendant welfare implications, we still know little about how and why financial innovations are initially developed. This remains an important area for further research.

References

Ackerberg, D. A. and Gowrisankaran, G. (2006). 'Quantifying Equilibrium Network Externalities in the ACH Banking Industry', *RAND Journal of Economics*, 37: 738–61.

Akhavein, J., Frame, W. S., and White, L. J. (2005). 'The Diffusion of Financial Innovation: An Examination of the Adoption of Small Business Credit Scoring by Large Banking Organizations', *Journal of Business*, 78: 577–96.

Alexander, W. P., Grimshaw, S. D., McQueen, G. R., and Slade, B. A. (2002). 'Some Loans Are More Equal than Others: Third-Party Originations and Defaults in the Subprime Mortgage Industry', *Real Estate Economics*, 30: 667–97.

Anguelov, C. E., Hilgert, M. A., and Hogarth, J. M. (2004). 'U.S. Consumers and Electronic Banking, 1995–2003', *Federal Reserve Bulletin*, 90: 1–18.

Ashcraft, Adam, B. and Til Schuermann (2008). 'Understanding the Securitization of Subprime Mortgage Credit'. Federal Reserve Bank of New York Staff Study 318.

ATM and Debit News (2007). *EFT Data Book*, 6(44). New York: SourceMedia.

Bauer, P. W. and Hancock, D. (1995). 'Scale Economies and Technological Change in Federal Reserve ACH Payment Processing', *Economic Review*, Federal Reserve Bank of Cleveland, Third Q III: 14–29.

—— and Ferrier, G. D. (1996). 'Scale Economies, Cost Efficiencies, and Technological Change in Federal Reserve Payment Processing', *Journal of Money, Credit, and Banking*, 28: 1004–39.

Ben-Horim, M. and Silber, W. L. (1977). 'Financial Innovation: A Linear Programming Approach', *Journal of Banking & Finance*, 1: 277–96.

Berger, A. N. (2003). 'The Economic Effects of Technological Progress: Evidence from the Banking Industry', *Journal of Money, Credit, and Banking*, 35: 141–76.

—— Cowan, A., and Frame, W. S. (2009). 'The Surprising Use of Credit Scoring in Small Business Lending by "Community Banks" and the Attendant Effects on Credit Availability and Risk', Federal Reserve Bank of Atlanta White Paper 2009-9.

—— Espinosa-Vega, M., Frame, W. S., and Miller, N. (2005). 'Debt Maturity, Risk, and Asymmetric Information', *Journal of Finance*, 60: 2895–923.

Berger, A. N., Espinosa-Vega, M., Frame, W. S., and Miller, N. (2006). 'Why Do Borrowers Pledge Collateral? New Empirical Evidence on the Role of Asymmetric Information', Federal Reserve Bank of Atlanta Working Paper 2006-29a.

—— Frame, W. S., and Miller, N. (2005). 'Credit Scoring and the Availability, Price, and Risk of Small Business Credit', *Journal of Money, Credit, and Banking*, 37: 191–222.

—— Kashyap, A. K., and Scalise, J. M. (1995). 'The Transformation of the U.S. Banking Industry: What a Long, Strange Trip It's Been', *Brookings Papers on Economic Activity,* 2: 55–218.

—— and Udell, G. F. (2006). 'A More Complete Conceptual Framework for SME Finance', *Journal of Banking & Finance*, 30: 2945–66.

Berkowitz, J. (1999–2000). 'A Coherent Framework for Stress Testing', *Journal of Risk*, 2: 5–15.

Boot, A. W. A. and Thakor, A. V. (1993). 'Security Design', *Journal of Finance*, 48: 1349–78.

Borzekowski, R. and Kiser, E. K. (2008). 'The Choice at the Checkout: Quantifying Demand Across Payment Instruments', *International Journal of Industrial Organization*, 26: 889–902.

—— —— and Ahmed, S. (2006). 'Consumers' Use of Debit Cards: Patterns, Preferences, and Price Response', Board of Governors of the Federal Reserve System Finance and Economics Discussion Series 16.

Calem, P. S., Gillen, K., and Wachter, S. (2004). 'The Neighborhood Distribution of Subprime Mortgage Lending', *Journal of Real Estate Finance and Economics*, 29: 393–410.

Cantor, R. and Hu, J. (2007). 'Deal Sponsor and Credit Risk of Asset-Backed and Mortgage-Backed Securities', SSRN Working Paper 996014.

Capozza, D. R. and Thomson, T. A. (2005). 'Optimal Stopping and Losses on Subprime Mortgages', *Journal of Real Estate Finance and Economics*, 30: 115–31.

—— —— (2006). 'Subprime Transitions: Lingering or Malignant in Default?', *Journal of Real Estate Finance and Economics*, 33: 241–58.

Cheney, J. S. and Rhine, S. L. W. (2006). 'Prepaid Cards: An Important Innovation in Financial Services', Federal Reserve Bank of Philadelphia Payment Card Discussion Paper 06-07.

Chinloy, P. and Macdonald, N. (2005). 'Subprime Lenders and Mortgage Market Completion, *Journal of Real Estate Finance and Economics*, 30: 153–65.

Chomsisengphet, S. and Pennington-Cross, A. (2006). 'The Evolution of the Subprime Mortgage Market', *Review*, Federal Reserve Bank of St. Louis, 88: 31–56.

Ciciretti, R., Hasan, I., and Zazzara, C. (2007). 'Do Internet Activities Add Value? Evidence from Traditional Banks', *Journal of Financial Services Research*, 35: 81–98.

Čihák, M. (2007). 'Introduction to Applied Stress Testing', IMF Working Paper 07/59.

Cohen, W. M. (1995). 'Empirical Studies of Innovative Activity', in *Handbook of the Economics of Innovation and Technological Change*, P. Stoneman (ed.), Cambridge: Blackwell, 182–264.

—— and Levin, R. C. (1989). 'Empirical Studies of Innovation and Market Structure', in *Handbook of Industrial Organization*, 2. Richard Schmalensee and Robert Willig (eds.), Amsterdam: North-Holland, 1059–107.

Crews-Cutts, A. and Van Order, R. (2005). 'On the Economics of Subprime Lending', *Journal of Real Estate Finance and Economics*, 30: 167–96.

Danis, M. A. and Pennington-Cross, A. (2005a). 'A Dynamic Look at Subprime Loan Performance', Federal Reserve Bank of St. Louis Working Paper 2005-29.

DANIS, M. A. and PENNINGTON-CROSS, A. (2005b). 'The Delinquency of Subprime Mortgages', Federal Reserve Bank of St. Louis Working Paper 2005-22.

DELGADO, J., HERNANDO, I., and NIETO, M. J. (2007). 'Do European Primarily Internet Banks Show Scale and Experience Efficiencies?', *European Financial Management*, 13: 643–71.

DELL'ARICCIA, G., IGAN, D., and LAEVEN, L. (2008). 'Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market', International Monetary Fund Working Paper 2008-106.

DEMARZO, P. M. (2005). 'The Pooling and Tranching of Securities: A Model of Informed Intermediation', *Review of Financial Studies*, 18: 1–35.

—— and DUFFIE, D. (1999). 'A Liquidity-Based Model of Security Design', *Econometrica*, 67: 65–99.

DEMYANYK, Y. and VAN HEMERT, O. (2008). 'Understanding the Subprime Mortgage Crisis', available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1020396.

DEYOUNG, R. (2001). 'The Financial Performance of Pure Play Internet Banks', *Economic Perspectives*, 25: 60–75.

—— (2005). 'The Performance of Internet-Based Business Models: Evidence from the Banking Industry', *Journal of Business*, 78: 893–947.

—— GLENNON, D., and NIGRO, P. (2008). 'Borrower-Lender Distance, Credit Scoring, and Loan Performance: Evidence from Informational-Opaque Small Business Borrowers', *Journal of Financial Intermediation*, 17: 113–43.

—— LANG, W. W., and NOLLE, D. L. (2007). 'How the Internet Affects Output and Performance at Community Banks', *Journal of Banking & Finance*, 31: 1033–60.

DUFFIE, D. and GARLEANU, N. (2001). 'Risk and Valuation of Collateralized Debt Obligations', *Financial Analysts Journal*, 57: 41–59.

ELLIEHAUSEN, G. and STATEN, M. E. (2004). 'Regulation of Subprime Mortgage Products: An Analysis of North Carolina's Predatory Lending Law', *Journal of Real Estate Finance and Economics*, 29: 411–33.

ELSLINGER, H., LEHAR, A., and SUMMER, M. (2006). 'Using Market Information for Banking System Risk Assessment', *International Journal of Central Banking*, 2: 137–65.

FARRIS, J. and RICHARDSON, C. A. (2004). 'The Geography of Subprime Mortgage Prepayment Penalty Patterns', *Housing Policy Debate*, 15: 687–714.

FENDER, I. and GIBSON, M. S. (2001). 'Stress Testing in Practice: A Survey of 43 Major Financial Institutions', *Bank for International Settlements Quarterly Review*, June: 58–62.

FOOTE, C. GERARDI, L. K., and WILLEN, P. (2008). 'Negative Equity and Foreclosure: Theory and Evidence', *Journal of Urban Economics*, 64: 234–45.

FRAME, W. S., PADHI, M., and WOOLSEY, L. (2004). 'The Effect of Credit Scoring on Small Business Lending in Low- and Moderate-Income Areas', *Financial Review*, 39: 35–54.

—— SRINIVASAN, A., and WOOSLEY, L. (2001). 'The Effect of Credit Scoring on Small Business Lending', *Journal of Money, Credit, and Banking*, 33: 813–25.

—— and WHITE, L. J. (2004). 'Empirical Studies of Financial Innovation: Lots of Talk, Little Action?', *Journal of Economic Literature*, 42: 116–44.

FURLETTI, M. and SMITH, S. (2005). 'The Laws, Regulations, and Industry Practices that Protect Consumers Who Use Electronic Payment Systems: ACH E-checks and Prepaid Cards', Federal Reserve Bank of Philadelphia Payment Card Discussion Paper 05-04.

FURST, K., LANG, W., and NOLLE, D. (2002). 'Internet Banking', *Journal of Financial Services Research*, 22: 95–117.

Gerardi, K., Lehnert, A., Sherlund, S. M., and Willen, P. (2008). 'Making Sense of the Subprime Mortgage Crisis', *Brookings Papers on Economic Activity*, 2008: 69–159.

Gibson, M. S. (2004). 'Understanding the Risk of Synthetic CDOs', Board of Governors of the Federal Reserve System Finance and Economics Discussion Series 2004-36.

Glasserman, P. and Suchintabandid, S. (2007). 'Correlation Expansions for CDO Pricing', *Journal of Banking & Finance*, 31: 1375–98.

Goderis, B., Marsh, I. W., Castello, J. V., and Wagner, W. (2007). 'Bank Behavior with Access to Credit Risk Transfer Markets', Bank of Finland Research Discussion Paper 4-2007.

Goodhart, C. A. E. (2006). 'A Framework for Assessing Financial Stability?', *Journal of Banking & Finance*, 30: 3415–22.

Goodman, L. S. (2002). 'Synthetic CDOs: An Introduction', *Journal of Derivatives*, 9: 60–72.

Gorton, G. and Souleles, N. (2005). 'Special Purpose Vehicles and Securitization', National Bureau of Economic Research Working Paper 11190.

Gowrisankaran, G. and Stavins, J. (2004). 'Network Externalities and Technology Adoption: Lessons from Electronic Payments', *RAND Journal of Economics*, 35: 260–76.

Harvey, K. D. and Nigro, P. J. (2003). 'How Do Predatory Lending Laws Influence Mortgage Lending in Urban Areas? A Tale of Two Cities', *Journal of Real Estate Research*, 25: 479–508.

——— (2004). 'Do Predatory Lending Laws Influence Mortgage Lending? An Analysis of the North Carolina Predatory Lending Law', *Journal of Real Estate Finance and Economics*, 29: 435–56.

Hayashi, F. and Klee, E. (2003). 'Technology Adoption and Consumer Payments: Evidence from Survey Data', *Review of Network Economics*, 2: 175–90.

Hernando, I. and Nieto, M. J. (2007). 'Is the Internet Delivery Channel Changing Banks' Performance? The Case of Spanish Banks', *Journal of Banking & Finance*, 31: 1083–99.

Higgins, E. J. and Mason, J. R. (2004). 'What Is the Value of Recourse to Asset-Backed Securities? A Clinical Study of Credit Card Banks, *Journal of Banking & Finance*, 28: 875–99.

Ho, G. and Pennington-Cross, A. (2006a). 'The Impact of Local Predatory Lending Laws on the Flow of Subprime Credit', *Journal of Urban Economics*, 60: 210–28.

——— (2006b). 'Loan Servicer Heterogeneity and the Termination of Subprime Mortgages', Federal Reserve Bank of St. Louis Working Paper 2006-24.

——— (2006c). 'The Termination of Subprime Hybrid Fixed-Rate Mortgages', Federal Reserve Bank of St. Louis Working Paper 2006-42.

Hull, J. and White, A. (2004). 'Valuation of a CDO and an n-th to Default CDS without Monte Carlo Simulation', *Journal of Derivatives*, 12: 8–23.

Inside Mortgage Finance (2008). *The 2007 Mortgage Market Statistical Annual, I: The Primary Market*. Bethesda, MD: Inside Mortgage Finance Publications.

Jonker, N. (2005). 'Payment Instruments as Perceived by Consumers: A Public Survey', De Nederlandsche Bank Working Paper 53.

Kaniovski, Y. M. and Pflug, G. C. (2007). 'Risk Assessment for Credit Portfolios: A Coupled Markov Chain Model', *Journal of Banking & Finance*, 31: 2303–23.

Keys, B. J., Mukherjee, T. K., Seru, A., and Vig, V. (2008). 'Did Securitization lead to Lax Screening? Evidence from Subprime loans', available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1093137.

KLEE, E. (2006). 'Families' Use of Payment Instruments during a Decade of Change in the U.S. Payment System', Federal Reserve Board, Finance and Economics Discussion Series, 2006-01.

KRAHNEN, J. P. and WILDE, C. (2006). 'Risk Transfer with CDOs and Systemic Risk in Banking', Center for Economic Policy Research Discussion Paper 5618.

KUPIEC, P. H. (2000). 'Stress Tests and Risk Capital', *Journal of Risk*, 2: 27–39.

LAX, H, MANTI, M., RACA, P. and ZORN, P. (2004). 'Subprime Lending: An Investigation of Economic Efficiency', *Housing Policy Debate*, 15: 533–71.

LELAND, H. E. and PYLE, D. H. (1977). 'Informational Asymmetries, Financial Structure, and Financial Intermediation', *Journal of Finance*, 32: 371–87.

LERNER, J. (2002). 'Where Does State Street Lead? A First Look at Finance Patents, 1971–2000', *Journal of Finance*, 57: 901–930.

—— (2006). 'The New New Financial Thing: The Origins of Financial Innovations', *Journal of Financial Economics*, 79: 223–55.

LEVINE, R. (1997). 'Financial Development and Economic Growth: Views and Agenda', *Journal of Economic Literature*, 35: 688–726.

LOCKWOOD L. J., RUTHERFORD, R. C., and HERRERA, M. J. (1996). 'Wealth Effects of Asset Securitization', *Journal of Banking & Finance*, 20: 151–64.

LOIX, E., PEPERMANS, R., and VANHOVE, L. (2005). 'Who's Afraid of the Cashless Society?', Belgian Survey Evidence. Unpublished Manuscript.

LONGSTAFF, A. and RAJAN, A. (2006). 'An Empirical Analysis of the Pricing of Collateralized Debt Obligations', National Bureau of Economic Research Working Paper 12210.

LUCAS, D. J., GOODMAN, L. S., and FABOZZI, F. J. (2006). *Collateralized Debt Obligations: Structures and Analysis*, 2nd edn. Hoboken, NJ: John Wiley and Sons.

—— —— —— (2007). 'Collateralized Debt Obligations and Credit Risk Transfer', Yale International Center for Finance Working Paper 07-06.

MAJNONI, G., MARTINEZ-PERIA, M. S., BLASCHKE, W., and JONES, M. T. (2001). Stress Testing of Financial Systems: An Overview of Issues, Methodologies, and Financial Sector Assessment Program experiences', IMF Working Paper 01/88.

MANCHESTER, P. and POTERBA, J. M. (1989). 'Second Mortgages and Household Saving', *Regional Science and Urban Economics*, 19: 325–46.

MANTEL, B. (2000). 'Why Do Consumers Pay Bills Electronically? An Empirical Analysis', *Economic Perspectives*, Federal Reserve Bank of Chicago, Q IV: 32–47.

—— and MCHUGH, T. (2001). 'Competition and Innovation in the Consumer e-Payments Market? Considering Demand, Supply, and Public Policy Issues', Federal Reserve Bank of Chicago, Emerging Payments Occasional Working Paper EPS-2001-4.

MAYER, C. and PENCE, K. (2008). 'Subprime Mortgages: What, Where, and to Whom?' Federal Reserve Board Working Paper 2008-29.

—— —— and SHERLUND, S. (2009). 'The Rise in Mortgage Defaults', *Journal of Economic Perspectives*, 23: 27–50.

MCGRATH, J. C. (2005). 'The Cost-Effectiveness of Stored Value Cards for Unbanked Consumers', Federal Reserve Bank of Philadelphia Payment Card Discussion Paper 05-06.

MENEGUZZO, D. and VECCIATO, W. (2004). 'Copula Sensitivity in Collateralized Debt Obligations and Basket Default Swaps', *Journal of Futures Markets*, 24: 37–70.

MERTON, R. C. (1992). 'Financial Innovation and Economic Performance', *Journal of Applied Corporate Finance*, 4: 12–22.

MERTON, R. C. (1995). 'Financial Innovation and the Management and Regulation of Financial Institutions', *Journal of Banking & Finance*, 19: 461–81.

MILLER, M. H. (1986). 'Financial Innovation: The Last Twenty Years and the Next', *Journal of Financial and Quantitative Analysis*, 21: 459–71.

—— (1992). 'Financial Innovation: Achievements and Prospects', *Journal of Applied Corporate Finance*, 4: 4–12.

MYERS, S. C. and MAJLUF, N. S. (1984). 'Corporate Financing and Investment Decisions When Firms Have Information that Investors Do Not Have', *Journal of Financial Economics*, 13: 187–221.

NICHOLS, J., PENNINGTON-CROSS, A., and YEZER, A. (2005). 'Borrower Self-Selection, Underwriting Costs, and Subprime Mortgage Credit Supply', *Journal of Real Estate Finance and Economics*, 30: 197–219.

PENNATHUR, A. K. (2001). 'Clicks and Bricks: E-risk Management for Banks in the Age of the Internet', *Journal of Banking & Finance*, 25: 2103–23.

PENNINGTON-CROSS, A. (2003). 'Credit History and the Performance of Prime and Subprime Mortgages', *Journal of Real Estate Finance and Economics*, 27: 279–301.

—— (2006). 'The Value of Foreclosed Property', *Journal of Real Estate Research*, 28: 193–214.

—— and CHOMSISENGPHET, S. (2007). 'Subprime Refinancing: Equity Extraction and Mortgage Termination', *Real Estate Economics*, 35: 233–63.

PLANTIN, G. (2004). 'Tranching', SSRN Working Paper 650839.

QUERCIA, R. G., STEGMAN, M. A., and DAVIS, W. R. (2004). 'Assessing the Impact of North Carolina's Predatory Lending Law', *Housing Policy Debate*, 15: 573–601.

RAJAN, U., SERU, A., and VIG, V. (2008). 'The Failure of Models that Predict Failure: Distance, Incentives, and Defaults', available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1296982.

SCHUMPETER, J. A. (1950). *Capitalism, Socialism, and Democracy*, 3rd edn. New York: Harper & Bros.

SIENKIEWICZ, S. (2007). 'Prepaid Cards: Vulnerable to Money Laundering?', Federal Reserve Bank of Philadelphia Payment Card Discussion Paper 07-02.

STAVINS, J. (2001). 'Effect of Consumer Characteristics on the Use of Payment Instruments', *New England Economic Review*, Federal Reserve Bank of Boston, 2001: 19–31.

—— and BAUER, P. W. (1999). 'The Effect of Pricing on Demand and Revenue in Federal Reserve Automated Clearing House Payment Processing', *Journal of Financial Services Research*, 16: 27–45.

STRUNK, N. and CASE, F. (1988). *Where Deregulation Went Wrong: A Look at the Causes behind Savings and Loan Failures in the 1980s*. Chicago: U.S. League of Savings Institutions.

SULLIVAN, R. J. (2000). 'How Has the Adoption of Internet Banking Affected Performance and Risk in Banks?', *Financial Industry Perspectives*, Federal Reserve Bank of Kansas City, December: 1–16.

THOMAS, H. (1999). 'A Preliminary Look at the Gains from Asset Securitization', *Journal of International Financial Markets, Institutions, and Money*, 9: 321–33.

—— (2001). 'The Effect of Asset Securitization on Seller Claimants', *Journal of Financial Intermediation*, 10: 306–30.

TUFANO, P. (2003). 'Financial Innovation', in *Handbook of the Economics of Finance: 1A Corporate Finance*, G. M. Constantinides, M. Harris, and R. Stulz (eds.), Amsterdam: North Holland, 307–35.

VAN HORNE, J. C. (1980). 'Of Financial Innovations and Excesses', *Journal of Finance*, 40: 621–36.

WHITE, L. J. (1991). *The S&L Debacle: Public Policy Lessons for Bank and Thrift Regulation*. New York: Oxford University Press.

YANG, J., HURD, T. R., and ZHANG, X. (2006). 'Saddlepoint Approximation Method for Pricing CDOs', *Journal of Computational Finance*, 10: 1–20.

ZINMAN, J. (2009). 'Debit or Credit?', *Journal of Banking & Finance*, 33: 358–66.

CHAPTER 20

BANKING GLOBALIZATION

INTERNATIONAL CONSOLIDATION AND MERGERS IN BANKING

CLAUDIA M. BUCH
GAYLE L. DELONG[1]

Introduction

The internationalization of the financial services industry has become a buzzword and almost a synonym for the globalization process. Cross-border capital flows have risen tenfold in the past two decades,[2] and financial institutions are among the largest multinational companies worldwide.

Still, observers have long been puzzled by the relatively small number and frequency of cross-border mergers and acquisitions in banking. Compared to the number of domestic mergers, cross-border mergers have been relatively few,

[1] Sabrina Keller provided efficient research assistance. All errors and inaccuracies are solely our own responsibility.

[2] These data refer to the periods 1980–4 vs. 2000–4 and are taken from Kose, et al. (2006).

suggesting that implicit or explicit barriers to the integration of markets exist. Political and regulatory barriers are a natural candidate, as the banking industry is typically considered of strategic importance for the real economy and for financial stability. Yet, there has been a growing awareness that non-political obstacles such as cultural barriers might be holding back bank mergers as well. These barriers, in turn, could also affect the risk and efficiency effects of bank mergers.

In this chapter, we want to provide an overview of research on the causes and effects of international bank mergers. Considering the vast number of issues at stake, we naturally have to be selective. See Kose at al. (2006) for an encompassing overview of financial globalization, including macroeconomic aspects. Berger, Demsetz, and Strahan (1999) and Berger, et al. (2000) review earlier literature on the consolidation and globalization of financial institutions. Our focus is on three main questions: First, what are the determinants and driving forces of cross-border bank mergers? Second, what are the effects of cross-border bank mergers on the efficiency and competitiveness of financial institutions and the financial system? Third, what are the implications of bank mergers for risks in banking? Our focus is on empirical studies of the commercial banking industry. We begin with a brief review of the stylized facts on international bank mergers.

International mergers and acquisitions in banking: Still the rare animal?

International mergers between financial institutions, it may seem, are one feature of the globalization of financial markets. Headline cases—such as the takeover of the German bank Bayerische Hypo- und Vereinsbank by the Italian bank Unicredito in 2005, the inroads of US investment banks into Europe, or the presence of foreign banks in many emerging markets—show that the banking industry is currently operating at a global scale. Yet, a more careful examination of the numbers suggests that international mergers of financial institutions are relatively recent phenomena and tend to occur mainly between certain countries.

Figure 20.1 (Graph 1) shows how domestic and international bank mergers have evolved over time. We examine cross-border mergers that were announced *and* completed between 1985 and 2006 where at least one of the partners was a commercial bank and the other partner was any type of firm. Usually, the other partner was in financial services, that is, commercial banking, securities, or insurance. We define a cross-border merger as any merger whereby the headquarters of the target are not located in the same country as the ultimate parent of the acquirer. We obtain the

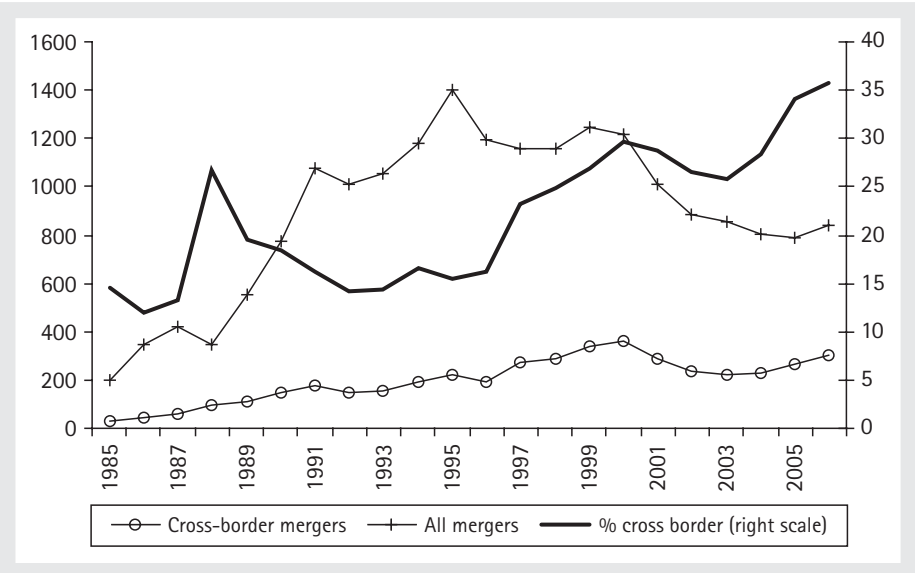

Fig. 20.1. Bank mergers (by year) 1985–2006

The study consists of 3,131 completed cross-border mergers announced between 1985 and 2006 where at least one partner is a commercial bank. The graph shows the number of international merges as well as the total number of bank mergers announced by year. This graph has been updated from Buch and DeLong (2004).

names of merger partners from Thomson Financial Securities Data. Up to 1992, the database includes all deals with values of at least $1 million. After 1992, deals of any value are covered. Also included are transactions with undisclosed values as well as public and private transactions. Thomson Financial Securities Data identifies 3,131 mergers that meet our criteria. The Graph shows that the number of international bank mergers has steadily increased over time, but the percentage of bank mergers that are cross-border has been small. The percentage climbed during the late 1980s to reach a plateau around 15 percent in the early 1990s. Between the mid-1990s and 2000, the share grew steadily to reach almost 30 percent in 2000. After a dip between 2001 and 2003, the percentage of cross-border mergers grew to over 35 percent in 2006.

Table 20.1 looks further into the regional structure of cross-border mergers and acquisitions (M&A) in banking. It shows that Europe and the Americas experienced a significant growth in the share of cross-border bank mergers in the years 1996–2006 compared with the years 1985–95. Asia, Africa/Middle East, and Australasia saw no significant change in the percentage of bank mergers represented by cross-border transactions. Table 20.1 also shows that cross-border mergers increasingly occurred *between* continents. This increase is different from the results of Buch and DeLong (2004), who compared cross-border mergers from 1985 to 2001. When they compared mergers during the two halves of their study—1985–93 vs. 1994–2001—they found banks chose targets within their continents more

Table 20.1. Cross-border bank mergers by continent

The table shows the number of cross-border mergers announced and completed between 1985 and 2006 where at least one partner is a commercial bank. It also reports results of splitting the sample according to year of announcement. The first time period is from 1985 to 1995, and the second is from 1996 to 2006. The statistical significance of the difference between the two time periods is measured using the following statistic: $z = \hat{\pi}_1 - \hat{\pi}_2 \Big/ \sqrt{\hat{\pi}(1-\hat{\pi})(\frac{1}{n_1}+\frac{1}{n_2})}$ where $\hat{\pi} = \frac{x_1+x_2}{n_1+n_2}$ and where $\hat{\pi}_1$ and $\hat{\pi}_2$ are the sample proportions, n_1 and n_2 are the total number of observations in each sample, and x_1 and x_2 are the number of observations that possess the characteristic. Worldwide figures are less than the sum of the continents due to mergers between banks headquartered in two nations that are located on the same continent.

	Europe	Americas	Africa/Middle East	Asia	Australasia	Total
			Panel A: 1985 to 2006			
Number of bank mergers	7,774	10,318	514	1,737	413	19,506
Cross-border mergers	3,131	1,347	266	784	220	5,748
Cross-border as % of total	40.3	13.1	51.8	45.1	53.3	29.5
Intra-continental as % of total	59.6	91.6	61.1	72.1	63.0	81.6
			Panel B: 1985 to 1995			
Number of bank mergers	3,064	5,082	137	440	158	8,356
Cross-border mergers	1,005	442	65	221	90	1,823
Cross-border as % of total	20.7	2.8	12.5	16.7	15.3	10.5
Intra-continental as % of total	87.0	94.0	57.7	65.7	57.0	70.5
			Panel C: 1996 to 2006			
Number of bank mergers	4,710	5,236	377	1,297	255	11,147
Cross-border mergers	2,216	905	201	563	130	3,925
Cross-border as % of total	27.1	5.6	15.1	13.7	17.2	15.9
Intra-continental as % of total	84.5	89.2	62.3	74.3	66.7	89.9
			Differences between panels B and C			
Cross-border as a % of total (z-statistic)	6.4*** (6.58)	2.8*** (7.01)	2.6 (0.76)	− 3.0 (− 1.47)	1.9 (0.51)	5.4*** (11.24)
Intra-continental as % of total (z-statistic)	− 2.5*** (− 3.07)	− 4.7*** (− 8.72)	4.7 (0.95)	8.6*** (3.37)	9.7** (1.97)	− 4.6*** (− 12.21)

Notes: *** Statistically significant at the 1% level. This table has been updated from Buch and DeLong (2004).

during the second half. Banks tended to acquire firms in neighboring countries when countries in Eastern Europe and Latin America first opened their markets. That is, banks in Western Europe and North America tended to acquire institutions in their own continents as opposed to traveling across oceans to find a merger partner. The updated results suggest that as banks polish the skills needed for cross-border mergers, they venture farther from their homes.

Evidence from the UNCTAD's World Investment Report[3] supports the finding that banks from advanced market economies dominate the global banking industry. Ranking the top fifty multinational firms in the financial services sector by an index based on the number and the location of their foreign affiliates shows that ten out of fifty firms were headquartered in the US in 2004. Only a handful of firms were headquartered outside the European Union or the US. The uneven degree of internationalization of banks from different continents is also supported by a recent empirical study of Schoenmaker and van Laecke (2007). Interestingly, they find that economic integration does not only stimulate integration within the region but also beyond.

The crisis that struck international financial markets beginning in 2007 can be expected to have a major impact on the incentives for banks to merge and to form alliances, both across borders and domestically. Defaults on subprime loans in the US forced investors to question the value of all their assets, which triggered a downward spiral of revaluations and declines in asset prices across the globe. Banks have been forced to sell parts of their assets and to raise new capital. (See Hellwig, 2008 for a detailed account of the causes and potential consequences of the crisis.) In Europe, for instance, the governments of Belgium, Luxembourg, and the Netherlands have initiated the divesture of Fortis and the sale of its assets to other financial institutions in Europe. The British mortgage lender Bradford and Bingley sold its deposits and branches to the Spanish bank Santander. Part of the resolution of the crisis have thus been mergers of domestic banks, be it through market forces, moral suasion of the supervisors, or outright interventions of policymakers.

In many countries, rescue packages that have been put into place to recapitalize the banking systems foresee provisions to restructure and to merge financial institutions. Hence, one might expect that the number of bank mergers could have increased. According to *Thomson Financial* though, the number of domestic bank mergers worldwide in fact increased by nearly 11 percent from 565 in 2007 to 626 in 2008. However, the number of cross-border bank mergers fell by 30 percent from 362 in 2007 to 254 in 2008. At the time of writing, we can only speculate about the reasons for changing numbers of bank mergers, and a careful analysis of the individual cases is needed. One reason for the decline in cross-border mergers though could be that mergers that are not urgently needed to rescue a failing bank have been curtailed. That is, about the only motive for mergers during such

[3] See <http://www.unctad.org/sect.s/dite_dir/docs/wir2006top50_spreadindex_en.pdf>.

economic turmoil is macroeconomic necessity. At the same time, banks seem to have tried to join forces domestically. As a result, the number of domestic mergers has increased occurred while international mergers have become even more rare.

At the time of writing, the turmoil is far from over. Hence, this chapter cannot give an overall assessment of the merger activity that has been triggered by the crisis and the long-run effects on banking markets. It is very clear though that both the activities that we have seen to far as well as the changes in international banking regulations that are currently under discussion will have implications for the structure of banking markets in the future. Banks will be subjected to more regulatory scrutiny, perhaps requiring more justification for mergers, especially cross-border mergers, which we will show can affect the riskiness of a bank. In this sense, the analysis of this chapter will provide some broad guidelines and assessments of what to expect based on experiences with cross-border mergers in the past.

More specifically, we will focus on an explanation of the relatively modest increase in cross-border M&A activity in the past, the regional concentration, and the dominance of a few large countries. We will also address the implications of cross-border bank mergers for efficiency and risks in banking.

Determinants of cross-border bank mergers[4]

Why should banks merge across border? Early theoretical literature on the determinants of international banking has taken a fairly eclectic approach to answer this question. Traditionally, this literature distinguishes between location- and ownership-specific factors (Sagari, 1992; and Williams, 1997). There has been relatively little formal theoretical work providing an encompassing model of the international banking firm. (See textbooks covering banking theory such as Allen and Gale, 2000 or Freixas and Rochet, 1998.) Typically, theoretical work focuses on specific aspects of international banking such as regulatory consequences (Repullo, 2000; Harr and Ronde, 2003; Dalen and Olsen, 2003) or the determinants of entry into markets in Eastern Europe (Claeys and Hainz, 2007). Gray and Gray (1981) and Berger, et al. (2004b) thus suggest borrowing from the literature on cross-border foreign investment of non-financial firms to explain cross-border banking activity. Goldberg (2004) also discusses whether multinational activities of banks and of non-financial firms can be treated in parallel. While she identifies parallels in the two literatures, she also notes differences between foreign direct investment (FDI)

[4] This section partly draws on Buch and DeLong (2004) and Berger, et al. (2004b).

in financial services and manufacturing, especially with respect to the implications for local institution building and business cycles.

In the remainder of this section, we review the empirical literature on the determinants of cross-border bank mergers more carefully. We structure the discussion around the main determinants—information costs, regulations, bank-specific variables, and other mostly macroeconomic factors. From a policy perspective, the distinction between efficiency barriers caused by regulations and by information costs is important. While the former can eventually be removed, the latter will remain even in (legally) integrated markets.

Most of the studies we review make use of a gravity-type model, which essentially relates bilateral economic activities between two countries to the size of markets and geographic distance. Studies using bank-level data additionally take into account the entry decision by estimating limited dependent variable models of banks' foreign expansions.

Information costs

Operating a financial institution in a foreign country raises a number of performance challenges for financial institution managers (Berger, et al., 2004b). Managers must grapple with differences in languages, laws, social practices, regulations, and customer expectations, as well as the sheer geographic distance between the home and host countries. These cross-border managerial challenges add to the usual difficulties of operating an acquired institution during the post-merger transition period. Hence, Berger, et al. (2001a) argue that 'efficiency' barriers such as distance as well as differences in language, culture, currency, and regulatory or supervisory structures could inhibit cross-border bank mergers.

One important impediment to cross-border bank mergers could thus be information costs. These can be proxied by geographical distance, a common language, or a common legal system. The motivation for the use of the distance variable is related to a strand of the literature that applies gravity-type models to international investment decisions. In this literature, distance is typically considered to capture transportation costs. In contrast, international finance literature interprets distance in terms of information costs. Empirical applications by Ahearne, Griever, and Warnock (2000), Buch (2003; 2005), Buch and Lipponer (2006; 2007), or Portes and Rey (1999) show that distance influences international capital flows and investment decisions of banks in a similar way as it influences international trade. Besides geographic proximity, sharing a common language is likely to lower the costs of melding two corporate cultures. Information needs to be communicated in only one language, and, more indirectly, sharing a common language can be seen as a proxy for common cultural links. Also, the presence of a common legal system should have a positive impact on cross-border M&A.

Buch and DeLong (2004) use bilateral data on the number of bank mergers between countries and find that information costs and regulations in fact impede cross-border mergers. At the same time, large, efficient banks located in countries with developed banking markets can overcome these barriers and tend to be the banks that expand abroad. Alibux (2007) uses more recent data (1995–2005) and confirms the importance of information costs and regulations as impediments to cross-border mergers. Focarelli and Pozzolo (2001b) look at where banks expand their cross-border shareholdings and find the most important determinants are potential profit opportunities as well as regulatory environments. The paper uses bank-level data on foreign investment for a representative sample of 260 large banks from the OECD countries. Cross-border shareholdings in their analysis include both mergers and greenfield investment.

Another barrier may be preferences for domestically owned institutions due to the 'concierge' services that they can provide in terms of knowledge of the local conditions and information about local non-financial suppliers and customers. Berger, et al. (2003) find that foreign affiliates of multinational corporations operating in European countries usually choose domestically owned banks for cash management services, consistent with the 'concierge' effect.

Regulations[5]

While information costs measure indirect, implicit barriers to the integration of banking markets, regulations of banking activities can erect direct, explicit barriers. The empirical literature on the determinants of bank mergers generally supports the hypothesis that deregulation has a substantial impact on merger decisions. Jayartne and Strahan (1998) and Saunders (1999) discuss the influence of deregulation in a domestic setting. Clearly, the presence of an international financial center in the target country makes countries more attractive destinations for international mergers (Choi, et al., 1986; Ter Wengel 1995). Also, foreign banks have often found it easy to make inroads into domestic banking systems that have undergone major privatization programs. Guillén and Tschoegl (2000) show that privatization has paved the way for many Spanish banks into Latin America, and Bonin and Abel (2000) show that privatization has been one of the reasons for the high market shares of foreign banks in the transition economies of Central and Eastern Europe. Generally, evidence on the experience with foreign banks in transition economies can be found in de Haas and Ilko Naaborg (2006); de Haas and Lelyveld (2006); or Haselmann (2006). Claeys and Hainz (2007) study the effects of different modes of entry for lending rates. Berger (2007b) supports this finding and adds net comparative

[5] A detailed database on banking regulations around the world can be found at <http://www.worldbank.org/research/interest/2003_bank_survey/2003_bank_regulation_database.htm>. See Barth, Dan Brambaugh, and Yago (2001) for a description of these data.

advantages for foreign banks, coupled with low government entry barriers, as explanations for high ratios of foreign bank ownership in some emerging markets.

Buch and DeLong (2004) provide evidence for the importance of the regulatory environment for cross-border bank mergers. They find that national banking regulations affect the probability of being an acquirer or target in cross-border bank mergers. Looking at changes in merger characteristics over time, they find that regulatory changes made to encourage regional integration produced mixed results. The number of cross-border bank mergers within the European Union following the European Union's Single Market Program in 1992 did not increase significantly, but the number of cross-border bank mergers among Canada, Mexico, and the US did increase after the implementation of the North American Free Trade Agreement in 1994.

The integration of banking markets in Europe indeed provides an interesting case study for the effects of regulations (see also Berglöf, et al., 2005). While official restrictions to the cross-border entry of banks have largely been abolished, implicit barriers through the 'misuse of supervisory power' (European Commission, 2005: 4) remain prevalent. According to a survey by the European Commission (2005), savings of fixed costs resulting from cross-border mergers are relatively small compared to those that can be achieved through domestic mergers, and, in particular, smaller financial institutions find it difficult to sell the same product in different markets. Hence, cross-border consolidation may be deterred by political factors, differences in institutions and cultures, the use of different payment and settlement systems, and remaining differences in capital markets, taxes, and regulations across countries (Giddy, Saunders, and Walter, 1996; Lannoo and Gros, 1998; Boot, 1999; Blandon, 2000; and Goddard, Molyneux, and Wilson, 2001).

Fecht and Grüner (2006) provide an alternative explanation for a relatively limited degree of pan-European bank mergers. In a theoretical model, they argue that the allocation of liquidity shocks may constitute a natural limit to the merger of banks. In their model, benefits from diversification and the costs of contagion may be traded off optimally when banks from some but not all regions merge. Carletti, Hartmann, and Spagnolo (2006) propose a theoretical model discussing the liquidity effects of bank mergers focusing on the trade-off between an internalization and a diversification effect.

Bank-specific factors

Bank-specific characteristics that increase the likelihood of entering into a merger include efficiency, experience in a competitive environment, economies of scale and scope, and domestic clients that have international operations. Using various measures of efficiency and profitability, studies find that stronger banks take over weaker ones in that acquirers tend to be more cost efficient (Berger and Humphrey, 1992), more profitable (Peristiani, 1993), or better capitalized (Wheelock and

Wilson, 2000) than their targets. For European banks, Vander Vennet (1998) confirms that acquiring banks tend to be larger and more efficient than their targets. Acquirers in cross-border mergers are generally large institutions from countries with developed financial markets (Focarelli and Pozzolo, 2001a; 2001b).

Macroeconomic factors

Macroeconomic factors such as a high-growth potential of host countries (so-called pull factors) or lagging growth in the home country (push factors) affect cross-border capital flows and foreign direct investments of banks. Also, the demand for differentiated financial services tends to increase with the level of economic development. The heightened demand increases the incentives for banks to form cross-border alliances and to jointly provide financial services. A high GDP per capita and large market size could also generate economies of scale and hence create motives for international mergers (Berger, Hunter, and Timme, 1993; Benston, Hunter, and Wall, 1995; and Berger, et al., 2000). Consistent with these hypotheses, empirical literature finds a positive effect of market size and GDP per capita on cross-border bank mergers (Buch and DeLong 2004; and Focarrelli and Pozzolo 2001b).

In addition to standard push and pull factors, literature on international banking has also borrowed possible determinants of the foreign expansion of banks from the theory of multinational firms (Goldberg, 2004). One implication of these theories is that, as two countries become more similar in size, relative factor endowments, and technical efficiency, foreign direct investment will increase relative to trade between the two countries (Markusen and Venables, 1995). Moreover, trade literature predicts that in industries like banking, for which intangible, firm-specific, and knowledge-based assets are important, international firms are more likely to export their management expertise via foreign direct investment rather than exporting the goods and services themselves. Thus, trade theory would predict significant cross-border financial institution M&A primarily between country pairs with similar national characteristics.

Berger, et al. (2004b) test the relevance of the new trade theory and the traditional theory of comparative advantage for explaining the geographic patterns of international M&A of financial institutions between 1985 and 2000. Their data provide statistically significant support for both theories. They also find evidence that the US has idiosyncratic comparative advantages at both exporting and importing financial institutions management. Claessens and van Horen (2007) use data on foreign direct investments of banks and confirm the importance of comparative advantages and institutional familiarity.

Below, we review literature stressing the importance of relative efficiency at the *bank*-level for the probability of becoming an acquirer in an international bank merger. Literature also shows the importance of profitability at the *country*-level.

Focarelli and Pozzolo (2001a) study the pattern of cross-border M&A in the banking industry relative to the non-financial sector. Using data on almost 2,500 banks from twenty-nine OECD countries, they find that banks tend to expand into countries where banking systems are inefficient.

A large literature on FDI in banking has dealt with the question whether trade and finance are linked. According to this literature, banking organizations engage in a 'follow-your-customer' strategy of setting up offices in countries where their home country customers have foreign affiliates (Goldberg and Saunders, 1981; and Brealey and Kaplanis, 1996). However, other researchers point out that foreign-owned banks lend mostly to borrowers other than customers from the home country, which suggests that 'follow-your-customer' may not be the dominant motivation behind cross-border M&A (Stanley, Roger, and McManis, 1993; and Seth, Nolle, and Mohanty, 1998). Focarreli and Pozzolo (2001b) support the 'follow-your-customer' hypothesis, especially for branches. However, they also find that other factors such as institutions and profit opportunities are relatively more important. Ultimately, however, firm- or bank-level evidence would be necessary to disentangle causality between the foreign expansions of banks and non-financial firms and thus to ultimately resolve the 'follow-your-customer' hypothesis.

Summary

The determinants of international bank mergers and cross-border banking in general are one of the most intensively research areas in the context of international bank mergers. A couple of stylized facts stand out. At the bank level, the probability of becoming an acquirer in an international merger is positively linked to size and profitability. At the country level, mergers are more frequent between large and developed market economies and countries with similar cultural background. In addition, regulatory entry barriers deter entry. The fact that implicit regulatory and cultural entry barriers into foreign markets still prevail is likely to have implications for the efficiency and risk effects of bank mergers. This is an issue to which we turn next.

Effects of cross-border bank mergers: Efficiency and competition

Cross-border bank mergers can affect the efficiency of banks through a number of channels. The merged entity could be able to exploit economies of scale and scope, or management and corporate governance practices could be improved. At the

same time, however, managing an increasingly large and complex organization operating in several countries may also lead to managerial inefficiencies and lower performance—the largest multinational banks, for instance, operate affiliates in up to seventy host countries, according to data collected by the United Nations Conference on Trade and Development. Which of the two effects dominates has been the subject of a large body of empirical literature.

Studies on the efficiency effects of cross-border bank mergers fall into two main groups. A first set of studies uses event studies to address the impact of mergers on banking performance. A second set of studies compares the efficiency of domestic versus foreign-owned banks. Since foreign ownership is often the result of mergers and acquisitions, these studies provide indirect evidence on efficiency effects of mergers. These studies also incorporate insights into the effects of foreign entry on competition in banking, thus addressing the potential trade-off between higher efficiency in the banks involved in cross-border mergers and the competitive structure of the banking system. (See Boot and Marinc, 2006 for a theoretical study of the trade-off between competition, efficiency, and the effectiveness of bank regulation.)

Event studies

Research on cross-border acquisitions of financial institutions in developed countries suggests, at best, mediocre post-merger financial performance. A study of cross-border M&A in Europe found that the associated combined bidder and target value changes were generally zero or negative, compared with domestic mergers, where combined values were positive on average (Beitel and Schiereck 2001). Similarly, a study of US domestic M&A found that mergers that combine two firms from different geographic areas create less shareholder value, consistent with fewer benefits from cross-border M&A (DeLong 2001). Cybo-Ottone and Murgia (2000) found that for fifty-four inter-European bank mergers between 1988 and 1997, the acquirers' abnormal returns were insignificantly different from zero.

DeLong (2003) goes one step further and compares market reactions to US bank mergers and to cross-border mergers. She examines abnormal returns of publicly traded partners upon the announcement of forty-one non-US bank mergers and compares the returns with a US control group. She finds acquirers of domestic bank mergers outside the US earn more on average than acquirers of domestic mergers in the US. Moreover, non-US targets tend to earn less than their US counterparts. However, for the subset of mergers in countries with relatively well-developed stock markets, she finds that partners both inside and outside the US earn similar returns. Ayadi and Pujals (2005) study bank M&A in Europe. They find that domestic mergers help cut costs but fail to achieve revenue synergies. Cross-border mergers, in contrast, generate revenue synergies, possibly due to improved geographical diversification.

Carletti, Hartmann, and Ongena (2007) analyze the impact of regulations on the effects of mergers. Using a new and unique dataset, they identify events that strengthen competition policy for nineteen countries and for the years 1987–2004. They find two positive effects of a more competition-oriented regime for merger control. First, the stock price increases for banks but not for non-financial firms. Second, targets of bank mergers become larger and more profitable.

Comparisons of bank efficiency

Whereas event studies compare the performance of merged banks before and after the merger, several studies also compare the efficiency of domestic and foreign-owned banks. (See Berger, 2007b for an encompassing survey.) Since M&A are a key channel for banks to enter foreign markets, these studies provide indirect evidence on the efficiency effects of bank mergers. In terms of the *host* country effects of M&A, we focus on competition and efficiency effects in the following. Other aspects such as the impact on lending to small and mid-sized firms are addressed, inter alia, in Berger, Klapper, and Udell (2001); Berger, et al. (2004b), or Goldberg, Dages, and Kinney (2000). Garcia Herrero and Simon (2003) survey the determinants and impact of financial sector FDI for the home economy.

Most of the efficiency studies of foreign-owned versus domestically owned banks within a developed country found the foreign-owned banks to be less efficient—with the possible exception of US banks operating abroad (DeYoung and Nolle, 1996; Chang, Hasan, and Hunter, 1998; and Berger, et al., 2000). However, a few studies found that foreign institutions have about the same efficiency on average as domestic institutions (Vander Vennet, 1996). Peek, Rosengren, and Kasirye (1999) argue that the poor performance of foreign bank subsidiaries is mainly due to pre-existing conditions. At the same time, foreign owners are also unable to turn around the banks they acquire.

Many of the alleged benefits of cross-border bank mergers are more prevalent for developing countries than for advanced market economies. Through foreign entry, emerging host countries can benefit from technology transfer, competition, and demonstration effects (Bank for International Settlements, 2004). Research on foreign banks in developing countries in fact finds results different from those in developed countries. For example, one study of foreign banks in over eighty countries found that foreign-owned banks in emerging markets have a relatively high profitability (Claessens, Demirgüç-Kunt, and Huizinga, 2001). This is consistent with disadvantages in financial institution management for local banks in these countries. Evidence in Demirgüç-Kunt and Huizinga (1999) supports the finding that foreign banks in emerging markets tend to outperform domestic banks.

Bank mergers that have a positive effect on efficiency at the bank level may have a negative effect on the competitive structure of the banking system. To assess the

overall welfare implications of bank mergers, we must examine the impact of mergers on market power as well.

Huizinga, Nelissen, and Vander Vennet (2001) show that the trade-off between efficiency and competition need not be steep. Using a sample of fifty-two horizontal bank mergers in Europe and studying the pre-euro period, they find evidence for unexploited scale economies and X-inefficiencies in European banking. To some extent, these inefficiencies are reduced through cross-border mergers. However, the authors do not find evidence for a greater market power of the merged banks.

Summary

The influence of foreign banks on a country's banking system is not well established. Theoretically, the added competition should increase efficiency and lower costs. Empirically, foreign-owned banks have been found to be less efficient than domestic banks in developed countries, suggesting they do not add much competition. In contrast to these general findings, banks from developed countries expanding into developing countries tend to be more efficient than their domestic counterparts. While operating in a developing country may add to a bank's risk, there may be benefits both to the bank and to the host banking system. We now look at aspects of that risk.

Effects of cross-border bank mergers: Risk[6]

A large set of studies looks at the determinants of risks in banking, but only a handful of these addresses the impact of the internationalization of banks (Nier and Baumann, 2003; De Nicolò, 2001; González, 2005; and Buch, DeLong, and Neugebauer, 2007). Yet, there is a growing awareness that cross-border banking activities could affect the risk and thus the stability of the domestic banking system.

A common argument in banking is that cross-border (geographic) mergers have the potential to reduce bank (and thus regulators') risk of insolvency (Segal, 1974; Vander Vennet, 1996; and Berger, 2000). This conventional wisdom is based on the notion that it is better for a bank not to put 'all its eggs in one basket' and thus geographic diversification is a naturally risk-reducing activity.

[6] This section draws partly on Amihud, DeLong, and Saunders (2002).

However, offsetting these perceived benefits are at least two potential costs that may well enhance the risk of bank insolvency and ultimately the risk exposure of bank regulators. The first risk-increasing effect comes from the incentives banks have to shift risk when the regulatory 'safety net' and its associated implicit and explicit guarantees are underpriced. As discussed by John, John, and Senbet (1991) and John, Saunders, and Senbet (2000), banks have incentives to increase their risk exposure beyond the level that would be privately optimal in a world in which there are no safety net guarantees or in which the safety net—deposit insurance, capital requirements, and implicitly, bank closure—is fairly priced. One way in which the safety net might be exploited is for a bank to acquire other (risky) banks by cross-border expansion. If the risky investment pays off, then the acquiring bank has the potential to keep any upside returns. If the acquisition of the foreign target fails and the domestic bank's (acquirer's) solvency is threatened, then the acquiring bank may be bailed out either by its own home or domestic regulator or perhaps by the host regulator (the regulator of the target bank). As a result, cross-border mergers may increase the insolvency risk exposure of either one or both the domestic (acquirer) and host (target) bank regulators.

A second reason why cross-border acquisitions may increase an acquirer's risk concerns 'who is watching the eggs in the basket' (Winton, 1999). Specifically, by extending its operations into new overseas markets, the (domestic) bank is confronted with potentially new and risk increasing monitoring problems related to the loan customer base, the operating cost structure, etc., of the target bank. If monitoring costs are high, these problems may also increase the insolvency risk of the domestic acquiring bank and implicitly the risk of domestic (and foreign) regulators.

The question is whether and to whom cross-border mergers are net beneficial. For example, if cross-border mergers do not raise the risk of acquiring banks relative to other domestic (home country) banks, or indeed, reduce their risk, then domestic regulators may encourage domestic banks to expand abroad. By contrast, if cross-border mergers increase the relative domestic riskiness of the acquiring bank, then domestic regulators may wish to scrutinize such mergers more carefully and may even seek to restrict them in an effort to reduce safety net subsidies and to reduce risk shifting behavior.

To gain insight into these issues, Amihud, DeLong, and Saunders (2002) examine risk effects of cross-border bank mergers. They analyze the change in total risk of an acquiring bank as a result of a cross-border banking merger, changes in the systematic risk of acquiring banks relative to home, foreign, and world market bank indexes, and the reaction of stock prices to news about the acquisition and examine the relationship between this stock price reaction and changes in risk brought about by cross-border bank mergers. They find that, on average, cross-border bank mergers do not change the risk of acquiring banks in any significant way. This finding has important regulatory policy implications in that the effect of an overseas acquisition is highly bank dependent or idiosyncratic. On average the

risk decreasing effects of cross-border bank mergers are offset by risk increasing effects, and the nature of the merging partners' operation changes in a way so as to leave the acquirer's risk unchanged. In a follow-up study, Buch and DeLong (2007) look further into the determinants of risk following bank mergers. They find that strong bank supervision is associated with banks reducing risk after mergers. The results suggest banks subject to strong supervision use cross-border mergers to diversify risk rather than to shift risk to banks in countries with weaker supervision.

Whereas the focus in Amihud, DeLong, and Saunders (2002) is on the exposure of banks to market risks, recent research has also addressed the exposure of banks to macroeconomic risks. Méon and Weill (2005) study the impact of mergers among large banks in Europe on the banks' exposure to macroeconomic risk. They find that loan portfolios of European banks provide a sub-optimal risk–return trade-off. Hence, there are potential gains in risk diversification from cross-border mergers even within the European Union due to imperfect correlations of business cycles.

A quite comprehensive theoretical literature also deals with the risk and regulatory consequences of international bank mergers. Repullo (2001), for instance, has a theoretical model in which a foreign bank becomes a branch of the domestic banks. Each bank is initially supervised by the domestic supervisory agency. Under home country control, the takeover moves responsibility to the domestic agency. The model shows that while cross-border bank acquisitions may reduce risk due to diversification, they also shift responsibility for supervision and deposit insurance to the domestic regulatory agency. Harr and Ronde (2003) study the regulatory implications of banks' organizational choice between branch and subsidiary, which fall under home and country supervision, respectively. Their results do not show incentives for regulators to engage in a 'race to the bottom' by relaxing banking regulations. Dalen and Olsen (2003) show that the link between multinational banking and risk taking is not clear cut. On the one hand, a lack of international coordination of supervisory responsibility toward subsidiaries of foreign banks tends to lower capital requirements. On the other hand, regulators respond by increasing incentives to improve asset quality.

Questions of supervisory responsibility are becoming more urgent as banks form more complex cross-border mergers. Dermine (2006), for instance, details the regulatory treatment and challenges presented by the Scandinavian bank Nordea, which was formed by banks from four different Nordic countries. Nordea adopted the Societas Europaea, a corporate structure that is governed by European Union law. The structure allows banks in the European Union to branch across national borders. Nordea is incorporated in Sweden, and Swedish supervisors are responsible for the supervision and deposit insurance of the entire entity. The result is that a branch operating in Finland, Denmark, or Norway has different supervisory and deposit insurance systems than host country banks with which they are competing.

Conclusions

In this chapter, we have reviewed the empirical literature on determinants and effects of cross-border mergers in banking. Our results can be summarized as followed.

First, the determinants of international banking activities are relatively well understood. Implicit and explicit barriers to the integration of markets can hold back cross-border merger activity. Implicit barriers include information costs as well as regulations impeding the market access of foreign banks. As explicit, direct barriers to the integration of markets have been lowered significantly in developed market economies, these indirect barriers have gained in relative importance. Also, bank mergers tend to take place mostly between large and developed countries, between countries in close regional proximity, and between countries which share a common cultural background. Considering the bank-level determinants of cross-border bank mergers, there is clear evidence for larger and more profitable banks to be the acquirers.

Second, several studies have looked into the effects of international bank mergers in terms of competition and efficiency. One common finding of this literature is that foreign-owned banks—which are often the result of mergers and takeovers—outperform domestically owned banks in developing countries. The comparative advantages of foreign banks in developed countries are less evident.

Third, despite the growing recognition that international banking can have an important impact on (international) financial stability, relatively few studies analyze the risk effects of bank mergers. At the bank level, studies find little evidence for a systematic change in risk following bank mergers.

In terms of future research in the field, we see three main gaps in the literature. First, testing the determinants of mergers and acquisition in banking based on a fully fledged model of the international bank would be desirable. Applying the literature on multinational firms to international banking while taking into account that 'banks are special' seems a potentially fruitful avenue for future theoretical work. The empirical banking literature provides ample evidence on the stylized facts that such a theoretical model should be able to match.

Second, improving our understanding of the risk–return trade-offs in international bank mergers requires the more extensive use of bank-level data. Looking into the portfolio structures of internationally active banks and assessing the impact of foreign activities on risks and returns could provide important insights into the effects of cross-border mergers.

Third, in the theoretical literature, there are a number of papers which analyze the regulatory consequences of multinational banking. At the bank level, these papers focus on the organizational choice between branches and subsidiaries; at the supervisory level, the focus is on the costs and benefits between home and

host-country control. To the best of our knowledge though, few of these papers have been put to a structural empirical test. This would be the natural next step.

Ultimately, linking more closely together empirical and theoretical work on international bank mergers and making use of new bank-level datasets is not only of academic interest. It will also pay in terms of improved information for policymakers. The increase in cross-border financial institutions raises some important policy issues, such as the transmission of systemic risk across borders, the governance and supervision of multinational financial institutions, and the extent to which foreign-owned institutions will provide sufficient services in times of local crises. Reacting to these challenges in an appropriate way is particularly important for developing countries. These countries enjoy relatively great benefits of foreign bank entry in terms of efficiency but they may also encounter potentially greater risks. Supervisory must react via adjustments in the supervisory framework, information sharing among supervisors, and developing supervisory skills.

References

Ahearne, A., Griever, W. and Warnock, F. (2000). 'Information Costs and Home Bias: An Analysis of U.S. Holdings of Foreign Equities', International Finance Discussion Paper, 691, Federal Reserve Board, Washington.

Alibux, A. N. R. N. (2007). 'Cross-Border Mergers and Acquisitions in the European Banking Sector'. Dissertation. Erasmus University of Rotterdam.

Allen, F. and Gale, D. (2000). *Comparing Financial Systems*. Cambridge, MA: MIT Press.

Amihud, Y., DeLong, G., and Saunders, A. (2002). 'The Effects of Cross-Border Bank Mergers on Bank Risk and Value'. *Journal of International Money and Finance*, 21: 857–77.

Ayadi, R., and Pujals, G. (2005). 'Banking Mergers and Acquisitions in the EU: Overview, Assessment and Prospects'. Société Universitaire Européenne de Recherches Financiéres (SUERF) Study 3.

Bank for International Settlements (BIS) (2004). 'Foreign Direct Investment in the Financial Sector of Emerging Market Economies'. *CGFS Publications*, 22. Basel.

Barth, J. R., Caprio, G. Jr., and Levine, R. (2001). 'The Regulation and Supervision of Banks around the World: A New Database, in Litan R. E. and Herring, R. (eds.), *Integrating Emerging Market Countries into the Global Financial system*, Washington: Brookings Institution Press.

Beitel, P. and Schiereck, D. (2001). 'Value Creation at the Ongoing Consolidation of the European Banking Market'. Institute for Mergers and Acquisitions Working Paper.

Benston, G., Hunter, C., and Wall, L. (1995). 'Motivations for Bank Mergers and Acquisitions: Enhancing the Deposit Insurance Put Option versus Earnings Diversification', *Journal of Money, Credit and Banking*, 27: 777–88.

Berger, A. N. (2000). 'The Big Picture of Bank Diversification'. Proceedings, *Federal Reserve of Chicago*, May, 162–74.

Berger, A. N. (2007a). 'International Comparisons of Banking Efficiency'. *Financial Markets, Institutions & Instruments,* 16: 119–44.
—— (2007b). 'Obstacles to a Global Banking System: "Old Europe" versus "New Europe",' *Journal of Banking & Finance,* 31: 1955–73.
—— Buch, C. M., DeLong, G., and DeYoung, R. (2004a). 'Exporting Financial Institutions Management via Foreign Direct Investment Mergers and Acquisitions'. *Journal of International Money and Finance,* 23: 333–66.
—— Dai, Q., Ongena, S., and Smith, D. C. (2003). 'To What Extent Will the Banking Industry Be Globalized? A Study of Bank Nationality and Reach in 20 European Nations'. *Journal of Banking & Finance,* 27: 383–415.
—— Demsetz, R. S., and Strahan, P. E. (1999). 'The Consolidation of the Financial Services Industry: Causes, Consequences, and Implications for the Future'. *Journal of Banking & Finance,* 23: 135–94.
—— DeYoung, R., Genay, H., and Udell, G. F. (2000). 'The Globalization of Financial Institutions: Evidence from Cross-Border Banking Performance'. *Brookings-Wharton Papers on Financial Services,* 3: 23–158.
—— —— and Udell, G. F. (2001a). 'Efficiency Barriers to the Consolidation of the European Financial Services Industry', *European Financial Management,* 7: 117–130.
—— Hanweck, G. A., and Humphrey, D. B. (1987). 'Competitive Viability in Banking: Scale, Scope, and Product Mix Economies'. *Journal of Monetary Economics,* 20: 501–20.
—— and Humphrey, D. B. (1992). 'Megamergers in Banking and the Use of cost Efficiency as an Antitrust Defense', *Antitrust Bulletin,* 37: 541–600.
—— Hunter, W. C., and Timme, S. G. (1993). 'The Efficiency of Financial Institutions: A Review and Preview of Research Past, Present, and Future', *Journal of Banking & Finance,* 17: 389–405.
—— Klapper, L. F., and Udell, G. F. (2001b). 'The Ability of Banks to Lend to Informationally Opaque Small Businesses'. *Journal of Banking & Finance,* 25: 2127–67.
Berglöf, E., Fulghieri, P., Jordi, G., Colin, M., Barros, P. L., and Pita and Vives, X. (2005). 'Integration of European Banking: The Way Forward'. Centre for European Policy Studies. London.
Blandon, J. G. (2000). 'Cross-Border Banking in Europe: An Empirical Investigation'. Universitat Pompeu Fabra Working Paper.
Bonin, J. P. and Abel, I. (2000). 'Retail Banking in Hungary: A Foreign Affair?', Working Paper, Economics Department, Wesleyan University, Middletown.
Boot, A. W. A. (1999). 'European Lessons on Consolidation in Banking'. *Journal of Banking & Finance,* 23: 609–13.
—— and Marinc, M. (2007). 'Competition and Entry in Banking: Implications for Capital Regulation'. University of Amsterdam and CEPR. Mimeo.
Brealey, R. A. and Kaplanis, E. C. (1996). 'The Determination of Foreign Banking Location'. *Journal of International Money and Finance,* 15: 577–97.
Buch, C. M. (2003). 'Information versus Regulation: What Drives the International Activities of Commercial Banks'. *Journal of Money, Credit, and Banking,* 35: 851–69.
—— (2005). 'Distance and International Banking'. *Review of International Economics,* 13: 787–804.
—— and DeLong, G. L. (2004). 'Cross-Border Bank Mergers: What Lures the Rare Animal?' *Journal of Banking & Finance,* 28: 2077–102.

——— ——— (2006). 'Do Weak Supervisory Systems Encourage Bank Risk-Taking?' Baruch College and University of Tübingen. Working Paper.
——— ——— and Neugebauer, K. (2007). 'International Banking and the Allocation of Risk'. Institute for Applied Economic Research (IAW). *Discussion Paper*, 32. Tübingen.
——— and Lipponer, A. (2007). 'FDI versus Exports: Evidence from German Banks'. *Journal of Banking & Finance,* 31: 805–26.
——— ——— (2006). 'Clustering or Competition? The Foreign Investment Behaviour of German Banks'. *International Journal of Central Banking,* 5: 135–68.
Carletti, E., Hartmann, P., and Spagnolo, G. (2006). 'Bank Mergers, Competition and Liquidity'. Center for Financial Studies, European Central Bank, and Stockholm School of Economics. Mimeo.
——— ——— and Ongena, S. (2007). 'The Economic Impact of Merger Control: What Is Special About Banking?' Center for Financial Studies, European Central Bank, and Tilburg University. Mimeo.
Chang, C. E., Hasan, I., Hunter, W. C. (1998). 'Efficiency of Multinational Banks: An Empirical Investigation'. *Applied Financial Economics,* 8: 1–8.
Choi, S.-R., Tschoegl, A. E., and Yu, E.-M. (1986). 'Banks and the World's Major Financial Centers, 1970–1980', *Weltwirtschaftliches Archiv,* 122: 48–64.
——— Park, D., and Tschoegl, A. E. (2002). 'Banks and the World's Major Banking Centers, 2000', Working Paper, Hanyang University, Seoul, and Wharton School of the University of Pennsylvania, Philadelphia.
Claessens, S., Demirgüç-Kunt, A., and Huizinga, H. (2001). 'How Does Foreign Entry Affect Domestic Banking Markets?' *Journal of Banking & Finance,* 25: 891–911.
——— and van Horen, N. (2007). 'Location Decisions of Foreign Banks and Competitive Advantage'. International Monetary Fund, World Bank, and University of Amsterdam. Mimeo.
Claeys, S. and Hainz, C. (2007). 'Acquisition versus Greenfield: The Impact of the Mode of Foreign Bank Entry on Information and Bank Lending Rates', Sveriges Riksbank, Working Paper Series, 210. Stockholm.
Commission of the European Communities (2005). 'Cross-Border Consolidation in the EU Financial Sector'. Securities and Exchange Commission Staff Document 1398. Brussels.
Cybo-Ottone, A. and Murgia, M. (2000). 'Mergers and Shareholder Wealth in European Banking'. *Journal of Banking & Finance,* 24: 831–59.
Dalen, D. M. and Olsen, T. E. (2003). 'Regulatory Competition and Multinational Banking', CESifo Working Paper Series, 971. Munich.
de Haas, R. and Naaborg, I. (2006). 'Foreign Banks in Transition Countries: To Whom Do They Lend and How Are They Financed?' *Financial Markets, Institutions & Instruments,* 15: 159–99.
——— and van Lelyveld, I. (2006). 'Foreign Banks and Credit Stability in Central and Eastern Europe: A Panel Data Analysis'. *Journal of Banking & Finance,* 30: 1927–52.
De Nicolò, G. (2001). 'Size, Charter Value and Risk in Banking: An International Perspective', in *The Financial Safety Net: Costs, Benefits and Implications for Regulation*, Proceedings of the 37th Annual Conference on Bank Structure and Competition, Federal Reserve of Chicago, 197–215.
DeLong, G. L. (2001). 'Stockholder Gains from Focusing versus Diversifying Bank Mergers'. *Journal of Financial Economics,* 59: 221–52.

DeLong, G. L. (2003). 'Does Long-Term Performance of Mergers Match Market Expectations?' Evidence from the US banking industry. *Financial Management,* 32: 5–26.

Demirgüç-Kunt, A. and Huizinga, H. (1999). 'Determinants of Commercial Bank Interest Margins and Profitability: Some International Evidence', *World Bank Economic Review,* 13: 379–408.

——— (2001). 'The Taxation of Domestic and Foreign Banking', *Journal of Public Economics,* 79: 429–53.

Dermine, J. (2006). 'European Banking Integration: Don't Put the Cart before the Horse', *Financial Markets, Institutions & Instruments,* 15: 57–106.

DeYoung, R. and Nolle, D. E. (1996). 'Foreign-Owned Banks in the U.S.: Earning Market Share or Buying it?' *Journal of Money, Credit, and Banking,* 28: 622–36.

European Central Bank (ECB) (2003). 'Structural Analysis of the EU Banking Sector: Year 2002'. Frankfurt a.M.

Fecht, F. and Grüner, H. P. (2006). 'Limits to International Banking Consolidation'. Deutsche Bundesbank. Discussion Paper 11. Frankfurt a.M.

Focarelli, D. and Pozzolo, A. F. (2001a). 'The Patterns of Cross-Border Bank Mergers and Shareholdings in OECD Countries', *Journal of Banking & Finance,* 25: 2305–37.

——— (2001b). 'Where Do Banks Expand Abroad?', Working Paper, Banca d'Italia, Rome.

Freixas, X. and J.-C. Rochet (1998). *Microeconomics of Banking.* Cambridge, MA: MIT Press.

Giddy, I., Saunders, A., Walter, I. (1996). 'Alternative Models for Clearance and Settlement: The Case of the Single European Capital Market'. *Journal of Money, Credit and Banking,* 28: 986–1000.

Goddard, J., Molyneux, P., Wilson, J. O. S. (2001). *European Banking: Efficiency, Technology and Growth.* Chichester, UK: John Wiley and Sons.

——— Dages, B. G., Kinney, D. (2000). 'Foreign and Domestic Bank Participation in Emerging Markets: Lessons from Mexico and Argentina'. *Federal Reserve Bank of New York Economic Policy Review,* 6: 17–36.

——— Saunders, A. (1981). 'The Determinants of Foreign Banking Activity in the United States'. *Journal of Banking & Finance,* 5: 17–32.

Goldberg, L. G. (2004). 'Financial-Sector FDI and Host Countries: New and Old Lessons', National Bureau of Economic Research Working Paper 10441, Cambridge, MA.

Gonzáles, F. (2005). 'Bank Regulation and Risk-Taking Incentives: An International Comparison of Bank Risk'. *Journal of Banking & Finance,* 29: 1153–84.

Gray, J. M. and Gray, H. P. (1981). 'The Multinational Bank: A Financial MNC?', *Journal of Banking & Finance,* 5: 1–153.

Guillén, M. and Tschoegl, A. E. (2000), 'The Internationalization of Retail Banking: The Case of the Spanish Banks in Latin America', *Transnational Corporations,* 9: 63–97.

Harr, T. and Ronde, T. (2003). 'Branch or Subsidiary? Capital Regulation of Multinational Banks'. University of Copenhagen and CEPR. Mimeo.

Haselmann, R. (2006). *Performance and Strategies of Banks in Transition Economies.* Marburg: Metropolis Verlag.

Herrero, A. G. and Simón, D. N. (2003). 'Determinants and Impact of Financial Sector FDI to Emerging Economies: A Home country's perspective', Banco de España documentos ocasionales 8. Madrid.

Huizinga, H. P., Nelissen, J. H. M., and Vander Vennet, R. (2001). 'Efficiency Effects of Bank Mergers and Acquisitions in Europe'. Tinbergen Institute Discussion Paper 088/3. Rotterdam.

Jayaratne, J. and Strahan, P. (1998). 'Entry Restrictions, Industry Evolution, and Dynamic Efficiency: Evidence from Commercial Banking', *Journal of Law & Economics*, 41: 239–73.

John, K., John, T., and Senbet, L. (1991). 'Risk Shifting Incentives of Depository Institutions: A New Perspective on Federal Deposit Insurance Reform', *Journal of Banking & Finance,* 15: 895–915.

—— Saunders, A., Senbet, L. (2000). 'A Theory of Bank Compensation and Management Regulation', *Review of Financial Studies,* 3: 95–126.

Kose, M. A., Prasad, E., Rogoff, K., and Wei, S.-J. (2006). 'Financial Globalization: A Reappraisal'. IMF Working Paper 06/189. Washington, DC.

Lannoo, K. and Gros, D. (1998). 'Capital Markets and EMU: Report of a CEPS Working Party'. Centre for European Policy Studies.

Markusen, J. R. and Venables, A. J. (1995). 'The Increased Importance of Multinationals in North American Economic Relationships: A Convergence Hypothesis', in Canzoneri, M. W., Ethier, W. J., and Grilli, V. (eds.), *The New Transatlantic Economy*. London: Cambridge University Press.

Méon, P. G. and Weill, L. (2005). 'Can Mergers in Europe Help Banks Hedge against Macroeconomic Risk?' *Applied Financial Economics*, 15: 315–26.

Nier, E. and U. Baumann (2003). 'Market Discipline, Disclosure and Moral Hazard in Banking', in *Corporate Governance: Implications for Financial Services Firms*, Proceedings of the 39th Annual Conference on Bank Structure and Competition, Federal Reserve Bank of Chicago, 261–77.

Peek, J., Rosengren, E. S., and Kasirye, F. (1999). 'The Poor Performance of Foreign Bank Subsidiaries: Were the Problems Acquired or Created?', *Journal of Banking & Finance,* 23: 579–604.

Peristiani, S., (1993). 'The Effects of Mergers on Bank Performance', in Studies on Excess Capacity in the Financial sector, Federal Reserve Bank of New York, New York.

Portes, R. and Rey, H. (1999). 'The Determinants of Cross-Border Equity Flows', Working Paper, 7336 (National Bureau of Economic Research, Cambridge).

Repullo, R., (2000). 'A Model of Takeovers of Foreign Banks', *Spanish Economic Review,* 3: 1–21.

Sagari, S. B. (1992). 'United States Foreign Direct Investment in the Banking Industry', *Transnational Corporations*, 1: 93–123.

Saunders, A. (1999). 'Consolidation and Universal Banking', *Journal of Banking & Finance,* 23: 693–5.

Schoenmaker, D. and van Laecke, C. (2007). 'Determinants of International Banking: Evidence from the World's Largest Banks'. Government of the Netherlands. Ministry of Finance. Mimeo.

Segal, Z. (1974). 'Market and Industry Factors Affecting Commercial Bank Stocks: An Analysis of the Price Behavior of Bank Stock Prices'. Unpublished Doctoral Dissertation, New York University, New York.

Seth, R., Nolle, D. E., Mohanty, S. K. (1998). 'Do Banks Follow their Customers Abroad?' *Financial Markets, Institutions & Instruments,* 7: 1–25.

Stanley, T. O., Roger, C., and McManis, B. (1993). 'The Effects of Foreign Ownership of U.S. Banks on the Availability of Loanable Funds to Small Business'. *Journal of Small Business Management,* 31: 51–66.

Ter Wengel, J. (1995). 'International Trade in Banking Services', *Journal of International Money and Finance,* 14: 47–64.

Vander Vennet, R. (1996). 'The Effect of Mergers and Acquisitions on the Efficiency and Profitability of EC Credit Institutions'. *Journal of Banking & Finance*, 20: 1531–58.

—— (1998). 'Causes and Consequences of EU Bank Takeovers, in S. Eijffinger, K. Koedijk, M. Pagano, and R. Portes (eds.), *The Changing European Landscape*, Brussels: Centre for Economic Policy Research, 45–61.

Wheelock, D. C. and Wilson, P. W. (2000). 'Why Do Banks Disappear? The Determinants of U.S. Bank Failures and Acquisitions', *Review of Economics and Statistics*, 82: 127–38.

Williams, B. (1997). 'Positive Theories of Multinational Banking: Eclectic Theory versus Internalisation Theory', *Journal of Economic Surveys*, 11: 71–100.

Winton, A. (1999). 'Don't Put All Your Eggs in One Basket? Diversification and Specialization in Lending'. Working Paper, University of Minnesota.

CHAPTER 21

SMALL BUSINESS LENDING

ALLEN N. BERGER[1]

Introduction

Small businesses are engines of growth in the modern economy and bank lending provides much of the fuel for this growth. Banks are the single largest provider of external finance to small businesses, funding about 19 percent of all small business assets in the US (Berger and Udell, 1998). Small business lending differs from lending to large corporations because small businesses are generally much more informationally opaque, lacking certified audited financial statements, and market prices for traded equity or debt. As a result of this opacity, small firms often face significant difficulties in accessing funding for positive net present value projects. To address the informational opacity problem, banks use a number of different lending technologies.

The basic research model for analyzing bank small business lending has evolved considerably since the early 1990s. Early theoretical studies (e.g., Sharpe, 1990; and Rajan, 1992) and subsequent empirical analyses (e.g., Petersen and Rajan, 1994; and Berger and Udell, 1995) helped to bring about a broad recognition of the special methods that banks use to lend to small businesses.

[1] The author thanks Lamont Black, Ken Brevoort, Nate Miller, Greg Udell, and John Wolken for helpful comments and suggestions.

These studies focused on the differences between relationship lending and transactions lending technologies.

Under the technology of relationship lending, a loan officer collects proprietary information through contact over time with the small business, its owner, and other members of its local community. The bank then uses these data to make future decisions on underwriting and contract terms with the firm. Thus, a bank with an existing relationship with the firm may be able to provide credit that other banks cannot because of the relationship bank's informational advantage from past contact. Under transactions lending technologies, by contrast, loans are underwritten primarily on the basis of information collected at the time of the loan application.

The research model for small business lending has become much more sophisticated since its inception. Perhaps the most important change is that the model has been broadened to include many more lending technologies that may be used to lend to opaque small businesses, each based on a different combination of 'hard' (quantitative) and 'soft' (qualitative) information. For example, small business credit scoring can be used to provide credit to very opaque firms by relying on hard data derived largely from consumer information about the owners of these firms. Notably, however, this technology also allows for judgmental overrides based on soft information known to the loan officer. The lending technologies may be thought of as the basic building blocks of the modern research model of small business lending. Virtually all of the analyses of credit availability, contract terms, and type of bank that provides the funding revolve around the lending technologies employed in the underwriting process.

The research on small business lending raises a number of issues of research and policy importance. One such issue is the effect of the consolidation of the banking industry on small business credit availability, particularly for opaque firms that might rely on relationship lending. As discussed below, the early research suggested potentially serious declines in small business lending as a result of consolidation, but the current research on this topic yields more ambiguous predictions.

Another key issue is the effect of technological progress on small business lending. Technological progress may increase overall small business credit availability through the innovation of new lending technologies, such as small business credit scoring, and through improvements in existing lending technologies. Enhancements in information technology more broadly defined may also improve the ability of banks to process and transmit hard information over distances between the bank and the small business, and between the loan officer and the management of the bank. This may allow for more and better use of hard-information lending technologies to serve opaque small businesses, particularly those located distantly from the bank.

We acknowledge that there are many other important issues regarding small business lending, including the effects of banking crises and credit crunches, the

business cycle, monetary policy, the interest rate cycle, and the regulatory and legal rules and conditions regarding bank lending. However, space constraints restrict the number of topics that can be adequately covered here.

The remainder of the chapter is organized as follows. The second section discusses the lending technologies that form the building blocks of the modern research model. We also discuss the information on which these technologies are based and the types of borrowers these technologies are designed to serve.

The third section discusses the issue of bank consolidation and small business lending. We include summaries of research on the comparative advantages of different types of banks in small business lending—small versus large banks, single-market versus multimarket institutions, and domestically-owned versus multinational organizations. Consolidation tends to shift banking resources from the former set of banks (small, single-market, and domestically-owned) to the latter set of organizations (large, multimarket, and multinational, respectively). In addition, we cover research on the effects of changes in competitiveness associated with mergers and acquisitions (M&A) on small business credit availability. The discussion in this section is also designed to help illustrate some of the main elements of the current research model of small business lending.

The fourth section discusses the 'hardening' of small business lending information over time. We cover research that shows that distances between small businesses and their lending institutions has generally increased over time and the method of contact between the firm and its institution has become more impersonal over time, consistent with a greater use of hard information in lending. The fifth section concludes.

Lending technologies

Classification of the technologies

A lending technology is a unique combination of the primary information source used in the underwriting process, a set of screening and underwriting policies and procedures, a loan contract structure, and monitoring strategies and mechanisms. A technology is typically identified by the primary source of information employed in the credit underwriting process, but we do not rule out that some important information may be generated using other technologies in secondary roles. Thus, a commercial mortgage would generally be classified as generated using the commercial real estate lending technology, even if the bank also obtained a credit score and used information from an existing relationship as secondary information in

underwriting the credit. This taxonomy of lending technologies is largely based on Berger and Udell (2006).

We also distinguish between hard- and soft-information technologies. Hard-information technologies are based principally on quantitative data that may be relatively easily processed and transmitted within a banking organization. Examples of hard information include valuations of collateral, financial ratios from certified audited financial statements, and credit scores generated by outside parties. Soft-information technologies, in contrast, are based mainly on qualitative information that may not be easily processed and transmitted beyond the loan officer or other bank employee that collects it. Examples include information on the character and reliability of the owner of the firm, and the personal experience and training of the loan officer that helps in judging the firm's creditworthiness.

There are at least ten lending technologies used by banks to lend to small businesses—leasing, commercial real estate lending, residential real estate lending, motor vehicle lending, equipment lending, asset-based lending, financial statement lending, small business credit scoring, relationship lending, and judgment lending. All of these technologies employ some combination of both hard and soft information. At a minimum, underwriting any loan requires some numbers about the firm, the owner, and/or the collateral (hard information), and some judgment of the loan officer based on experience and training (soft information). A key implication of this fact is that even if large banks have a comparative advantage in processing and transmitting hard information due to economies of scale, large institutions will not necessarily have an advantage in all of the hard-information technologies. The reason is that their comparative advantage in using the hard-information component of a technology may be offset by their comparative disadvantage in using the soft-information component. Similarly, an advantage for small banks in processing soft information due having fewer layers of management over which to transmit the soft information may not always translate into an advantage for small institutions in every soft-information technology because of a disadvantage in the hard-information component. The comparative advantages may also depend on the type of firm being served. It may be expected that the hard-information component would generally be greater for larger firms and the soft-information component would tend to be greater for smaller firms. It is an empirical question as to whether large or small banks have a comparative advantage in using a given technology to lend to given class size of firms.

Turning to the specific technologies, fixed-asset lending is a set of technologies that are based primarily on the values of fixed assets that are leased or pledged as collateral. Fixed assets are long-lived assets that are not sold in the normal course of business (i.e., are 'immovable'), and are uniquely identified by a serial number or a deed. These include commercial and residential real estate, motor vehicles, and equipment. Leasing is considered to be a fixed-asset lending technology, because the leased assets are generally fixed. Other fixed-asset technologies include

commercial real estate lending, residential real estate lending, motor vehicle lending, and equipment lending, which are based primarily on the valuations of the corresponding fixed assets pledged as collateral. Fixed-asset lending technologies may be applied to both transparent and opaque small businesses, as long as the firms have easily valued fixed assets to pledge.

Asset-based lending is another hard-information technology based principally on the value of collateral, in this case accounts receivable and/or inventory. The amount of credit extended is linked to the estimated liquidation value of the assets, so that the credit exposure is always below the estimated liquidation value. This technology also may be applied to both transparent and opaque small businesses, as it is based primarily on the collateral, rather than on the ability of the firm to generate cash flow to repay the loan.

Financial statement lending is a hard-information technology based primarily on the strength of a borrower's financial statements and the quality of those statements. The latter condition generally implies that the statements must be audited by an outside accounting firm. In addition to having informative financial statements, the borrower must have a sufficiently strong financial condition as reflected in the financial ratios calculated from these statements to justify credit. Unlike the fixed-asset and asset-based lending technologies, financial statement lending is based primarily on an assessment of the firm's ability to repay, rather than the value of the collateral that may be taken in the event of non-payment. Unlike all the other lending technologies, financial statementlending is limited to relatively transparent firms.

Small business credit scoring is a technology based primarily on hard information about the firm's owner as well as the firm. Owner information from consumer credit bureaus is combined with data on the firm collected by the bank and often from commercial credit bureaus to produce a score, or summary statistic for the loan. The models are usually designed for credits up to $250,000, but many institutions use them only for credits up to $100,000. In most cases, the scores are purchased from an outside party, rather than generated by the bank.[2] As already discussed, small business credit scoring may be applied to very opaque small firms.

Relationship lending relies primarily on soft information gathered through contact over time with the firm, its owner, and its local community. Information may be acquired from the provision of loans, deposits, and other services to the firm and to the owner over time. Much of the soft information may be acquired through personal contact by the loan officer with the firm, its owner, local suppliers, and customers, and so forth. Relationship lending may be used for relatively opaque small businesses without significant hard information available.

[2] These credit limits and details are for the US and may vary from country to country.

Another soft-information technology employed by banks to lend to small businesses is judgment lending, which is based primarily on the loan officer's training and personal experience. When firms do not have sufficient hard information on which to base their credit and they have not established a strong relationship to generate soft information, their loans may require a high degree of judgment on the part of the loan officer. The officer makes a judgment based on whatever limited information is available about the firm, plus the officer's training and personal experience with regard to the type of business, location, local demand for the product, and so forth. The training and experience of the loan officer are primarily soft information, as they generally cannot be reduced to credible hard numbers that may be easily communicated. Similar to relationship lending, judgment lending may be applied to opaque small businesses without significant available hard information.[3]

Empirical studies of the technologies

Most empirical studies of small business lending identify only one or two lending technologies or simply analyze small business lending without identifying the technologies employed. Two of the lending technologies are extensively studied. A number of studies examine the effects of relationship strength, generally measured by the duration, breadth, or exclusivity of the relationship, or whether the institution is the firm's 'main' bank. The technology of relationship lending is generally not explicitly identified, but stronger relationships are assumed to be more often associated with this technology. Most of these studies find benefits to borrowers from stronger relationships. The research often finds that stronger relationships are associated with better credit availability (e.g., Petersen and Rajan, 1994; Petersen and Rajan, 1995; Berger and Udell, 1995; Cole, 1998; Elsas and Krahnen, 1998; Harhoff and Korting, 1998; and Machauer and Weber, 2000), although the effects on loan interest rates are mixed (e.g., Petersen and Rajan, 1994; Berger and Udell, 1995; Elsas and Krahnen, 1998; Harhoff and Korting, 1998; Machauer and Weber, 2000; and Degryse and Cayseele, 2000). Other studies find benefits from strong relationships during banking crises (e.g., Horiuchi and Shimuzu, 1998; Watanabe, 2006; Park, Shin, and Udell, 2007; and Jiangli, Unal, and Yom, 2008). Some recent studies also discover favorable effects of strong relationships on firm performance in terms of recovery from distress and bankruptcy, and fostering innovation (Dahiya, et al., 2003; Herrera and Minetti, 2007; and Rosenfeld, 2007).

Strong relationships—particularly when they are exclusive—may also involve costs. The private information generated by an exclusive banking relationship

[3] Judgment lending was first introduced by Berger and Black (2008).

may give the bank market power over the firm, yielding a 'hold-up' problem and extraction of rents from the firm (e.g., Sharpe, 1990; and Rajan, 1992). Firms may mitigate the rent extraction by engaging in multiple relationships (e.g., von Thadden, 1992; Boot, 2000; and Elsas, Heinemann, and Tyrell, 2004), by adding a relationship at the margin (Farinha and Santos, 2002), and/or by paying a higher interest rate at a different bank (Degryse and van Cayseele, 2000).

The small business credit scoring technology has also been extensively studied. Banks that use this technology are identified based on survey data regarding whether, when, and how US banks employ this lending technology (Frame, Srinivasan, and Woosley, 2001; and Cowan and Cowan, 2006). Banks appear to differ significantly in how they use credit scoring. Some institutions essentially follow 'rules' and use the scores automatically to accept or reject loan applications and set loan terms (subject to judgmental overrides). Other banks use more 'discretion' and combine the scores with information generated using other technologies. The use of 'rules' probably reduces underwriting costs significantly, and the use of 'discretion' may add costs, but also provide more information. Some studies find an increase in lending associated with the technology, but this increase appears to be primarily by 'rules' banks, and is probably driven by lower costs (e.g., Frame, Srinivasan, and Woosley, 2001; Frame, Padhi, and Woosley, 2004; and Berger, Frame, and Miller, 2005). Several studies also find results consistent with the hypothesis that use of small business credit scoring with 'discretion' significantly reduces informational opacity—specifically, 'discretion' banks may be associated with reduced borrower risk (Berger, Frame, and Miller, 2005), longer maturities (Berger, et al., 2005), and reduced use of collateral (Berger, et al., 2007). Finally, some studies also find that small business credit scoring tends to be used for more distant or 'out-of-market' borrowers, consistent with the use of hard information that requires relatively little personal contact (e.g., Frame, Padhi, and Woosley, 2004; and DeYoung, Glennon, and Nigro, 2008). Another study finds that the recent increase in out-of-market small business lending is concentrated in loans of $100,000 or less—the limit on small business credit scoring amounts imposed by many banks—consistent with the use of small business credit scoring as a key technology for providing small business credit at a distance (Brevoort, 2006).

Some recent studies identify the use of multiple technologies. One study identifies all five of the fixed-asset lending technologies—leasing, commercial real estate lending, residential real estate lending, motor vehicle lending, and equipment lending—from the loan contract data in the 1998 Survey of Small Business Finance (SSBF) (Berger and Black, 2008). The authors find that more than half of the loans in the survey can be identified as made using the fixed-asset lending technologies. The identification procedure uses only information on whether the contract type was a lease and the type of fixed asset pledged as

collateral. Similar to the relationship lending literature, the authors also examine the effects of relationship strength using lines of credit that are not scored by fixed assets. This method may be more accurate in identifying the effect of relationship strength than the conventional relationship lending literature because of the removal from the sample of the loans made using the fixed-asset lending technologies. A study of Japanese firms identifies six lending technologies—financial statement lending, equipment lending, real-estate-based lending, relationship lending, leasing, and factoring using information from the borrowing firms (Uchida, Udell, and Yamori, 2006).

Effects of bank consolidation on small business credit availability

Early research on consolidation and small business lending

Some of the early research on small business lending suggested a generally unfavorable effect of the consolidation of the banking industry on small business credit availability, particularly for opaque firms that might rely on relationship lending. Studies testing the effect of bank size on the supply of small business credit find that large banks allocate much lower proportions of their assets to small business loans than do small banks (e.g., Berger, Kashyap, and Scalise, 1995; Keeton, 1995; and Strahan and Weston, 1996). A second key finding is that the ratio of small business loans to assets declines after large banks are involved in M&A (e.g., Berger, et al., 1998; Peek and Rosengren, 1998; and Strahan and Weston, 1998). Thus, large banks were thought to be disadvantaged in relationship lending, with a potential consequence of significantly reduced credit availability to informationally opaque small businesses as a result of consolidation.

This early research may be misleading for at least two main reasons. First, although the research finds that consolidating institutions often substantially reduce their ratios of small business loans to total assets, this does not necessarily imply that small business lending by these banks declines significantly. The ratios may decrease primarily because of an increase in other assets in the denominators of the ratios, such as large business loans. Consistent with this possibility, some evidence suggests that small businesses' ability to borrow is unrelated to the presence of large banks in their markets (Jayarante and Wolken, 1999). Other findings suggest that the likelihood of borrowing from a large bank is roughly proportional to the local deposit market share of large banks, consistent with small

businesses simply borrowing from the most convenient bank, independent of bank size (Berger, Rosen, and Udell, 2007).

Second, even if M&A do significantly reduce the supplies of small business credit of the consolidating banks, there may be offsetting 'external effects' or general equilibrium effects in the local market. Empirical evidence suggests that other incumbent banks in the same local market substantially increase their supplies of small business credit after M&A (e.g., Berger, et al., 1998; and Avery and Samolyk, 2004). In addition, newly chartered banks—which tend to specialize in small business lending—often enter the market after M&A activity, potentially offsetting any cutbacks in small business lending by consolidating banks (e.g., Berger, et al., 2004).[4]

Current research on consolidation and small business lending

Under the current research model of small business lending, the effects of banking industry consolidation on credit availability to opaque small businesses is even more ambiguous for two additional reasons. First, even if the consolidated banks have a comparative disadvantage in relationship lending, they may have advantages in hard-information technologies that may be used to lend to opaque small businesses. Second, consolidation may affect the competitiveness of markets for small business borrowers, which may have either favorable or unfavorable effects on small business credit availability.

In remainder of this section, we first discuss the comparative advantages of different types of banks in the technologies used in small business lending. We look at the advantages of small versus large banks, single-market versus multimarket institutions, and domestically-owned versus multinational organizations. Consolidation often shifts banking resources from the former set of banks (small, single-market, domestically-owned) to the latter set of organizations (large, multimarket, multinational, respectively). We then review the findings on the effects of restrictions on competition on small business credit availability.

Large institutions are likely to have a comparative advantage in hard-information technologies and small institutions are likely to have the advantage in soft-information technologies. Large banks may be able to exploit economies of scale in the processing and transmission of hard information within the bank, but be

[4] Notably, these findings are—as are most of the results reported here—based on US data, and may not apply to other nations, particularly developing nations, where opacity problems are worse and hard information is more often lacking. An international comparison finds greater market shares for large banks are associated with lower small business employment and less overall bank lending (Berger, Hasan, and Klapper, 2004).

relatively poor at processing and transmitting soft information through the communication channels of large organizations (e.g., Stein, 2002). An additional problem for large banks with soft-information technologies may be the number of layers of management required for loan approval. This is because the loan officer is the prime repository of the soft information that cannot be easily communicated, giving a comparative advantage to small institutions with fewer layers of management (e.g., Berger and Udell, 2002) or less hierarchical distance between the loan officer and the manager that approves the loans (e.g., Liberti and Mian, forthcoming). Finally, large banks may suffer Williamson-type (Williamson, 1988) organizational diseconomies associated with providing hard-information loans to more transparent large businesses together with soft-information loans to less transparent small businesses.

Recent empirical research is consistent with the hypothesis that large and small banks have comparative advantages in using hard and soft information, respectively (e.g., Cole, Goldberg, and White, 2004; Scott, 2004; and Berger, et al., 2005). However, these advantages do not necessarily extend to all of the individual hard- and soft-information technologies, and do not necessarily apply to all types of firms. As discussed above, all technologies incorporate some hard and some soft information and the advantage of a large bank in the hard information may be overwhelmed by the advantage of a small bank in the soft information, and vice versa. To illustrate, one empirical study finds that large banks have a greater comparative advantage in leasing relative to other fixed-asset technologies but, this advantage is dissipated for the smallest firms in the sample (Berger and Black, 2008).

The arguments regarding single-market versus multimarket banks are similar to those regarding the size of banks. Single-market banks are likely to have a comparative advantage over multimarket banks in using soft information because of their physical proximity to small business customers (Degryse and Ongena 2005). Single-market institutions may be better able to play the role of 'community bank' that knows the local borrowers, their customers and suppliers, and local business conditions better than multimarket competitors (DeYoung, Hunter, and Udell, 2004). In addition, single-market banks may also have an advantage in processing soft information because of the physical proximity of the loan officer to the management of the bank that must approve the credits. It may be easier to transmit soft information to someone in the same location that may also have knowledge of the local conditions. Some recent empirical evidence is consistent with these arguments. One study of the lending of a US bank finds that borrower proximity facilitates the production of proprietary information by the bank, which gives the bank significant advantages over competitors (Agarwal and Hauswald, 2006). A study of an Argentine bank finds that soft information was most difficult to use when the loan officer and manager that approves the loans are located in different offices of the bank (Liberti and Mian, forthcoming). Similarly, it may be more difficult to transmit soft information between different regions of a country with

significant cultural differences. Consistent with this, one study of Italian banking finds more credit rationing of small businesses when their banks are headquartered in another province (Alessandrini, Presbitero, and Zazzaro, 2006). As discussed in the fourth section below, lending distances are increasing over time, consistent with a 'hardening' of information and reduced potential reduced importance of the distinction between single-market and multimarket banks over time.

The same arguments regarding size and geography generally apply to foreign bank ownership, as foreign-owned banks are generally quite large, have headquarters that are geographically distant, and often have different cultures and languages from the host nation. Thus, it is expected that foreign-owned banks have comparative advantages in hard-information technologies and disadvantages in soft-information technologies relative to domestically-owned banks. There is little evidence on the use of technologies, but the empirical research generally suggests that foreign banks make relatively few small business loans in developed nations, but may increase small business credit availability in developing nations due to access to superior hard-information technologies. See Berger and Udell (2006) for a summary of this research.

Finally, consolidation may affect the competitiveness of markets for small business borrowers, with M&A within markets probably reducing competition and M&A across markets more likely increasing competition. Reduced competition would restrict the supply of small business credit through any technology under the standard structure-conduct-performance hypothesis, but it may increase the supply through relationship lending. This is because limits on competition help banks enforce implicit contracts in which relationship borrowers receive subsidized rates in the short term, and pay higher rates in later periods (e.g., Sharpe, 1990; and Petersen and Rajan, 1995). The empirical evidence on this point is mixed, with some studies finding favorable effects of concentration and other restrictions on competitiveness on measures of credit availability, activity, and general economic performance (e.g., Petersen and Rajan, 1995; Cetorelli and Gambera, 2001; Bonaccorsi di Patti and Dell'Ariccia, 2004; and Cetorelli, 2004), others finding unfavorable effects (e.g., Black and Strahan, 2002; Berger, Hasan, and Klapper, 2004; Karceski, Ongena, and Smith, 2005; and Cetorelli and Strahan, 2006), and some finding different effects based on alternative measures of competition (e.g., Carbo-Valverde, Rodriguez-Fernandez, and Udell, 2007).[5]

[5] Consolidation may also affect the lending technologies used by local rivals. One study of Belgian banking finds that when local banks are larger, their rival bank tends to lend over a smaller geographic area, consistent with a focus on soft-information technologies (Degryse, Laeven, and Ongena, 2006).

The 'hardening' of small business lending information over time

A potentially important development in small business lending is the 'hardening' of the information used in making these loans. As discussed below, there is evidence that on average, the distance between small business borrowers and their banks has increased and the percentage of borrowers that have personal contact with their banks has decreased. These findings are consistent with a greater use of hard information in lending or a 'hardening' of the information used, given that soft information is difficult to learn and transmit over long distances and through impersonal methods of contact. For example, a loan officer often needs to have face-to-face contact with the small business owner and members of the local community to gather soft information to use in the relationship lending technology.

One reason for the hardening of information is technological progress. Many studies have documented significant technological progress in the banking industry, as banks take advantage of improvements in information processing, telecommunications, and financial technologies (e.g., Berger, 2003). The improvements in information processing and telecommunication technologies probably have improved banks' abilities to process and transmit over longer distances hard quantitative information about loan customers. New financial technologies that use this information, such as small business credit scoring, may have further facilitated the ability of banks to expand their range of lending. Some research has specifically linked the use of the relatively new small business credit scoring technology to additional out-of-market lending and longer-distance lending (e.g., Frame, Padhi, and Woosley, 2004; DeYoung, et al., 2007; and DeYoung, et al., 2008) and reduced default rates on longer-distance small business loans (e.g., DeYoung, Glennon, and Nigro, 2008). It seems unlikely that technological change has had as much effect in improving soft-information technologies, which are by their nature more labor intensive, and the qualitative data are less subject to improvements in processing and transmission.

A second reason for the hardening of information is the consolidation of the banking industry. As discussed above, large financial institutions may have comparative advantages in using hard-information lending technologies. Thus, consolidation may have increased lending distances because the hard-information lending technologies in which large banks specialize tend to be associated with longer distances and more impersonal contact methods. Recent research is consistent with the notion that large banks are associated with longer-distance small business loans (e.g. Berger, et al., 2005; and Brevoort 2006).

Similarly, the hardening of information may be related to a shift in lending among the different types of financial institutions. Recent research suggests that increasing

proportions of small business loans are made by non-depository institutions (finance and factoring, brokerage and pension, leasing, insurance, and mortgage companies) rather than depositories (commercial banks, thrifts, and credit unions) and that non-depositories tend to lend at longer distances and use more impersonal means of contact with small businesses (Brevoort and Wolken, 2008).

The hardening of information and increase in lending distance over time may also be related to a shift in the mix of lending technologies. There is evidence that soft information is associated with relatively short distances between borrowers and lenders, as loan officers need a close geographic proximity to observe soft information such as owner character and reliability (e.g., Degryse and Ongena, 2005). Lending distances may be expected to increase to the extent that financial institutions shift from 'softer' lending technologies that are associated with relatively short distances to 'harder' technologies associated with longer distances.

There may also be important complementarities among technological progress, consolidation, and the shift to harder lending technologies. It is likely that large banks gained more from technological progress than small banks for a number of reasons, including that technological progress probably improved the hard-information lending technologies in which these banks specialize more than the soft-information technologies in which small banks specialize. Other empirical research on bank performance is consistent with this hypothesis, finding that, over time, large banks: (i) improved their productivity more than small banks (e.g., Berger and Mester, 2003); (ii) reduced their agency costs of managing affiliates at greater distances (e.g., Berger and DeYoung, 2006); and (iii) competed more effectively against small banks (e.g., Berger, et al., 2007).[6] It may also be argued that the shift to harder-lending technologies is related to both technological progress and consolidation. Technological progress may have resulted in new hard-information technologies, such as small business credit scoring, and lowered the relative cost of other hard-information technologies relative to soft-information technologies. Bank consolidation may also facilitate shifting into harder technologies, as large banks tend to have comparative advantages in these technologies.

In the remainder of this section, we briefly review some of the research on the changes in small business lending distance over time and give some recent updates from the 1993, 1998, and 2003 Surveys of Small Business Finances. For both the literature review and the recent data, more information may be found in Brevoort and Wolken (2008).

Petersen and Rajan (2002) use the 1993 SSBF to construct a synthetic panel based upon the year in which the lender–borrower relationship started and find that the average distance between small firms and their lenders increased by 3.4 percent

[6] However, there may be limits to consolidation. It may be argued that consolidation may not proceed beyond the point where there are sufficient small banks to provide relationship lending (DeYoung, Hunter, and Udell, 2004).

per year from the 1970s to the early 1990s. They also find that the most frequent method of contact between the small business and its lender was less often personal and more often by phone or mail. Wolken and Rohde (2002) compare lending distances between the 1993 and 1998 SSBFs and find that the average firm–lender distance increased from 115 miles in 1993 to 244 miles in 1998, an annual growth rate of 15 percent. However, the median distance only increased from nine miles in 1993 to ten miles in 1998, suggesting that the increase in mean distance was largely from increases in out-of-market distances that affected a minority of small businesses.

Brevoort and Hannan (2006) focus on in-market distances using Community Reinvestment Act data from 1997 to 2001, and find little change in distance over the sample period. Hannan (2003) uses Community Reinvestment Act data from 1996 to 2001 and finds a significant increase in out-of-market lending. He also finds that more loans in metropolitan markets are extended by out-of-market lenders over time. Brevoort (2006) uses Community Reinvestment Act data from 1998 to 2003 and finds a large increase in out-of-market commercial lending. However, Brevoort also finds that the effects are limited to large banks and loans of $100,000 or less, consistent with the effects of small business credit scoring by large banks to lend to small borrowers.

Using data on US Small Business Administrations 7(a) loan program and the 1998 Atlanta Federal Reserve Survey on the use of the small business credit scoring technology, DeYoung, et al. (2007) and DeYoung, Glennon, and Nigro (2008) find that average distances between small business borrowers and their lenders grew between 1984 and 2001 and that these observed increases were larger at banks that had adopted credit scoring by the time of the 1998 survey. These results are consistent both with the earlier observed increases in mean distances and with the likelihood that the adoption of the small business credit scoring technology has played an important role in the increase.

Finally, we turn to the most recent information on changes in lending distance and personal contact with small businesses from Brevoort and Wolken (2008), who provide detailed comparisons of the data from the 1993, 1998, and 2003 SSBFs. They find that mean lending distance more than doubled in five years, from 110.6 miles in 1993 to 242.9 miles in 1998, and then surprisingly fell to 180.6 miles in 2003. The exact reason for the drop between 1998 and 2003 is not known, but apparently it is concentrated at the top end of the distribution, since the median lending distance rose modestly from eight miles to eleven miles, and the loans of over thirty miles increased slightly from 34 percent to 34.6 percent of the total. The proportion of financial institutions conducting business in person with small business borrowers dropped from 49 percent in 1993 to 48 percent in 1998 to 44 percent in 2003, consistent with a continuing hardening of information over time. As noted above, the data also show important differences in lending distances and personal versus impersonal contact between depositories (commercial banks, thrifts, and credit unions) and non-depository institutions (finance and factoring, brokerage and pension, leasing, insurance, and mortgage companies). For example, in 2003, the mean distance to

depository lenders was 74.6 miles versus 357.4 miles for non-depositories, and the percent conducting business in person was 71 percent for depositories and only 15 percent for non-depositories, consistent with a much greater use of soft information by depositories and a much greater use of hard information by non-depositories. This is again consistent with expectations that loan officers at commercial banks and other depositories tend to specialize in relationship lending and judgment lending, while other types of financial institutions rely more on hard-information techniques. Finally, the data in Brevoort and Wolken (2008) suggest some interesting differences in the use of hard and soft information in different lending products that are consistent with expectations. For example, in 2003, the mean distance for lines of credit was 77.1 miles, while the mean distance for leases was 438 miles. This is consistent with prior findings that lines of credit are associated with relationship lending (e.g., Berger and Udell, 1995) and with prior arguments that leasing is one of the 'hardest' lending technologies (e.g., Berger and Black, 2008).

Conclusions

This chapter covers some of the issues regarding bank lending to small businesses. We briefly discuss the lending technologies used by banks to lend to small businesses. These technologies form the building blocks of the modern research model of small business lending. We also look at the effects of banking industry consolidation and technological progress on the use of the lending technologies and their effects on small business credit. We find, for example, that the effects of consolidation on small business credit availability is ambiguous for several reasons, including the possibility that the consolidated banks may have comparative advantages in hard-information lending technologies that may be used to lend to opaque small businesses. We also find that consolidation and technological progress and their interactions appear to have resulted in a 'hardening' of small business lending information over time. This is reflected in greater distances between financial institutions and their loan customers and greater use of impersonal methods of contact between the parties.

References

Agarwal, S. and Hauswald, R. (2006). 'Distance and Information Asymmetries in Lending Decisions', Working Paper. Federal Reserve Bank of Chicago.

Alessandrini, P., Presbitero, A. F., and Zazzaro, A. (2006). 'Banks, Distances and Financing Constraints for Firms'. Working Paper. Università Politecnica delle Marche.

Avery, R. B. and Samolyk, K. A. (2004). 'Bank Consolidation and the Provision of Banking Services: Small Commercial Loans', *Journal of Financial Services Research*, 25: 291–325.

Berger, A. N. (2003). 'The Economic Effects of Technological Progress: Evidence from the Banking Industry', *Journal of Money, Credit, and Banking*, 35: 141–76.

—— and Black, L. K. (2008). 'Bank Size, Lending Technologies, and Small Business Finance.' Working Paper, University of South Carolina.

—— Bonime, S. D., Goldberg, L. G., and White, L. J. (2004). 'The Dynamics of Market Entry: The Effects of Mergers and Acquisitions on Entry in the Banking Industry', *Journal of Business*, 77: 797–834.

—— and DeYoung R. (2006). 'Technological Progress and the Geographic Expansion of the Banking Industry', *Journal of Money, Credit, and Banking*, 38: 1483–513.

—— Dick, A. A., Goldberg, L. G., and White, L. J. (2006). 'Competition from Large, Multimarket Firms and the Performance of Small, Single-Market Firms: Evidence from the Banking Industry', *Journal of Money, Credit, and Banking*, 39: 331–68.

—— Espinosa-Vega, M. A., Frame, W. S., and Miller, N. H. (2005). 'Debt, Maturity, Risk, and Asymmetric Information', *Journal of Finance*, 60: 2895–923.

———— (2007). 'Why Do Borrowers Pledge Collateral? New Empirical Evidence on the Role of Asymmetric Information'. Working Paper. Federal Reserve Bank of Atlanta.

—— Frame, W. S., and Miller, N. H. (2005). 'Credit Scoring and the Availability, Price, and Risk of Small Business Credit', *Journal of Money, Credit, and Banking*, 37: 191–222.

—— Hasan, I., and Klapper, L. F. (2004). 'Further Evidence on the Link between Finance and Growth: An International Analysis of Community Banking and Economic Performance', *Journal of Financial Services Research*, 25: 169–202.

—— Kashyap, A. K., and Scalise, J. M. (1995). 'The Transformation of the U.S. Banking Industry: What a Long, Strange Trip It's Been', *Brookings Papers on Economic Activity*, 2: 55–218.

—— and Mester, L. J. (2003). 'What Explains the Dramatic Changes in Cost and Profit Performance of the U.S. Banking Industry?', *Journal of Financial Intermediation*, 12: 57–95.

—— Miller, N. H., Petersen, M. A., Rajan, R. G., and Stein, J. C. (2005). 'Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks', *Journal of Financial Economics*, 76: 237–69.

—— Rosen, R. J., and Udell, G. F. (2007). 'Does Market Size Structure Affect Competition? The Case of Small Business Lending', *Journal of Banking & Finance*, 31: 11–33.

—— Saunders, A., Scalise, J. M., and Udell, G. F. (1998). 'The Effects of Bank Mergers and Acquisitions on Small Business Lending', *Journal of Financial Economics*, 50: 187–229.

—— and Udell, G. F. (1995). 'Relationship Lending and Lines of Credit in Small Firm Finance', *Journal of Business*, 68: 351–81.

———— (1998). 'The Economics of Small Business Finance: The Roles of Private Equity and Debt Markets in the Financial Growth Cycle', *Journal of Banking & Finance*, 22: 613–73.

———— (2002). 'Small Business Credit Availability and Relationship Lending: The Importance of Bank Organizational Structure', *Economic Journal*, 112: 32–53.

———— (2006). 'A More Complete Conceptual Framework for SME Finance', *Journal of Banking & Finance*. 30: 2945–66.

Black S. E. and Strahan, P. E. (2002). 'Entrepreneurship and Bank Credit Availability', *Journal of Finance*, 57: 2807–33.

Boot, A. W. A. (2000). 'Relationship Banking: What Do We Know?', *Journal of Financial Intermediation,* 9: 7–25.

Bonaccorsi di Patti, E. and Dell'Ariccia, G. (2004). 'Bank Competition and Firm Creation', *Journal of Money, Credit, and Banking*, 36: 225–51.

Brevoort, K. P. (2006). An Empirical Examination of the Growth in Out-of-Market Lending: The Changing Competitive Landscape and the Role of Asymmetric Information. Working Paper. Federal Reserve Board.

—— and Hannan, T. H. (2006). 'Commercial Lending and Distance: Evidence from Community Reinvestment Act Data', *Journal of Money, Credit, and Banking*, 38: 1991–2012.

—— and Wolken, J. D. (2008). 'Does Distance Matter in Banking?' Working Paper. Federal Reserve Board.

Carbo-Valverde, S., Rodriguez-Fernandez, F., and Udell, G. F. (2007). 'Bank Lending, Financing Constraints and SME Investment', Indiana University Working Paper.

Cetorelli, N. (2004). 'Bank Concentration and Competition in Europe', *Journal of Money, Credit, and Banking*, 36: 543–58.

—— and Gambera, M. (2001). 'Banking Market Structure, Financial Dependence and Growth: International Evidence from Industry Data', *Journal of Finance*, 56: 617–48.

—— and Strahan, P. E. (2006). 'Finance as a Barrier to Entry: Bank Competition and Industry Structure in Local U.S. Markets', *Journal of Finance*, 61: 437–61.

Cole, R. A. (1998). 'The Importance of Relationships to the Availability of Credit', *Journal of Banking & Finance*, 22: 959–77.

—— Goldberg, L. G., and White, L. J. (2004). 'Cookie-Cutter versus Character: The Micro Structure of Small Business Lending by Large and Small Banks', *Journal of Financial and Quantitative Analysis*, 39: 227–51.

Cowan, C. D. and Cowan, A. M. (2006). 'A Survey-Based Assessment of Financial Institution Use of Credit Scoring for Small Business Lending', Small Business Administration, Office of Advocacy Report 283.

Dahiya, S., Bharath, S., Saunders, A., and Srinivasan, A. (2007). 'So What Do I Get? The Bank's View of Lending', *Journal of Financial Economics*, 85: 368–419.

—— John, K., Puri, M., and Ramirez, G. (2003). 'Debtor-in-Possession Financing and Bankruptcy Resolution: Empirical Evidence', *Journal of Financial Economics,* 69: 259–80.

Degryse, H., Laeven, L., and Ongena, S. (2006). 'The Impact of Organizational Structure and Lending Technology on Banking Competition'. Working Paper. Tilburg University.

—— and Ongena, S. (2005). 'Distance, Lending Relationships, and Competition', *Journal of Finance*, 60: 231–66.

—— and Van Cayseele, P. (2000). 'Relationship Lending within a Bank-Based System: Evidence from European Small Business Data', *Journal of Financial Intermediation*, 9: 90–109.

DeYoung, R., Frame, W. S., Glennon, D., McMillen, D. P., and Nigro, P. (2008). 'Commercial Lending Distance and Historically Underserved Areas', *Journal of Economics & Business*, 60: 149–64.

—— —— —— —— (2007). 'What's Driving Small Borrower-Lender Distance?' Working Paper. Federal Reserve Bank of Atlanta.

—— Glennon, D., and Nigro, P. (2008). 'Borrower-Lender Distance, Credit Scoring, and the Performance of Small Business Loans', *Journal of Financial Intermediation*, 17: 113–43.

—— Hunter, W. C., Udell, G. F. (2004). 'The Past, Present, and Probable Future for Community Banks', *Journal of Financial Services Research,* 25: 85–133.

Elsas, R., Heinemann, F., and Tyrell, M. (2004). 'Multiple but Asymmetric Bank Financing: The Case of Relationship Lending'. Working Paper. CESifo Series 1251.

—— and Krahnen, J. P. (1998). 'Is Relationship Lending Special? Evidence from Credit-File Data in Germany', *Journal of Banking & Finance*, 22: 1283–316.

Farinha, L. A., and Santos, J. A. C. (2002). 'Switching from Single to Multiple Bank Lending Relationships: Determinants and Implications', *Journal of Financial Intermediation*, 11: 124–51.

Frame, W. S., Padhi, M., and Woolsey, L. (2004). 'The Effect of Credit Scoring on Small Business Lending in Low- and Moderate-Income Areas', *Financial Review*, 39: 35–54.

—— Srinivasan, A., and Woosley, L. (2001). 'The Effect of Credit Scoring on Small Business Lending', *Journal of Money, Credit, and Banking*, 33: 813–25.

Hannan, T. H. (2003). 'Changes in Non-Local Lending to Small Business', *Journal of Financial Services Research*, 24: 31–46.

Harhoff, D. and Korting, T. (1998). 'Lending Relationships in Germany: Empirical Evidence from Survey Data', *Journal of Banking & Finance*, 22: 1317–53.

Herrera, A. M., and Minetti, R. (2007). 'Informed Finance and Technological Change: Evidence from Credit Relationships', *Journal of Financial Economics*, 83: 223–69.

Horiuchi, A. and Shimizu, K. (1998). 'The Deterioration of Bank Balance Sheets in Japan: Risk-taking and Recapitalization', *Pacific-Basin Finance Journal*, 6: 1–26.

Jiangli, W., Unal, H., and Yom, C. (2008). 'Relationship Lending, Accounting Disclosure and Credit Availability during the Asian Financial Crisis'. *Journal of Money, Credit, and Banking*, 40: 25–55.

Karceski, J., Ongena, S., and Smith, D. (2005). 'The Impact of Bank Consolidation on Commercial Borrower Welfare', *Journal of Finance*, 60: 2043–82.

Liberti, J. M. and Mian, A. R. (forthcoming). 'Estimating the Effect of Hierarchies on Information Use'. *Review of Financial Studies*.

Machauer, A. and Weber, M. (2000). 'Number of Bank Relationships: An Indicator of Competition, Borrower Quality or Just Size?' Working Paper 2000/06. Johan Wolfgang Goethe-Universitat Center for Financial Studies.

Park, S. Y., Shin, B. S., and Udell, G. F. (2007). 'Lending Relationships, Credit Availability and Banking Crises'. Working Paper. Indiana University.

Petersen, M.A. and Rajan, R. G. (1994). 'The Benefits of Lending Relationships: Evidence from Small Business Data', *Journal of Finance*, 49: 3–37.

—— —— (1995). 'The Effect of Credit Market Competition on Lending Relationships', *Quarterly Journal of Economics*, 110: 407–43.

—— —— (2002). 'The Information Revolution and Small Business Lending: Does Distance Still Matter?' *Journal of Finance*, 57: 2533–70.

Rajan, R. G. (1992). 'Insiders and Outsiders: The Choice between Informed and Arms Length Debt'. *Journal of Finance*, 47: 1367–400.

Rosenfeld, C. M. (2007). 'The Effect of Banking Relationships on the Future of Financially Distressed Firms'. Working Paper. University of Minnesota.

Scott, J. A. (2004). 'Small Business and Value of Community Financial Institutions'. *Journal of Financial Services Research*, 25: 207–30.

Sharpe, S. A. (1990). 'Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships', *Journal of Finance*, 45: 1069–87.

Stein, J. C. (2002). 'Information Production and Capital Allocation: Decentralized vs. Hierarchical Firms', *Journal of Finance*, 57: 1891–921.

Uchida, H., Udell, G. F., and Yamori, N. (2006). 'SME Financing and the Choice of Lending Technology', Working Paper. Research Institute of Economy, Trade, and Industry (REITI).

Von Thadden, E. (1992). 'The Commitment of Finance, Duplicated Monitoring, and the Investment Horizon'. Working Paper. Center for Economic Policy Research 27.

Watanabe, W. (2006). 'Does a Large Loss of Bank Capital Cause Evergreening? Evidence from Japan'. Working Paper. Tohoku University.

Williamson, O. (1988). 'Corporate Finance and Corporate Governance', *Journal of Finance*, 43: 567–91.

Wolken, J. and Rohde, D. (2002). 'Changes in the Location of Small Businesses' Financial Services Suppliers between 1993 and 1998', Memo. Federal Reserve Board.

CHAPTER 22

CONSUMER LENDING

THOMAS A. DURKIN
GREGORY ELLIEHAUSEN

Introduction

Beyond deposit services and monetary transfers, the most prevalent financial service in developed countries is consumer lending. Terminology sometimes differs according to the user, but most often the term 'consumer lending' refers to the advance of cash to a consumer by a financial institution or permission by a retail seller of goods and services to delay payment for a purchase. The term normally does not include credit for purchase of a residence or collateralized by real estate or by specific financial assets like stocks and bonds or extended for business financing. Most consumer lending involves repayment in periodic payments sometimes called 'installments', at set intervals such as monthly. Besides loans of cash, credit for purchase of substantial goods and services like automobiles, home improvements, appliances, recreational goods like boats, movable housing, and educations all fall within this definition, as does credit on credit cards. Credit of this kind has been growing worldwide in recent decades; in the US alone more than $2½ trillion of credit arising from consumer lending was outstanding at the end of 2008. This amount is in addition to more than $10 trillion of credit outstanding on real estate collateral there and not counted in 'consumer lending' as the term is used here.

Within this definition of consumer lending, there are many ways to classify the loans, including by purpose of credit use (automobile lending, student education loans, etc.), by institutional source of the funds (e.g., bank, credit union, or store), according to method of credit generation and repayment (closed-end single advances versus multiple-advance revolving credit arrangements like credit cards), and by mechanics of extension (directly from the financial institution or indirectly from a seller of goods that relies on a financial institution for funding). In earlier decades, another classification method was according to agreed timing of repayment; at that time, a further common differentiation was between non-installment credit and installment credit. Non-installment credit referred to single-payment loans, charge accounts at retail stores and dealers without an extended payments feature, and service credit granted by physicians, hospitals, lawyers, and other professionals where payment was expected in one lump sum. Today, credit cards substitute for many kinds of non-installment credit and most consumer lending is installment credit.

Consumer lending is sometimes controversial among people who believe use is merely an attempt by consumers to live beyond their means, but most informed observers agree that consumer lending provides a number of important economic benefits. First, consumer credit use makes purchasing household investment goods and services like automobiles and education easier and more timely for many families. In this context, the term 'household investments' using credit does not refer to financial investment in such assets as stocks or bonds. Rather, it means making expenditures for high-value goods or services that provide their benefits over a period of time and whose cash purchase does not usually fit comfortably into monthly budgets. By facilitating such investment spending, consumer lending enables consumers to change the timing of their saving and consumption flows to a preferred pattern. Specifically, rather than postponing the purchase of household investment goods and services and the consumption benefits they provide until funds are available from savings (a difficult task for many families, especially in the earlier stages of their earning years), consumers have been able to use credit to purchase the investment goods and services first and pay for them while using them. In effect, they can save for them by making payments while actually using the goods and services.

Second, consumer lending has contributed to the growth of durable goods industries where new technologies, mass production, and economies of scale historically have produced employment growth and new wealth. It is simply hard to imagine development of the suburbs or the automobile and appliance industries in the twentieth century, or for that matter the higher education system as it now exists in many places, without the simultaneous rise of consumer credit to facilitate sale of the output.

Third, consumer lending provides an important outlet for employing financial resources available from net surplus components of the economy, notably from

consumers themselves, through the financial intermediation process. Ultimately, the source of funds for consumers who borrow is other (or even the same) consumers who have a financial surplus they can hold as deposits, as life insurance and pension reserves, or as portfolios of securities including bonds, stocks, and mutual fund shares.

Demand for consumer lending

The intertemporal investment-consumption economic model developed by Irving Fisher (1930) and extended by Hirschleifer (1958) and Juster and Shay (1964) provides the neoclassical analytical framework for consumers' borrowing decisions. The investment-consumption framework explored by these economists, including their extensions to encompass uncertainty and credit rationing, relates consumer investment opportunities, time preference, the possibility of lending and borrowing, and the market interest rate to solve the problem of maximizing and allocating consumption over time. It also shows formally when borrowing is a rational economic decision for consumers, as well as for investors in commercial and industrial enterprises (the latter the main focus of much of the theoretical economics derived from the Fisher approach). Since this theory also shows that there are many common circumstances when credit use by consumers is rational, it leads immediately to the inference that there will be widespread rational economic demand for consumer lending. (Durkin, et al., 2010, especially chapters 3–5, discusses many of these issues at greater length.)

The investment-consumption theory based on the work of Fisher provides the formal basis in economics for consumer lending demand, but the fundamentals are also intuitive: an individual will borrow to purchase investment-type goods and services if doing so has a favorable impact on consumption possibilities after repaying the loan with any necessary interest. This is merely an informal rendering of the Fisher/Hirschleifer/Juster–Shay theoretical conclusion: borrowing to undertake the investment is rational for consumers, as for business enterprises, if there is a positive net present value from the investment under consideration. Under the condition of a positive net present value from borrowing and investing, the individual is better off by undertaking the transaction. If, in contrast, borrowing to purchase the investment goods or services does not produce a positive net return, then the rational choice is not to undertake the investment. Limited amounts of empirical work on uses of consumer loans show that returns can be quite high (see, e.g., discussion in Poapst and Waters, 1964; Dunkelberg and Stephenson, 1974; and Elliehausen and Lawrence, 2001).

During the post-World War II period, and certainly since at least the 1960s, when Juster and Shay were formally analyzing the rationality conditions behind consumers' credit use behavior, the view that consumer credit use is a normal development in a modern economy seems also to have gained traction with the public at large. Consumer lending is not without its problems and its critics, however, including analysts advancing hypotheses of fundamental consumer irrationality in credit use, especially focusing in recent years upon the modern phenomenon of credit cards.

Behavioral economists and psychologists actually have studied consumers' credit decisions for decades, especially using consumer survey methodologies pioneered by George Katona and his colleagues John B. Lansing, James N. Morgan, Eva Mueller, and others at the University of Michigan's Survey Research Center, founded by Katona in 1946. For consumer credit, passage of the US 'Truth in Lending' Act in 1968, designed to require transaction-specific information disclosures to borrowers, further stimulated research using psychology-based models to study the role of information in the credit decision process (see Day and Brandt, 1973; Day and Brandt, 1974; and Day, 1976). A bit later, the work by Tversky and Kahneman (1974) and Kahneman and Tversky (1979) on decision-making under uncertainty has further reawakened the interest of economists in psychological influences on consumer behavior, including credit use behavior.

Survey research on the processes of spending supports the theoretical economic analyses that treat consumer credit as a part of consumers' investment-consumption decisions. Although some consumer lending arises from the financial consequences of hardship or distress, such as medical expenses, paying recurring bills, or burden of already existing debts, surveys show that most arises in the consumer investment process involving acquisition of consumer durable goods and services like automobiles, home repairs, and education, which provide for both a return and a repayment process over time. Consistent with the theories of the economists, surveys have found that credit use is greatest in early family life-cycle stages when the rate of return on additional durable goods and services that might be financed using consumer credit is probably quite high.

Another major focus of the survey research has been to investigate the extent to which consumers' durable goods purchasing decisions are deliberative and rational. The research indicated that few purchases include all of the elements of rational decision making—namely planning for purchases, extensive search for information, formulation of evaluation criteria, and careful consideration of alternatives before making decisions. In fact, consumers often simplify, take shortcuts, or use heuristics. Consumers may focus on one or a few product characteristics or rely on the experience of friends or their own experience, for example.

Nevertheless, most consumers use one or more elements of deliberative behavior in decisions about consumer durables and credit. The research also

identified several circumstances that lead to more or less deliberation in durable goods purchases, including purchase of an item that is considered expensive or particularly important, purchase of a new or unfamiliar product, dissatisfaction with a previous purchase, and situations involving a strong new stimulus that causes uncertainty about previous attitudes or experience. In these situations, consumers are more likely to gather additional information, formulate or revise evaluative criteria, and deliberate on alternatives, although they may still take shortcuts, simplify, or use heuristics. Few consumers collect all available information or carefully consider all possible choices. But even in context of the optimizing models of traditional economics, consumers may not want to collect all available information. Consumers will collect additional information only as long as the perceived cost of search is less than its expected benefits (see Stigler, 1961 and Durkin and Elliehausen, 2010: chap. 2). Lowering the cost of search is the main argument in favor of disclosure rules like 'Truth in Lending'.

Supply of consumer lending

Production of consumer lending involves the transfer of funds from savers who have them to borrowers who have need of them, along with the subsequent collection of loan repayments from the borrowers. For consumer lending, the transfers from savers to borrowers and back are usually effected not directly from one to the other, but rather by financial firms through a production process called financial intermediation. As this term suggests, financial intermediaries are institutions that stand between the ultimate suppliers of funds (savers) and the ultimate users (investors). In common parlance, financial intermediaries are usually referred to simply as financial institutions; many kinds are broadly familiar including banks, credit unions, insurance companies, pension funds, mutual funds, and finance companies.

Many financial intermediaries are active in specialized areas of financial markets, and they do not all operate the same way. Among those undertaking consumer lending, banks and credit unions obtain funds directly from consumers and businesses by providing deposit accounts, and they lend the funds obtained back to consumers and businesses, sometimes the same ones. In contrast, other intermediaries involved in consumer lending obtain most or a good portion of their funds from other intermediaries. Finance companies, for example, obtain most of their funds in capital markets from other institutions like insurance companies and pension funds. Ultimate funds sources always are

savers, however, typically consumers, but also businesses and governments, both domestic and foreign.

Financial intermediaries perform several functions that facilitate the transfer of funds from savers to borrowers, none of which individuals probably want, or are able, to provide for themselves. These functions include: (i) information processing; (ii) risk intermediation; (iii) monitoring; (iv) temporal intermediation; and (v) size intermediation (see Benston and Smith, 1976). In performing these functions, financial intermediaries produce distinct financial products for one or both groups of market participants: borrowers, savers, or both. As indicated, banks and credit unions produce products for both savers (deposits) and investors (loans). Finance companies primarily produce products for borrowers, although the securities they issue to obtain their lending funds provide an outlet for other intermediaries. Mutual funds are examples of financial intermediaries that produce primarily a savings product, raising funds from many savers to purchase a diversified portfolio of securities. Through economies of scale and specialization, financial intermediaries are able to perform these functions in financial markets at a lower cost than individuals could do so on their own (see Gurley and Shaw, 1960; and Benston and Smith, 1976). That financial intermediaries use funds obtained either directly from savers or indirectly in financial markets and then use them as inputs to produce their own distinct products for borrowers is what distinguishes financial intermediaries from brokers. Brokers match sellers of a product with buyers—buyers and sellers of a house, for example. Financial intermediaries do not match borrowers and savers but rather obtain funds from one source for use by another, typically in much different form.

Whether or not it involves an intermediary, a lending transaction consists of an advance of funds to a borrower by a lender—in exchange, the lender receives from the borrower a promise to repay in the future the amount advanced plus a finance charge. At least on average, the amount of the finance charge must cover the lender's operating and non-operating expenses, including cost of funds.

Operating expenses in a lending transaction include costs of originating the loan, processing payments, and collection and bad debt expenses. All types of credit availability share the same basic activities, although the extent of specific activities depends on a variety of factors such as whether or not the credit is open end or closed end, the amount of credit and the term to maturity, whether or not collateral is taken, and the credit quality of the customers. For consumer lending, operating costs can be more substantial in relation to the size of the loan than on larger loans more typical of lending to business and governments, and research evidence suggests there are substantial economies of scale in operating costs associated with larger loan amounts.

Non-operating expenses of consumer lending include taxes, interest expense for share of the advance financed from borrowed funds, and a return on the owners' equity share of the advance. Although economic theory, as well as experience,

suggests that intermediation lowers the overall cost of the transfer of resources from ultimate savers to borrowers, it is obviously still true that the prices charged for loans cannot go below some minimum that must fully cover operating and non-operating costs of the transfer process, if the intermediary is to remain in business.

The importance of operating costs associated with loan origination, together with the high costs of losses on consumer lending, has kept the goal of reducing expenses constantly in the sight of the managers of lending intermediaries. It is, of course, possible to reduce both operating costs and losses to zero or close by not making any loans or making very few loans only to individuals who pose virtually no default risk. Not surprisingly, managers have not found this approach very useful, however, because it naturally also relegates profits to zero or very close. More useful over the years have been attempts to lessen operating costs and losses while keeping lending volume the same or increasing it. Managers have instituted a variety of approaches with this in mind, among them office automation, improved employee quality and training rather than more employees, and, especially, sophisticated statistical approaches to evaluating the risk of customers, an approach generally referred to as 'credit scoring'.

Default risk and the supply of credit

Because all credit transactions share the common feature of involving an intertemporal transaction in which the lender provides funds, there must be an expectation that future cash flows will be sufficient to replenish the funds and provide satisfactory capital return. For consumer lending, as with all lending, this means proper management of the possibility of default risk. Models of credit supply were originally developed to study the rationality of credit rationing, which was believed to be an important channel through which monetary policy transmitted to the economy. Credit rationing occurs when the price of credit is less than the equilibrium price. In such situations, the amount of credit demanded is more than the amount offered. Normally, an excess of demand over supply leads to a price increase. A model of credit supply was needed to explain why lenders would limit credit rather than raise the price of credit when monetary policy was tightened.

The default risk model of credit supply

The basic theoretical model of the supply of credit to an individual borrower (also known as the loan offer curve) starts with the quite reasonable assumption that the

borrower's final wealth, and thus his or her ability to repay, is limited and not known with certainty. Under these conditions, increasing the amount of credit extended increases the likelihood of default. Indeed, beyond a certain amount of credit, default may be virtually certain so that no offer to pay a higher interest rate would induce a lender to extend additional credit. As a consequence, the supply curve for an individual borrower that becomes completely inelastic or even backward bending at some rate of interest.

Several variants of this model have been developed. Perhaps the best-known variant is the model of Jaffee and Modigliani (1969). Other variants are by Hodgman (1960), Miller (1962), and Freimer and Gordon (1965). The default risk model was developed originally for commercial loans, but the critical feature of the model, the assumption that the borrower's ability to repay the loan is finite, clearly applies also to consumer loans.

Jaffee and Modigliani considered a lender's loan amount and interest rate decision when the borrower's wealth and hence ability to repay is a random variable. They demonstrated an optimal loan is one that equates the probability of default to the discounted difference between the loan interest rate and the opportunity rate. This result gives the loan-supply curve for an individual borrower its specific shape.

Normally, supply curves have a positive slope as higher prices elicit larger quantities supplied. In contrast, the loan-supply curve has several distinct features. For very small loan amounts, where repayment is virtually a certainty, the loan-supply curve is horizontal. That is, larger loan amounts do not entail higher interest rates. At some loan amount default risk becomes a consideration, however. Greater loan amounts entail greater default risk and hence higher interest rates. Thus, the supply curve has a positive slope. But, the maximum loan amount is limited because, as mentioned, the borrower's wealth is finite. A promise to pay a larger amount of interest is not credible. The borrower cannot possibly pay a larger amount even under the best of circumstances. Indeed, beyond the maximum loan amount higher interest rates entail smaller loan amounts. Consequently the loan offer bends backward.

The existence of a maximum loan amount is not credit rationing, though. Credit rationing requires consideration of demand and the determinants of the interest rate. Jaffee and Modigliani argued that credit rationing occurs because legal restrictions and considerations of good will and social mores prevent charging different rates to different customers. Instead, they suggested, lenders group customers in a small number of risk classes based on a few objective and verifiable criteria and charge a single rate to all customers in the class. Within these classes, borrowers whose individual rate is less than the common class rate will not be rationed, and borrowers whose individual rate more than the class rate will be rationed.

Asymmetric information and adverse selection

The basic default risk model of Jaffee and Modigliani does not consider the possibilities that lenders' information about borrowers may be imperfect or that the terms of a loan may affect borrowers' choices regarding risk or performance. Numerous models of credit markets with asymmetric information and adverse selection now exist. The models of Jaffee and Russell (1976) and Stiglitz and Weiss (1981) are among the best known and most influential.

Focusing on the latter, Stiglitz and Weiss demonstrate that adverse selection and moral hazard may cause credit rationing even in the absence of usury ceilings or community norms as suggested by Jaffee and Modigliani. Stiglitz and Weiss assume that the credit market is characterized by asymmetric information. Lenders observe only expected income but not the risk associated with income. In contrast, borrowers know both the expected value and risk.

Borrowers subsequently either realize income and repay the loan or default if income plus any assets pledged as collateral is less than interest and principal. The borrower keeps any surplus income above the amount of loan repayment but cannot lose more than the amount of assets pledged as collateral. This limit to the borrower's downside risk gives rise to the possibility of adverse selection and moral hazard.

Higher interest rates reduce the amount of income available after loan repayment, but lower-risk individuals are less likely to default and, therefore, would be less likely than higher-risk individuals to benefit from the limitation in downside risk or to receive large surpluses. Thus, rises in the interest rate would cause fewer lower-risk individuals to apply for loans. The resulting worsening of the risk distribution of applicants caused by rising interest rates is called adverse selection.

The lender cannot receive any more than the repayment amount of interest and principal if there is no default, and may lose up to the repayment amount less than the value of any assets pledged as collateral. Thus, greater risk due to adverse selection would increase the likelihood for the lender of receiving less than the contracted amount of principal and interest and, other things equal, would reduce the lender's profit per loan. Consequently, raising the lending interest rate might increase the lender's profit per loan for a while, but eventually a higher interest rate causes lower-risk borrowers to drop out of the market, worsening credit risk through adverse selection and thereby reducing profit overall. In other words, the lender's profit would not always rise with increases in the interest rate but may fall at some point because at a higher interest rate lower-risk borrowers do not apply for credit.

Credit rationing may then occur because lenders' supply of funds depends on lenders' profit, but borrowers' demand depends on the loan interest rate. Lenders will not increase interest rates to equilibrate supply and demand if doing so reduces their profits, which as described above may occur when higher interest rates cause lower-risk borrowers to leave the market.

Stiglitz and Weiss considered several extensions to their model. Among the extensions are the effects of the interest rate on borrowers' subsequent choices, differences among borrowers in attitudes toward risk, and collateral or equity requirements. For example, the interest rate may influence the subsequent behavior of a borrower. Specifically, a higher interest rate may induce a borrower to choose a riskier income prospect, a change in behavior called moral hazard. The reason is due fundamentally to the same difference in borrower and lender incentives that cause adverse selection. That is, the riskier income prospect becomes more attractive to the borrower as the interest rate rises but less profitable for the lender. The presence of moral hazard then provides another incentive for the lender to ration credit rather than raise the interest.

Credit scoring and current significance of models of default risk, asymmetric information, and adverse selection

The default risk model of loan supply for individual borrowers establishes the importance of default risk in determination of the interest rate, but the empirical significance of credit rationing is likely less today than it was when the model was developed. Interest rate ceilings have relaxed in many places and effectively have disappeared for some types of lenders. Special rate ceilings have been enacted in some places explicitly to allow small, short-term loans (payday loans, for example, in some American states and in other countries). Furthermore, information asymmetries between borrowers and lenders have been reduced as advances in technology made collection, storage, and analysis of comprehensive credit information possible and economical. Automated credit bureaus contain virtually complete credit use and payment performance information for nearly all credit users in some places, and the development of statistical credit bureau risk scores provide highly accurate predictions of future payment performance and can be available to any lender. Availability of comprehensive credit reports and credit bureau scores also facilitates risk-based pricing, which reduces the significance of rationing within broad risk classes as posited by Jaffee and Modigliani. Because of their usefulness for these purposes, public and private automated credit reporting agencies have developed in many countries (see Japelli and Pagano, 1999: Tables 1 and 2).

Through most of the twentieth century, lenders trying to assess a borrower's creditworthiness were guided by their own judgment and experience following industry folklore known as the five 'Cs' of lending: Character (of borrowers), their Capacity, Capital, and Collateral, and Conditions (largely economic conditions). Until fairly recently, consumer lending decisions were generally made individually by loan officers who exercised their individual judgment with each application. Loan officers gathered information from and about the applicant in each of the five

critical areas and applied lessons from their personal lending experience to decide whether an application should be approved.

As already indicated, more recently a number of factors have combined to push the consumer credit industry away from this 'judgmental' model of underwriting. Competitive pressures on lending institutions to process efficiently the rising tide of loan applications undermined the slow and typically labor-intensive judgmental credit evaluation process. The result was search for methods of automation, including statistical methodologies of credit evaluation that have come to be known as credit scoring, to take over the evaluation process from the older labor intensive methods. Statistical methods consisting of advanced forms of multiple regression and correlation analysis have become the norm in credit evaluation, along with extensive automated information sources, to feed necessary information to the statistical evaluation models. Besides lowering the costs of the credit evaluation process, the automated statistical approaches also have the advantage of consistent application across loan applicants in a way that simplifies management of an intermediation/lending enterprise. On this basis, they are unlike judgmental lending approaches that in the past were at least somewhat idiosyncratic to each individual loan officer. Statistical approaches do not invalidate the need for application of judgment to development of overall systems, however. Over-reliance on data about past relationships in mortgage lending and excess faith in relatively new forms of financial engineering contributed to credit market difficulties in 2007–9.

Regulation of consumer lending

Lending to individuals is as old as recorded human history and the interest of governments in regulating it is at least as old. For centuries this meant either absolute prohibition of consumer lending at interest or legal ceilings on the legal rate of interest, known respectively as usury laws and usury ceilings. In modern times, as interest and the economics of commercial activities have come to be better understood, usury restrictions have come to be less influential in many countries, although still important for consumer lending in some jurisdictions including France, Italy, Portugal, and Switzerland (see Masciandaro, 2001).

Of more significance in recent decades is growing governmental interest in ensuring transparency of consumer lending transactions and in regulating specific lending practices of financial institutions. An example of transparency regulation is the massive Truth in Lending Act of 1968 in the US, which covers most lending transactions involving consumers, including real estate transactions not generally

within the definition of 'consumer lending' employed here. Truth in Lending and similar statutes elsewhere have a number of distinct advantages over other forms of government regulation. Improving transparency refers to more than generic educational or credit-related educational materials in the lending area but rather to the governmental requirement of specific disclosures on details of lending transactions.

It seems that the importance of improved transparency and specific lending disclosures to government policy for protecting consumers in important transactions stems ultimately from at least three potential advantages over other methods of regulation (see Durkin and Elliehausen, 2010 for more extended discussion). First, information protections often are compatible with existing market forces already at work to protect consumers. Financial services providers with good reputations and favorable pricing have an incentive to make these facts known, and required disclosures can provide for common standards and terminology, such as the finance charge and annual percentage rate (APR) under Truth in Lending in the US. Mandatory standards can then enhance the power of existing market incentives to provide information, advancing consumers' learning process, lowering its cost, and making it more efficient. Under the circumstances, required disclosures in a standard format help highlight the performance of the best institutions and expose the inadequacies of the poorer ones.

Second, if what consumers really lack is information in particular areas, then it seems logical that consumer protection should focus upon providing what is missing rather than engaging in some other protection method. If consumers need information about pricing or terms of consumer credit contracts, for example, then it seems more reasonable to require disclosure of the information than to regulate prices or contract terms. Providing information rather than directly intervening does not require that the government knows, or presumes to know, the product-feature preferences of all consumers. With disclosures, consumers can decide for themselves what their own preferences are for the trade-off between price and product features, and success of the disclosure approach does not depend on consumers' preferences being the same.

Third, required disclosures may be relatively lower in cost—both in terms of market disruption and out-of-pocket government expenditures—than other approaches to consumer protection, although some observers may argue this point. Lower expected costs of this sort from disclosure schemes undoubtedly have been instrumental in encouraging their adoption in some places as political compromises between those demanding greater consumer protection and those arguing that more substantive market interference is too wrenching and costly or too harmful to the benefits that arise from a market-based system.

In contrast to transparency initiatives, an example of restrictions on lending practices to protect consumers is the Equal Credit Opportunity Act in the US, which prohibits taking into account in any credit-granting evaluation system (judgmental or statistical) certain individual characteristics such as sex, marital

status, race, or national origin. Although implementation of this law in the 1970s required lenders to engage in costly review all their evaluation practices and record-keeping procedures, few lenders today either disagree with the principles of Equal Credit Opportunity or find compliance especially difficult. Although there are occasional claims that correlations between measures of lending experience like lending turndowns or loan pricing with personal characteristics like race or national origin are indicative of illegal discrimination, most observers of lending markets believe there are different reasons, including differential income and assets that support lending (see Avery, Brevoort, and Canner, 2006).

There also are many additional governmental restrictions on lending practices in the US and elsewhere, including the US federal Fair Credit Reporting Act (1970, with major revision in 2003) that regulates activities of credit-reporting agencies, the Fair Debt Collection Practices Act (1978) that governs third-party debt collection agencies, the Federal Trade Commission's rule on Credit Practices (1984), and the consumer lending codes of the fifty individual states. The federal requirements are quite extensive in their coverage, as are many of the state laws.

Conclusion

Consumers have taken on debt obligations since antiquity, and lenders and the consumer lending marketplace have evolved over the centuries into very modern and sophisticated financial providers today. The essential elements of consumer lending and borrowing are well established in economics, and it is possible to study many features of the lending process and its governmental regulation with the tools of modern economic analysis. Because so many people today use the products of consumer lending, however, and because of ongoing political interest in the nature of markets and institutions in this area, widespread public discussion of both the benefits and costs of consumer lending seems likely to continue, despite their familiarity. For this reason, continued attention to this area from economic analysts should continue to prove beneficial.

References

Avery, R. B., Brevoort, K. P., and Canner, G. B. (2006). 'Higher-Priced Home Lending and the 2005 HMDA Data', *Federal Reserve Bulletin*, 92: A123–A166.

Benston, G. J. and Smith, C. W., Jr. (1976). 'A Transactions Cost Approach to the Theory of Financial Intermediation', *Journal of Finance*, 31: 215–31.

DAY, G. S. (1976). 'Assessing the Effects of Information Disclosure Requirements', *Journal of Marketing*, 40: 42–52.

—— and BRANDT, W. K. (1973). 'A Study of Consumer Credit Decisions: Implications for Present and Prospective Legislation', in National Commission on Consumer Finance, *Technical Studies of the National Commission on Consumer Finance*, Vol. I, 2. Washington, DC: US Government Printing Office.

—— —— (1974). 'Consumer Research and the Evaluation of Information Disclosure Requirements: The Case of Truth in Lending', *Journal of Consumer Research*, 1: 21–32.

DUNKELBERG, W. C. and STEPHENSON, J. (1974). 'Durable Goods Ownership and the Rate of Return', in National Commission on Consumer Finance, *Technical Studies of the National Commission on Consumer Finance*, Vol. VI. Washington, DC: US Government Printing Office.

DURKIN, T. A. and ELLIEHAUSEN, G. E. (2010). *Financial Economics of Information Disclosure: Applications to Truth in Lending*. Oxford and New York: Oxford University Press.

—— —— STATEN, M. E., and ZYWICKI, T. J. (2010). *Consumer Credit and the American Economy*. Oxford and New York: Oxford University Press.

ELLIEHAUSEN, G. E. and LAWRENCE, E. C. (2001). *Payday Advance Credit in America: An Analysis of Customer Demand*, Credit Research Center Monograph 35. Washington, DC: Georgetown University, McDonough School of Business.

FISHER, I. (1930). *The Theory of Interest*. New York: The Macmillan Company.

FREIMER, M. and GORDON, M. J. (1965). 'Why Bankers Ration Credit', *Quarterly Journal of Economics*, 79: 397–416.

GURLEY, J. G. and SHAW, E. S. (1960). *Money in a Theory of Finance*. Washington, DC: The Brookings Institution.

HIRSCHLEIFER, J. (1958). 'On the Theory of Optimal Investment Decision', *Journal of Political Economy*, 66: 329–52.

HODGMAN, D. R. (1960). 'Credit Risk and Credit Rationing', *Quarterly Journal of Economics*, 74: 258–78.

JAFFEE, D. M. and MODIGLIANI, F. (1969). 'A Theory and Test of Credit Rationing', *American Economic Review*, 59: 850–72.

—— and RUSSELL, T. (1976). 'Imperfect Information, Uncertainty, and Credit Rationing', *Quarterly Journal of Economics*, 90: 651–66.

JAPELLI, T. and PAGANO, M. (1999). 'Information Sharing, Lending, and Defaults: Cross-Country Evidence'. University of Salerno, Centre for Studies in Economics and Finance, Working Paper 22.

JUSTER, F. T. and SHAY, R. P. (1964). 'Consumer Sensitivity to Finance Rates: An Empirical and Analytical Investigation'. New York National Bureau of Economic Research, Occasional Paper 88.

KAHNEMAN, D. and TVERSKY, A. (1979). 'Prospect Theory: An Analysis of Decision under Risk', *Econometrica*, 47: 263–92.

MASCIANDARO, D. (2001). 'In Offense of Usury Laws: Microfoundations of Illegal Credit Contracts', *European Journal of Law & Economics*, 12: 193–215.

MILLER, M. H. (1962). 'Credit Risk and Credit Rationing: Further Comment', *Quarterly Journal of Economics*, 76: 480–8.

POAPST, J. V. and WATERS, W. R. (1964). 'Rates of Return on Consumer Durables', *Journal of Finance*, 19: 673–77.

Stigler, G. J. (1961). 'The Economics of Information', *Journal of Political Economy*, 69: 213–25.

Stiglitz, J. E. and Weiss, A. (1981). 'Credit Rationing in Markets with Imperfect Information', *American Economic Review*, 71: 393–410.

Tversky, A. and Kahneman, D. (1974). 'Judgment under Uncertainty: Heuristics and Biases', *Science*, 185: 1124–31.

CHAPTER 23

RESIDENTIAL MORTGAGES

ANDREAS LEHNERT[1]

Introduction

A mortgage is a legal arrangement in which a lender (the mortgagee) has some kind of legal claim on an underlying piece of real property held by a borrower (the mortgagor). Interested parties other than the mortgagee may have claims to the property; such claims (as well as the mortgagor's) are often referred to as liens. A separate legal document, sometimes known as the note, sets out the terms and conditions under which the borrower will satisfy the lender and discharge the debt. In practice, most research tends to conflate the mortgage, lien, and note into a single entity known simply as a mortgage. Except where required for technical reasons, we will follow this convention.

The precise legal arrangements vary across jurisdictions, depending both on the underlying legal tradition of the jurisdiction (e.g., common vs. civil law) and on specific statutes enacted by the jurisdiction. For example, even jurisdictions with similar legal traditions, such as individual states in the US, differ markedly in their statutory treatment of 'foreclosure', the legal process by which lenders seize property from borrowers in default of the terms of their notes.

[1] The views expressed here are solely those of the author and do not reflect the view of the Federal Reserve Board or its staff. I thank Jay Brinkmann, Michael Gibson, and Michael Palumbo for comments and concrete suggestions, and Christina Pinkston for assistance with the First American LoanPerformance data. Any errors are my fault.

Mortgage banking traditionally consists of three related businesses: origination, funding, and servicing. Origination is the extension of new credit to borrowers; funding refers to the mix of debt, equity, and market-based instruments used to finance portfolios of mortgages; and servicing is the day-to-day business of managing payments from borrowers. These three businesses can be separated; indeed, many banks prefer to use their retail presence to concentrate on originating mortgages, then selling them to investors on the secondary market for loans rather than funding them internally. Mortgage servicing, because it features large fixed costs, can exhibit significant scale economies. Thus, many of the functions of mortgage banking can be done outside of traditional depository institutions.

The US mortgage market was the initial locus of the turmoil that hit financial markets beginning in August 2007. As of this writing, broader financial markets worldwide remained distressed and US mortgage markets had yet to find a new equilibrium. Nonetheless, researchers have had enough time to grapple with the issues and make major contributions to understanding the causes of the problems in US mortgage markets, and the important implications of these causes for mortgage markets worldwide.

This chapter aims to provide the reader with an introduction to residential mortgages and their relationship to banking. It will describe the common features of residential mortgages (second section); factors influencing households' behavior (third section); the market for mortgages as a financial asset (fourth section); bank capital regulation and its treatment of residential mortgages (fifth section). Finally, because standards and practices in mortgage markets worldwide have been affected by the turmoil that hit US mortgage markets in 2007, we describe the episode in some detail in the sixth section.

Residential mortgage features

Mortgages are a specialized asset class, with their own jargon and specific concerns. We describe how to compute payments on mortgages, the terminology and key concepts in the decision to extend credit to a borrower, the issues surrounding the seizure of collateral, and, finally, the little-studied back-office work of processing payments and contacting borrowers.

Mortgage payments

In principle, mortgages describe a series of payments to be made by the borrower; with the successful completion of these payments the borrower usually holds clear title to the property (hence the term 'mortgage', or 'dead pledge'). The process of

decreasing the principal balance owed on a mortgage is known as 'amortization'. In addition, mortgages carry an interest rate that can vary over time, usually in line with a published index. Mortgages can differ in how interest rates are determined and in their amortization schedules.

A mortgage will carry a 'note rate', r_t, the interest rate on the loan, usually expressed as an annual percent. The actual rate applied to the loan for a month's borrowing is, by convention, the annual rate divided by twelve. The mortgage will have a remaining maturity, T_t. Given a payment at the end of the month, x_t, and an unpaid principal balance at the end of the previous period, P_{t-1}, the unpaid principal balance at the end of period t is $(1 + r_t)P_{t-1} - x_t$. If the payment is designed to amortize the loan over the remaining maturity of T_t periods, it must be the case that $P_{T_t} = 0$. Substituting, the payment x_t must satisfy:

$$x_t = r_t P_{t-1} \frac{(1 + r_t)^{T_t}}{(1 + r_t)^{T_t} - 1}.$$

This formula assumes that the note rate r_t remains constant over time. As the note rate fluctuates, the payment will fluctuate, both to pay for the higher cost of borrowing a month's worth of the principal, but also to amortize the loan over the remaining maturity. Alternatively, the maturity could adjust to offset changes in interest rates to keep the payment constant, as in 'variable maturity mortgages'.

Mortgages may offer varying amounts of interest rate protection. A 'fixed rate' mortgage carries a constant note rate. An 'adjustable rate' mortgage (known as an ARM) carries a note rate that is usually computed as a fixed margin over a published index; ARMs vary in the frequency of adjustment. For example, the note rate on an ARM could adjust every six months and be computed as a 2 percent margin over the average value of the six-month LIBOR prevailing in the month prior to adjustment. 'Hybrid' mortgages carry a fixed rate for an extended period before converting to an ARM. The term 'variable rate' mortgage is sometimes used to encompass traditional ARMs and hybrids.

Mortgages may have provisions that allow the principal balance to remain constant ('interest-only payments' where $x_t = r_t P_t$) or even grow ('negative-amortization payments' where $x_t < r_t P_t$). Some mortgages allow borrowers to choose the amount of amortization from month to month. All of these loans have a cap on the principal balance; when the mortgage hits the cap it is said to 'recast'. Often, mortgage payments then rise to the so-called fully amortizing rate, sufficient to pay off the total principal balance over the remaining life of the loan.

Variable rate mortgages may have caps on how much the scheduled payment can rise at a time—however, these caps may be coupled with a provision that the difference between the fully indexed payment $r_t P_t$ and the capped payment $\bar{r}_t P_t$ be added to the principal.

As an alterative to non-amortizing loans, some loans amortize on a longer schedule than their contractual maturity. For example, a loan's principal payments

may be computed as if the principal were to be repaid over forty years, while the loan in fact is only set to last thirty years. The final payment of such a mortgage, consisting of all unpaid principal, is called a 'balloon payment'.

The maturity of the loan, T_p can vary over time. However, the traditional or typical maturity varies across countries as well. In the US, the standard maturity is thirty years, in many other countries it is shorter, while in some countries, notably Japan, it can be longer.

Mortgages can also have balloon payments due in a relative short amount of time—for example, five years after origination. Typically, borrowers do not actually make the balloon payment—instead, they take out a new loan to pay the balloon payment.

As a concrete example of the payment schedule and evolution of principal on a variable rate loan with negative amortization features, consider a hypothetical version of a popular mortgage made to borrowers with weaker credit histories in the US during the credit boom of 2004–7. (We discuss subprime mortgages in greater detail in the sixth section, below.) This mortgage has a variable interest rate indexed to the six-month LIBOR plus 6 percentage points, although the contract rate is fixed during its first two years at, say, 7 percent; the payment adjusts every six months thereafter so that the contract rate equals the average LIBOR over the month before adjustment. For simplicity, assume that the 6-month LIBOR is constant at 5 percent (roughly its average value over 2006 and 2007). Assume that the contract rate cannot change more than 2 percentage points at a time, although if the contract rate is below the fully indexed rate after the first rate reset the difference is rolled into the principal. Note that the initial rate (7 percent) is heavily discounted relative to the fully indexed rate of 11 percent; the initial rate is known as a 'teaser rate'. Finally, assume that the payments over the first two years are interest only.

Table 23.1 gives the contract rate, the fully indexed rate, the monthly payment, and the principal outstanding at several points in the loan's life. As shown in the table, the monthly payment jumps $466, or 40 percent, at the first rate reset date. Assuming that the initial mortgage payment accounted for 30 percent of the borrower's post-tax income, the fully indexed payment of $1,940.15 will account for 50 percent of his income. Further, because of the interest-only and negative amortization provisions, even by month 42, the principal outstanding is still higher than at the origination date of the mortgage. Such mortgages are designed to carry affordable payments for only a limited time before strongly encouraging the borrower to refinance.

The credit extension decision

Lenders are said to 'underwrite' a loan when they decide whether or not to extend credit (or 'originate' a loan) to a potential borrower, and, if so, on what terms.

Table 23.1. Payments on a hypothetical mortgage

Period	Contract rate (%)	Fully indexed rate (%)	Monthly payment	Principal outstanding (end of period, dollars)
Initial				200,000.00
Months 1–24	7.00	11.00	1,166.67	200,000.00
Months 25–30	9.00	11.00	1,632.60	201,232.34
Months 31–6	11.00	11.00	1,940.15	200,645.95
Months 37–42	11.00	11.00	1,940.15	200,236.80

Note: Table gives terms on a hypothetical mortgage. See text for details.

There are four main variables that commonly enter the underwriting decision, as well as a host of other considerations.

First, because mortgages are collateralized debt, the lender must value the property. Ideally, properties would be sold at auction, and the lender would value the property using the second-highest bid. In the event the borrower defaults, the lender could seize the property and sell it to the second-highest bidder. Yet, in most countries (with some notable exceptions), houses are not usually sold at auction. Further, borrowers may want to get a loan to refinance an existing mortgage rather than to purchase a home. To value the property, then, lenders must rely on an independent valuation, known as an 'appraisal'. Evidence suggests that, as one might expect, appraisers are under pressure to report a value high enough to enable the deal to go through (see LaCour-Little and Malpezzi, 2003). Indeed, Ben-David (2008) suggests that, in some instances, appraisers systemically over-report home values as part of a broader scheme to defraud lenders.

Appraisals can enter house price indexes as if they were true arm's-length transactions. Several major house price indexes now routinely strip out appraisals from their base data when constructing their indexes. However, Leventis (2006) points out that, even though flawed, appraisals may give some information on price movements and hence excluding them needlessly increases the standard error of the estimate. He describes a procedure for removing this 'appraisal bias'—the tendency of appraisals to overvalue properties—from house price indexes. Leventis argues, based on findings in the literature, that appraisals are likelier to be inflated for refinancings where the borrower liquidates equity—that is, cash-out refinancings, relative to refinancings where the borrower merely wants to take advantage of lower interest rates (so-called 'rate/term' refinancing). Leventis estimates a model in which reported house prices are inflated by a constant proportion each period depending on the type of refinancing. He finds that his improved price index has a lower variance and the same mean as a 'purchase-only' index.

Second, lenders must decide how much of an equity cushion to require. This is usually measured as the 'loan-to-price' or 'loan-to-value' (LTV) ratio, defined as the mortgage principal divided by the property's value. Given a low enough LTV

and foreclosure laws that permit the timely seizure of collateral, mortgage lending can be effectively riskless. For example, with an LTV at origination of 75 percent, even if house prices declined 20 percent and the borrower defaulted, the lender would still be unlikely to take a loss on the loan.

However, potential homebuyers often find the large down payments required to achieve low LTVs onerous (see Haurin, Herbert, and Rosenthal, 2007, among others). In principle, lenders should be, and have been, willing to accept the increased risk associated with higher LTVs in exchange for higher note rates (Edelberg, 2006). For institutional reasons, in the US, borrowers seeking an LTV above 80 percent often resort to either 'mortgage insurance', in which a third party guarantees repayment of principal to the lender in exchange for monthly insurance premiums paid by the borrower, or to 'piggyback mortgages' or 'junior liens', in which the borrower makes a downpayment of less than 20 percent, but splits the mortgage into a loan with an LTV of 80 percent and a second loan for the remaining amount. (This is sometimes also known, rather confusingly, as 'borrowing the downpayment'.) Operationally, investors and other market participants can find it difficult to determine whether a given mortgage has an associated junior lien, making it difficult for them to determine the total debt on a property.

Third, lenders consider the borrower's ability to make the scheduled mortgage payments. Usually, lenders compute various payment-to-income ratios (also known as debt-to-income or DTI ratios) and compare them to thresholds determined by underwriting guidelines. A commonly used ratio, known as the 'back-end ratio', compares monthly payments associated with all debts on a household's balance sheet (including property taxes and insurance, credit card, auto loans, and so on) to the household's post-tax income. However, measuring and verifying a borrower's income is not straightforward as some lenders consider expected future income such as unrealized bonuses while some borrowers prefer not to document certain income sources. This difficulty led to the rise of lending with incomplete verification of income and assets, known as 'low-doc' or 'no-doc' loans. In effect, the lender relies in part on the borrower's own estimation of his ability to make mortgage payments.

Fourth, lenders often consider a borrower's history of making debt payments on time. In the US, a popular summary measure of borrower credit quality developed by Fair Isaac Company, known as the 'FICO score',[2] has become a quick rule of thumb for determining whether a borrower is prime quality or below prime—that is, 'subprime'. (However, note that most underwriting engines use more than just the 'FICO score' to determine a borrower's credit risk.) Barakova, et al. (2003), using data on US households, find that wealth and

[2] FICO is a registered trademark of Fair, Isaac Corporation.

income apparently decreased in importance as barriers to homeownership over the 1990s, while credit scores increased in importance. This result underscores lenders' increased willingness to accept the risks associated with high LTV loans and reliance on credit histories.

Finally, lenders consider a variety of other factors when underwriting loans. They impose tighter restrictions on loans backed by non-owner occupied (i.e. investment) properties, loans with extremely high balances, and loans with non-standard amortization features.

Seizure of collateral

Mortgage lending is unlike other consumer lending because it is secured; however, if lenders cannot easily seize the underlying collateral, the difference between a mortgage and, say, a credit card loan starts to wane. However, policymakers may also want to design legal systems that delay collateral seizure in order to provide homeowners with some crude insurance and bargaining power to protect them against shocks to income and house prices.

Borrowers are typically not considered seriously delinquent, and hence at risk for having their home seized, until they have missed three consecutive mortgage payments. What happens at this point varies by state within the US, and, even more, across countries. Laws and regulations can be designed to hinder or ease the transfer of ownership from the borrower to the lender, a process known as 'foreclosure'. In the US, a lender typically starts the foreclosure process with a 'notice of default'. Using cross-state variation in the length of time between the foreclosure start and the final transfer of property, Pence (2006) shows that more defaulter-friendly foreclosure laws lead lenders to demand higher downpayments, and thus restrict credit access, because more defaulter-friendly laws delay repossession of the property and impose additional costs on the lender. Pence also cites a 1938 study that found higher foreclosure costs in states with defaulter-friendly foreclosure laws by the Home Owners' Loan Corporation (the original New Deal entity that became Fannie Mae). Clauretie and Herzog (1990) use data from mortgage insurers to estimate losses from more defaulter-friendly foreclosure laws. They find that lenders lose more money on loans that default in states with laws that hinder the foreclosure process.

Pence (2006) cites estimates of the financial losses to mortgage holders following a foreclosure ranging between 30 and 60 percent of the unpaid principal balance on the loan. Not only do lenders have to incur legal costs and expenses associated with maintaining a property, they have to face a sometimes substantial delay before they can repossess the property and resell it. An extensive literature documents these foreclosure costs (see, among others, Capone, 1996; Clauretie, 1989; and Ciochetti, 1997).

Mortgage servicing

Mortgage servicing is the business of computing scheduled payments on mortgages, collecting these payments from borrowers and transmitting the proceeds to the mortgage owners or holders. In addition, servicers monitor borrowers' credit records for events that can threaten the value of the collateral, such as failure to pay property insurance, personal bankruptcy filing, or liens filed by other creditors, such as 'mechanics liens' or liens filed by homeowners' associations. Finally, servicers are usually responsible for handling delinquent borrowers, whether by foreclosing on the property or some other course of action.

Mortgage servicers are usually paid by allowing them to retain a portion of the borrower's monthly payment—for example, a servicer's fee might be quoted as '25 basis points', indicating that, on a mortgage with a note rate of 6.75 percent, the servicer transmits 6.50 percent to the mortgage holder and retains 0.25 percent for itself. However, servicers are typically required to advance scheduled principal and interest payments to the mortgage holders even if the borrower has stopped paying. The servicer can recoup the value of these advances as well as out-of-pocket expenses incurred during a foreclosure proceeding (see Cordell, et al. 2008 for more information on the incentives faced by mortgage servicers).

Mortgage servicers charge more to handle loans that require them to contact borrowers more often. For example, borrowers with weaker credit histories can routinely miss one payment per year. While such borrowers are in technical violation of the terms of their mortgage, mortgage servicers typically respond by reminding the borrower of his missed obligation, and most of these borrowers subsequently make good on the missed payment. However, this kind of repeated contact is expensive, and explains part of the larger fee charged to service such loans.

When a borrower misses several payments in a row, servicers must decide whether to pursue a foreclosure to allow the borrower an opportunity to make good the missed payments. Stegman, et al. (2007) argue that such forbearance, and even more aggressive policies known as 'modifications', that involve permanent changes to the terms of a mortgage to decrease monthly payments, increase the net present value of a mortgage. Rather than seizing a property whose value has probably fallen, Stegman, et al. point out that by changing the terms of the mortgage to better suit a borrower's circumstances (which had perhaps changed since origination), the servicer is likelier to realize continued timely payments, albeit smaller than before, and avoid the expense of seizing the property. Eggert (2007) points out that while loan modifications may indeed make broad economic sense, the incentives of servicers are far from clear in this case and that the ultimate owners of the mortgage may disagree on how best to proceed.

More broadly, mortgage servicers find themselves at the center of the current subprime crisis in the US. With ownership of the mortgages often dispersed via securitization, servicers, rather the actual mortgage owners, can be left to decide

how much, if any, forbearance to offer borrowers. While they should act to maximize the present discounted value of the mortgage on behalf of the mortgage's ultimate owners, financial models and industry practice can give relatively little guidance in the current episode.

Household decision-making

In this section we analyze household decision-making regarding mortgages. First, we consider the various choices a borrower faces before taking on a mortgage: how much to borrow, how much interest rate risk to accept, what kind of amortization structure to use, and whether to accept a prepayment penalty. Second, we consider the various choices a borrower faces after taking on a mortgages: whether to refinance the loan or whether to default on the loan.

Choice of mortgage contract features

Households face a long menu of options when taking out mortgages, with the items on this menu differing radically across countries. Research has focused on choices made by consumers in the US because of the relatively large number of options available and because of the presence of several useful data sources with which to study decisions. Despite setbacks, the trend across countries appears to be toward ever-greater choice, suggesting that the US experience can be a useful lesson elsewhere. A comprehensive list of cross-country differences in mortgage systems, as well as differing household choices among options, is available from the Bank for International Settlements (2006).

The first choice a household faces is how much mortgage debt to take on. As shown in Figure 23.1, this choice is largely governed by house prices. However, unobserved common factors, such as expected future house price growth, probably influence both households' willingness to pay for a house *and* their willingness to borrow to finance the purchase. Thus, causation is likely to run in both directions.

Household borrowing may be constrained by caps on LTVs, as we discussed in the previous section. In principle, borrowers should face an interest rate trade-off between increased leverage and higher mortgage rates. Edelberg (2006) shows that such risk-based pricing increased in the 1990s in the US. Bucks, Kennickell, and Moore (2006) document that household leverage increased in recent years. Stein (1995) and Lamont and Stein (1999) argue that leverage constraints are binding for the marginal homebuyer, and that when buyers can use greater leverage, equilibrium house prices

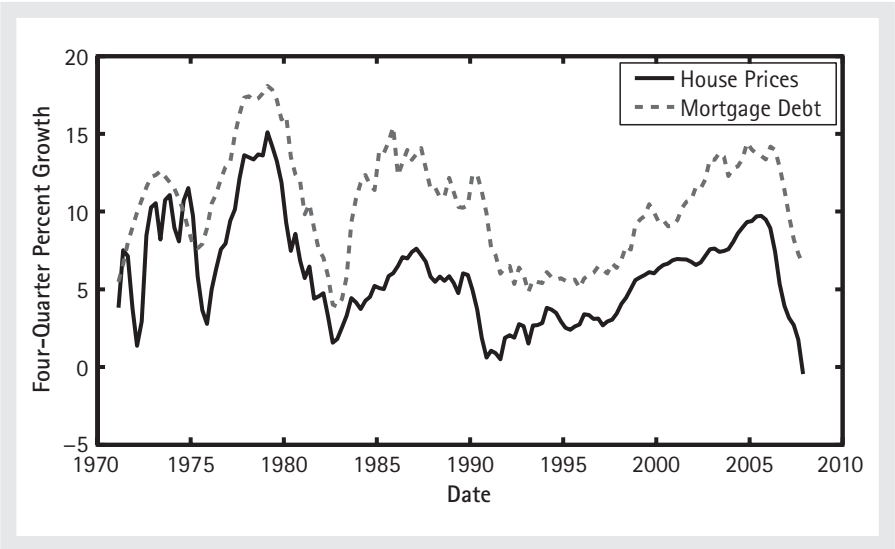

Fig. 23.1. US house price and mortgage debt growth

Note: Figure 23.1 shows the four-quarter changes in US residential mortgage debt outstanding (the dashed line) and house prices. The house price index used here is a composite of Freddie Mac's CMHPI, OFHEO's all-transactions index, and OFHEO's purchase-only index. Mortgage debt is taken from the flow of funds accounts of the US.

move more in response to a shock to fundamentals, such as unemployment. Thus, the mortgage market acts as an amplification mechanism.

In addition, some national tax codes, including that in the US, permit the deduction of mortgage interest, which may encourage households to use more debt than is optimal. (Technically, a mortgage interest deduction should not distort households' debt choice if it is coupled with a tax on the imputed income derived from the service flow of housing financed by the debt.) Amromin, Huang, and Sialm (2007) find that US households forgo a potentially lucrative tax arbitrage between tax-exempt debt (mortgages) and savings (401(k) plans). They ascribe this in part to debt-aversion, as described by Graham (2000). This suggests that, at the margin at least, the tax code may not exert as great an effect on household mortgage choice as one might expect.

The second choice faced by households concerns the level of interest rate risk they are willing to accept. National mortgage systems differ markedly in the amount of interest rate protection available. Mortgages with contract rates that are fixed for five or more years are widely available in Canada, Germany, France, Belgium, Switzerland, the Netherlands, and the US, although these mortgages may carry prepayment penalties and other features. The UK has attempted to encourage fixed rate borrowing, although with mixed success (see Miles, 2004).

To illustrate this choice, consider a hypothetical homebuyer who wishes to borrow $100,000 in January 1998 using either a hypothetical one-year Treasury ARM or a standard fixed rate mortgage: Here, the hypothetical ARM resets to the

1-year Treasury rate plus, by assumption, 2 percent every year; see Stanton and Wallace (1999) for more details of common features of ARMs. The homebuyer will remain in the home for seven years. Figure 23.2 shows the contract interest rates on his mortgage over the next seven years; the rate on the ARM adjusts essentially in line with US monetary policy over the period. The rate on the FRM is constant except when the borrower refinances during the extremely low rate environment of early 2003. (Here, refinancing is assumed to cost 1 percent of the loan balance and borrowers refinance when current rates fall 1.5 percent below their current contract rate; we discuss refinancing in greater detail below.) Assuming the borrower has a constant 3 percent discount rate, the present discounted value (PDV) of his interest and refinancing expenses would have totaled about $41,000 had he taken the fixed rate loan and $35,000 had he taken the adjustable rate loan.

Figure 23.3 shows the PDV of interest payments for the same hypothetical fixed and adjustable rate loans originated each month from January 1972 to June 2000. As shown, the relative advantage of one mortgage type over the other is dwarfed by the low-frequency trend in US interest rates (this is also visible in Figure 23.4). Further, this is an *ex post* exercise and does not reflect the *ex ante* conditions faced by the household when making the decision between fixed and adjustable rate mortgages. Finally, no consideration is made here for the very important issue of household risk aversion. That said, homebuyers in the 1970s would, generally speaking, have been much better off using fixed rate mortgages, while home buyers in the 1990s would have been slightly better off using adjustable rate mortgages (although this difference is relatively small, sensitive to assumptions, and, for the

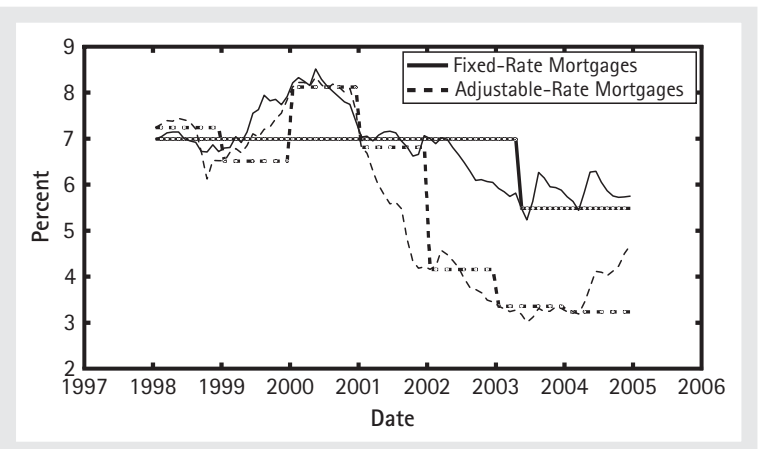

Fig. 23.2. Example interest rates on hypothetical adjustable and fixed rate mortgages

Note: Figure 23.2 shows the trajectories of rates paid by hypothetical borrowers getting a fixed rate (solid lines) and an adjustable rate (dashed lines) mortgage. Thick lines give the contract rates paid; thin lines give the prevailing mortgage rates.

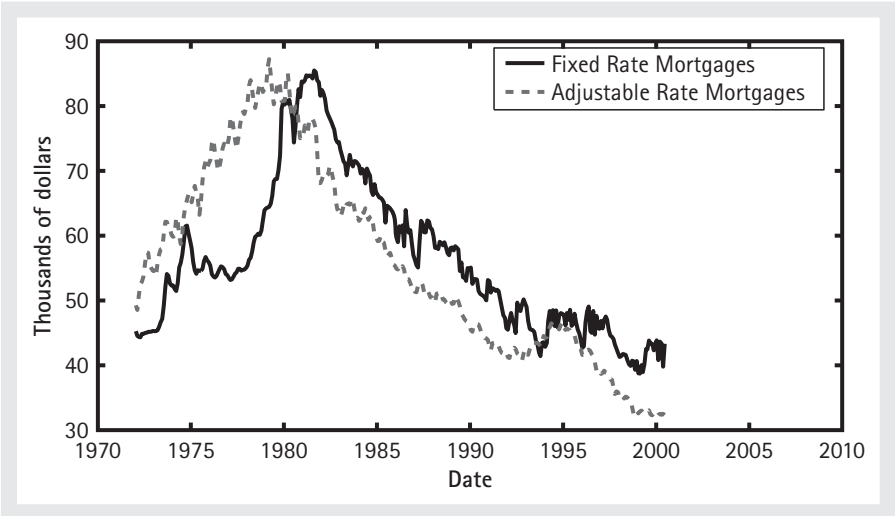

Fig. 23.3. Present discounted value of interest and refinancing expenses for hypothetical fixed and adjustable rate mortgages

Note: Figure 23.3 shows the present discounted value of interest payments for a hypothetical homebuyer financing $100,000 using either a fixed rate (solid line) or adjustable rate (dashed line) mortgage. See text for details on the hypothetical mortgages.

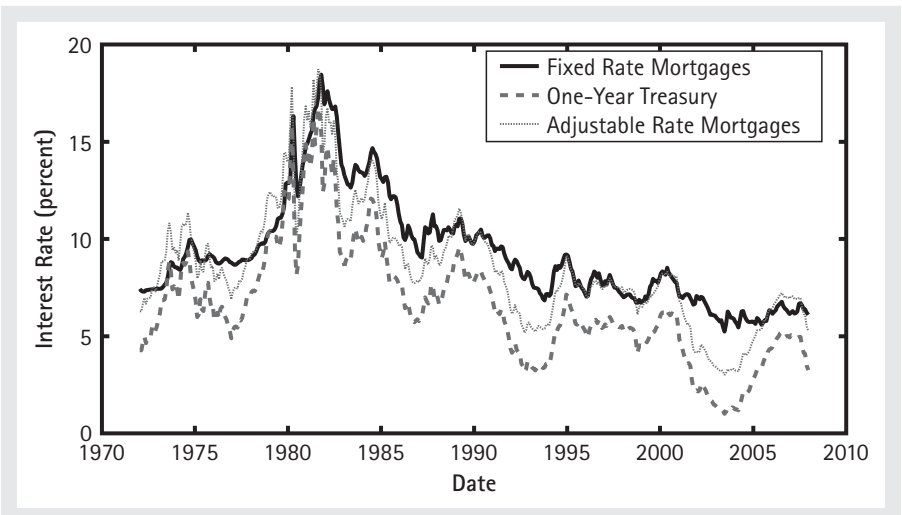

Fig. 23.4. Mortgage rates in the US

Note: Figure 23.4 shows various nominal interest rates. The solid line shows commitment rates on prime conforming 30-year fixed rate mortgages taken from Freddie Mac's PMMS, the dashed line shows the one-year Treasury rate, and the dotted line shows the hypothetical rate on a one-year Treasury ARM with a 200 basis point margin.

reasons outlined earlier, do not reflect all the considerations faced by households making decisions). Earlier studies (see Shilling, Dhillon, and Sirmans, 1987; Brueckner and Follain, 1988; and Brueckner, 1993) focused on the relative risks of the two mortgage types, while Campbell and Cocco (2003) find that the real cost of nominal fixed rate mortgages is extremely sensitive to realized inflation. They argue that inflation-indexed fixed rate mortgages offer the benefits of stable payments without requiring an inflation premium.

The third choice faced by households concerns the mortgages' amortization schedule. Amortization can be seen as a form of portfolio shuffling; households are building their home equity at the expense of other forms of saving. In this view, households pay down their mortgage taking into account the risk–return profiles of other investment opportunities; see Fu, LaCour-Little, and Vandell (1997). Alternatively, mortgages with deferred amortization schedules are marketed as 'affordability products', emphasizing their lower payments relative to fully amortizing mortgages. LaCour-Little and Yang (2008) find that borrowers with greater expected income growth, or those purchasing homes in areas with rapidly appreciating prices, are likelier to select loans with deferred amortization, suggesting that borrowers do indeed value the decreased payment burden associated with deferred amortization. LaCour-Little and Yang also argue that, over the course of many years, the lack of amortization on a mortgage leaves the borrower with less equity on the property and hence is more likely to default. Piskorski and Tchistyi (2007) argue that, within the context of a mechanism design problem, the optimal mortgage contract can be characterized as a loan with amortization under the control of the borrower, including negative amortization options, which are used to offset shocks to household income.

The final choice a household makes is whether to enter into a mortgage that carries some form of a prepayment penalty or other feature that makes refinancing less desirable. As we discuss below, lenders funding mortgages with non-callable liabilities (such as deposits or standard loans) must carefully manage prepayment risk when holding leveraged portfolios of fixed rate mortgage assets. In the absence of a well-developed market for fixed income derivatives, such lenders will require extra compensation to hold fixed rate mortgages without prepayment penalties. In countries such as the US with relatively liquid fixed income markets, prepayment penalties can spare the lender the cost of hedging. In practice, however, prepayment penalties are usually not even offered to prime borrowers, and there is considerable debate as to whether prepayment penalties offer any net benefits to borrowers. Elliehausen, Staten, and Steinbuks (2008) find that loans with prepayment penalties carry lower rates and fees than equivalent loans without them. However, if enough borrowers actually refinance, and hence must pay the fee, this benefit could be offset. Borrowers have the option of paying 'points'; that is, buying down the contract rate on their loan by paying an upfront fee. Brueckner (1994) argues that points are an effective signal of a borrower's unwillingness to refinance.

Given the various dimensions discussed in this section along which a mortgage contract can differ, it is reasonable to wonder whether the typical borrower understands the terms of his mortgage. Bucks and Pence (2006) find that most borrowers understand the broad terms of their mortgages. However, some borrowers with adjustable rate mortgages underestimated the size of the caps on potential changes in their note rates.

Household decisions to refinance or default

Borrowers who already have a mortgage can decide to 'refinance' or 'remortgage'; that is, take out a new mortgage under different terms than the existing loan and use the proceeds to pay off the existing loan. Borrowers may also decide, or be forced by a negative shock, to 'default'; that is, to violate the agreed-upon terms of the note by failing to make adequate or timely payments of principal or interest.

Standard models treat the refinance and default decisions as options embedded in the mortgage. In this view, refinancing corresponds to an option to call the mortgage at par (assuming no prepayment penalties), while default corresponds to an option to put the mortgage back to the lender at the value of the house. From period to period, mortgages terminate in a default, terminate in a refinancing, or continue to the next period. Thus, the standard empirical model is a modified version of the duration model known as a 'competing hazards' model. Deng, Quigley, and Van Order (2000) added unobserved borrower heterogeneity; their model, and its subsequent adumbrations, is the workhorse model used in mortgage-level analysis. Gerardi, Shapiro, and Willen (2008) constructed a dataset of homeownership experiences, rather than mortgages; in their paper, a homeowner goes through multiple mortgages, and the competing hazards are voluntary sale and loss of the home involuntarily through foreclosure.

Borrowers refinance for other reasons besides the desire to lower their note rate (i.e., at times when new mortgages carry substantially lower rates than existing mortgages); they may wish to increase their mortgage principal. Such 'cash out' refinancings often are the cheapest form of financing available to borrowers, even if the new mortgage carries a rate higher than the mortgage it is replacing. Hurst and Stafford (2004) document that households with few liquid assets are more likely to undertake a cash-out refinancing following an employment shock. Such refinancings allow households to tap their accumulated home equity and thus smooth consumption. Figure 23.5 shows a measure of total refinancing activity in the US as well as the fraction of refinancers taking cash out. Because most refinancing occurs when rates are relatively low and most refinancers in such time are simply seeking to lower their contracts, the two series exhibit an inverse correlation. When rates are relatively high, only borrowers seeking to take cash out will refinance. The Federal Reserve Board has conducted a series of surveys of borrowers undertaking a

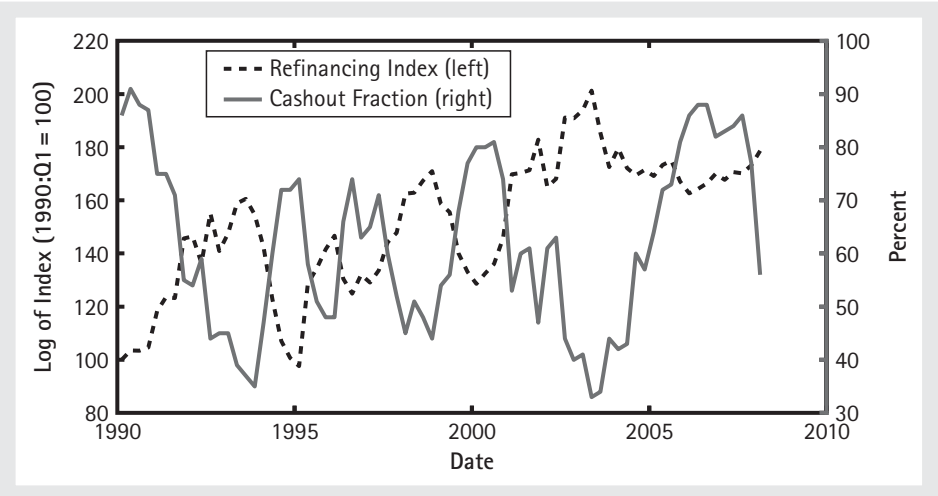

Fig. 23.5. Refinancing volume and percent taking cash out

Note: Figure 23.5 shows the log of the Mortgage Bankers Association index of refinancing applications (left scale), and the percent of refinancers increasing the loan balance five percent or more from Freddie Mac's quarterly survey of refinancers (right scale).

cash-out refinancing. As reported in Canner, Dynan, and Passmore (2002), the most popular reported uses of funds raised were, first, to repay other debts (presumably carrying higher interest rates), second, to finance a home improvement project, third, to purchase consumer items, and fourth, to invest the proceeds. Weighted by dollars raised, home improvement projects became the most popular use of funds. It is interesting to note that French mortgage law typically does not allow borrowers to increase their mortgage balance after the purchase of a property (Bank for International Settlements, 2006).

Mortgage defaults appear to be primarily driven by house prices or, more precisely, the borrower's equity in the home. Indeed, if a borrower truly has some equity in the home, default is extremely unlikely because the borrower would prefer to sell the home and realize the equity rather than have it repossessed by the lender. However, there is some debate over whether borrower behavior is best characterized as the exercise of the put option in the mortgage, with borrowers walking away from the property once its price dropped below a critical threshold, or whether defaults are ultimately driven by cash flow considerations. The former view is sometimes characterized as 'ruthless default' while the latter view is known as the 'double trigger' or 'borrower solvency' theory of default. In the double trigger view, negative equity is a necessary but not sufficient condition for default; the underlying pace of adverse life events such as job loss, uninsured medical expenses, divorce, and so on show through to default rates only when borrowers do not have an equity cushion to rely on. The two views can be reconciled if one assumes that there are substantial transactions costs from default, pushing the

optimal default trigger price extremely low. Indeed, Foote, Gerardi, and Willen (2008) document that a large number of homeowners in Massachusetts endured several years of negative equity in the early 1990s without defaulting.

Quigley and Van Order (1995) conduct one of the first empirical tests of the pure 'ruthless default' model and find that observed defaults imply fairly high transaction costs. Ambrose, Capone, and Deng (2001) estimate the trigger values of a put option and find that, again, a frictionless model of option exercise does a poor job of explaining the data, suggesting that borrowers weigh more than the narrow financial benefit of defaulting. However, the authors also find evidence that forward-looking borrowers consider the state of the housing market and thus emphasize the importance of expectations in the borrower decision.

Residential mortgages as a financial asset

In the same way that a mortgage is a household liability as discussed above, it is an asset held by the lender. Like any other asset, mortgages can be sold individually or used to back larger securities. Indeed, lenders, investors, and other financial market participants value mortgages using the standard tools for valuing any asset that generates a series of scheduled and largely fixed payments. A description of the fixed income valuation toolkit is beyond the scope of this chapter. Nonetheless, we discuss approaches to valuing the options embedded in residential mortgages and the inherent challenges of valuing individual mortgages. Because whole loans are relatively illiquid and hard-to-value financial instruments, investors prefer to buy securities backed by thousands of individual mortgages, often with extra protections to make valuation easier. Again, there is an enormous practical and scientific literature on mortgage-backed securities, which is outside the scope of this chapter. We instead describe some of the important organizing principles and institutional features of mortgage securitization.

Valuing a mortgage

The primary concern in valuing the cash stream generated by a mortgage is determining the probability that the mortgage will either default or prepay conditional on realizations of the appropriate aggregate variables thought to determine state prices (see Duffie, 1992). If cash flow in future states is valued based on just the prevailing risk-free rate, the key issue is how the mortgage cash flow varies with

changes in interest rates. This in turn requires modeling the behavior of the underlying borrower in reaction to aggregate variables. (State prices might also depend on house prices or other aggregate variables that can affect borrower behavior.)

The most common borrower choice is to refinance an existing mortgage when rates drop. Without refinancing, and ignoring default risk for a moment, fixed rate mortgages would tend to rise (fall) in value as spot rates fell (rose). If borrowers refinance when rates fall, however, the cash flow from the mortgage may terminate in precisely those states of the world when the mortgage is most valuable.

Figure 23.6 shows a scatterplot of the Mortgage Bankers Association's refinancing index relative to the 'refinancing incentive', defined as the difference, in basis points, between the currently prevailing rate on new thirty-year fixed rate prime mortgages and the average rate on all outstanding mortgages. When the refinancing incentive is large, most existing borrowers could lower their note rates by refinancing; when it is low only a few existing borrowers could lower their note rates. As shown in the figure, there is a strong positive relationship between refinancing volume and the refinancing incentive.

To describe the importance of the relationship between refinancing activity and interest rates, we first define some key terms. The change in an asset's value with respect to shifts in the risk-free rate—that is, the first derivative of the asset's value with respect to the spot rate, is known as an asset's 'duration'. The change in duration—that is, the second derivative of an asset's value with respect to the spot

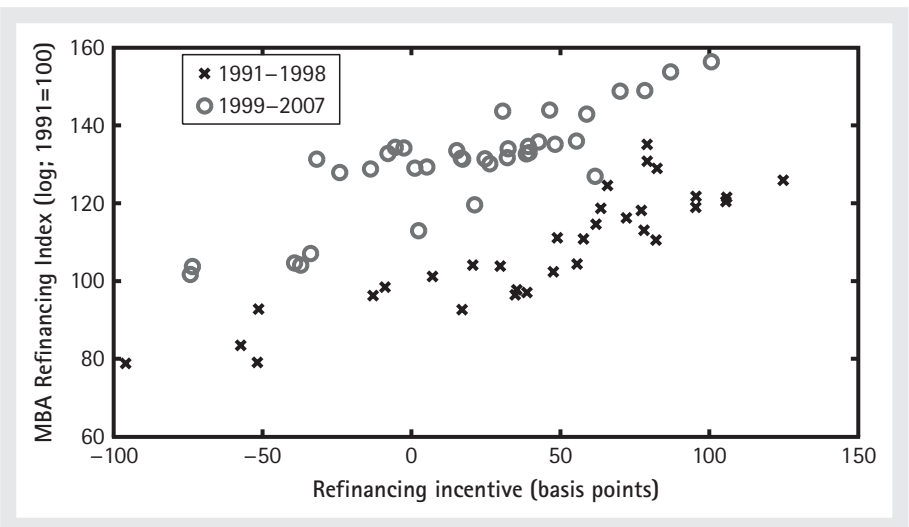

Fig. 23.6. Refinancing incentive and refinancing volume

Note: Figure 23.6 shows the log of the quarterly average Mortgage Bankers Association refinancing applications index plotted against the refinancing incentive, defined as the prevailing rate on new thirty-year fixed rate mortgages less the average rate on all outstanding mortgages.

rate, is known as an asset's 'convexity'. A mortgage's duration is lower (in absolute value) with respect to declines in interest rates than its duration with respect to increases in interest rates. Thus, fixed rate mortgages exhibit 'negative convexity'. Negative convexity is illustrated in Figure 23.7. The dashed line gives the value of a mortgage (actually, of a mortgage-backed security) with respect to parallel shifts in the yield curve.

This negative convexity underlies the difficulty faced by financial institutions when funding portfolios of fixed rate mortgages. If the portfolio is funded with non-callable debt (the solid line in the figure), *any* changes in interest rates will decrease the net value of portfolio because the liability's value falls less (rises more) than the asset's value when rates rise (fall). If the portfolio is funded with demand deposits the situation is more complex. In principle, the value of the liability is constant with respect to interest rate changes. Thus, the holder could benefit from falls in interest rates, but will suffer larger drops in net value if interest rates increase. Intuitively, the portfolio manager is paying the prevailing rate while receiving payments from the mortgage borrower that are insensitive to interest rates. If rates rise, the portfolio manager may have to pay more than he receives from the mortgage borrower, a situation known as 'negative carry'. Thus, holders of fixed rate mortgages must carefully manage their exposure to convexity. Perli and Sack (2003) argue that portfolio managers' desire to hedge US convexity risk is large enough to amplify shocks to interest rates.

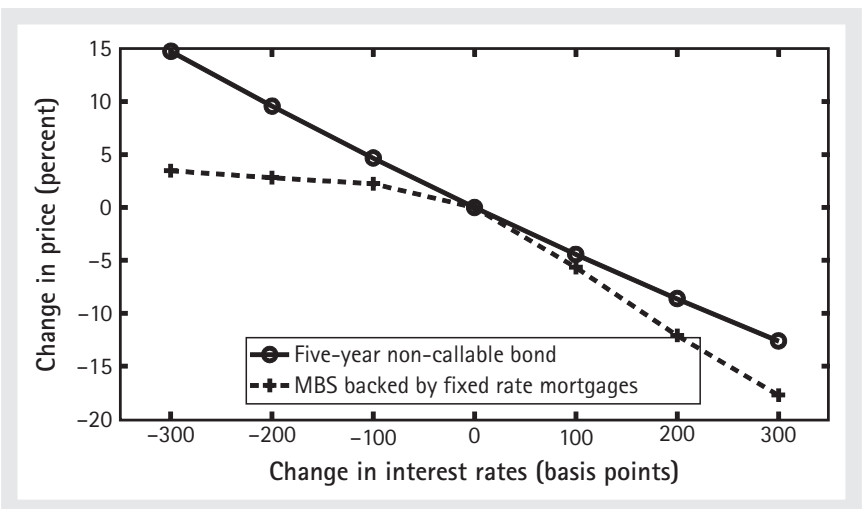

Fig. 23.7. Price response of mortgage-backed securities and corporate bonds to interest rates

Note: Figure 23.7 shows the percentage change in the value of a Fannie Mae MBS backed by fixed rate mortgage paying a coupon of 6.0 percent, and of a five-year Fannie Mae non-callable bond paying a coupon of 3.875 percent issued on 6 June 2008 to indicated parallel shifts in the Treasury yield curve. Results based on Bloomberg's duration model; values retrieved on 12 June 2008.

Note that the difficulty in funding fixed rate mortgages increases the more sharply kinked the value of the asset—that is, the greater the negative convexity. In turn, this implies that the more sensitive borrowers are to the incentive to refinance, the more difficult it is to hold unhedged portfolios of fixed rate mortgages. This is why prepayment penalties or other means of lessening a borrower's incentive to refinance when rates drop can be valuable to investors.

Mortgage-backed securities

Mortgage-backed securities (MBS) are simply financial assets backed by some claim on the cash flow from a group of mortgages. These come in a wide variety of flavors, depending on the differences in prevailing national mortgage institutions and financial regulations. We focus here on the institutions and markets in the US. As with mortgage valuation, an enormous and highly specialized literature has sprung up around MBS, including both theoretical contributions and those designed for industry participants (see Fabozzi and Modigliani, 1992). Here we focus on the relationship of MBS to banks.

Broadly speaking, there are two main types of MBS in the US. Mortgage-backed securities issued by Fannie Mae, Freddie Mac, and Ginnie Mae, which are known as 'agency securities'. Fannie Mae and Freddie Mac are private corporations with Congressional charters; they are known as the housing-related government-sponsored enterprises (GSEs). Agency MBS carry an additional guarantee beyond being collateralized by residential mortgages; Fannie Mae and Freddie Mac each guarantee their own MBS, while Ginnie Mae securities are guaranteed by the federal government. Thus, agency MBS are, to some degree, more like covered bonds than stand-alone asset-backed securities. 'Private-label securities', by contrast, are backed primarily by the pool of underlying mortgages. (Sometimes higher-rated tranches of these MBS will carry third-party guarantees from bond insurance companies.)

For historical reasons, the great majority of prime loans with balances below the 'conforming loan limit' ($417,000 in 2007) are securitized by Fannie Mae and Freddie Mac. (loans with balances in excess of the conforming loan limit are known as 'jumbos'). Thus subprime, near-prime, non-traditional, and prime jumbos are securitized by issuers in the private-label market. Figure 23.8 shows gross issuance of each type of security. As shown, agency security issuance far outpaced private-label issuance, except for a brief period around 2005 at the peak of the US credit boom (also note that issuance of private-label MBS fell essentially to zero in 2008).

Private-label MBS were typically divided into 'tranches' that varied in the seniority of their claims. The most senior tranches had first claim to any payment made by a borrower, while the most junior tranche had last claim. Alternatively, the most junior tranche took the first loss from any defaults. Tranches were rated by

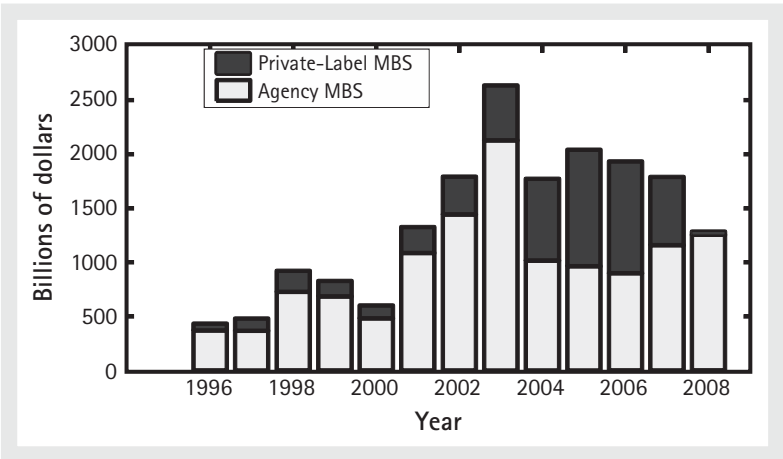

Fig. 23.8. Issuance of US mortgage-backed securities

Note: Figure 23.8 shows the total annual issuance of US mortgage-backed securities. Values for 2008 reflect data through 2008: Q1 at an annual rate.

Source: *Inside Mortgage Finance*.

the major rating agencies (Moody's, Standard & Poor's and Fitch) and marketed based on their rating. In 2007, securities backed by mortgages were rapidly downgraded as the rating agencies grappled with higher than expected defaults.

Figure 23.9 shows that an increasing fraction of mortgages were held in securities over time (either in private-label MBS, on GSE portfolios, or securitized by the GSEs), at the expense of whole loans held on bank portfolios. There are a few main reasons for this shift.

First, as we discuss below, MBS can carry lower capital charges than the equivalent portfolio of whole loans. Thus, banks can lower their regulatory capital charge by securitizing a portfolio of mortgages.

Second, as we discussed earlier, holding fixed rate mortgages on portfolio can be difficult for banks because of the mortgages' negative convexity. Of course, MBS have the same problem. However, MBS are easier and cheaper to sell than whole loans, thus allowing banks to liquidate their holdings if required. Further, pools of MBS can be structured into new securities with more desirable interest rate risk characteristics.

Third, securitization allowed a fundamental change in the industrial organization of mortgage lending. Previously, the three components of mortgage banking (origination, funding, and servicing) had been linked. Securitization allowed lenders to turn over the funding to financial markets.

Securitization could, in principle, lower the cost of funds to the ultimate mortgage borrower. Financial market participants differ in the premium they require to hold certain kinds of risk; by splitting out these risks into separate securities, securitization in theory allows mortgages to be funded at a lower cost

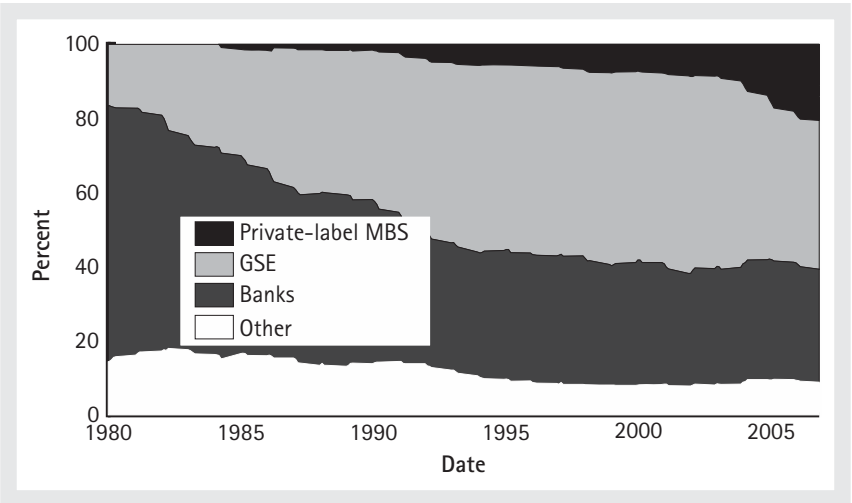

Fig. 23.9. Holders of US residential mortgage debt

Note: Figure 23.9 shows the percent of US residential mortgage debt held by or through various entities. 'GSE' refers to whole loans held by, and securities guaranteed by, the housing-related government-sponsored enterprises known as Fannie Mae and Freddie Mac, and Ginnie Mae; 'banks' refers to commercial banks, thrifts and credit unions; and 'other' includes whole loans held by finance companies, life insurance companies, pension plans, and other entities. Data are from the flow of funds accounts of the US, quarterly, 1970: Q1 through 2007: Q4.

than if a single institution had to hold all of the risks bundled with a mortgage. For example, a specialist hedge fund might feel that it can accurately predict defaults better than most market participants. Thus, it would be willing to pay a higher price than other participants for assets carrying credit risk. This hedge fund would be interested in buying the riskiest tranches of MBS.

Finally, most market participants require a liquidity premium, that is, extra compensation to hold an asset that they might not be able to sell quickly. To the extent that securitization permits mortgages to be traded in a more liquid environment, it can lower the cost of funds to the ultimate borrowers.

Bank capital regulation of mortgages

Financial institutions hold capital to ensure solvency in the face of undiversified, or undiversifiable, risk. Regulators require banks to hold capital, and closely monitor their capital, for the same reason, but also because, in part, they fear that their status as regulated institutions lessens market discipline. A complete discussion of bank capital regulation is beyond the scope of this chapter. Here, we focus on the key attributes that affect the economic capital required to back a portfolio of residential mortgages.

Bank capital regulation as described in the New Basel Capital Accord (or Basel II), adopts the terminology of portfolio credit risk management. Here, expected losses equal the probability of default (PD) times the loss given default (LGD) and the exposure at default (EAD), with a correction of the maturity of the obligation. For mortgages with a fixed maturity and a principal that cannot grow over time, the EAD is essentially fixed because the borrower cannot increase the principal value of his mortgage (this is not true for mortgages with a negative amortization option). Thus, we will focus on the PD and LGD associated with individual mortgages.

The first key attribute of residential mortgages is that credit risk (both PD and LGD) depends heavily on house prices. While geographic diversification can provide some protection for a portfolio, the current US experience, as well as other national house price cycles, suggests that during severe swings, local house price growth is driven by a national factor (see Del Negro, and Otrok, 2005).

Second, credit risk varies significantly by mortgage characteristics. As shown in Figure 23.10, default rates differ even within the narrow class of US subprime mortgages; default rates on variable rate subprime loans have increased far more than default rates on fixed rate subprime loans. More concretely, Calem and LaCour-Little describe a model for computing the economic capital required to back thirty-year fixed rate mortgages in the US. Hancock, et al. (2006) adapt the Calem and LaCour-Little model to compute the 'prudent economic capital' required to back mortgages with a variety of borrower credit scores and loan-to-value ratios. Portfolios of mortgages backed by 'prudent economic capital' are typically rated BBB+ to A–. Hancock,

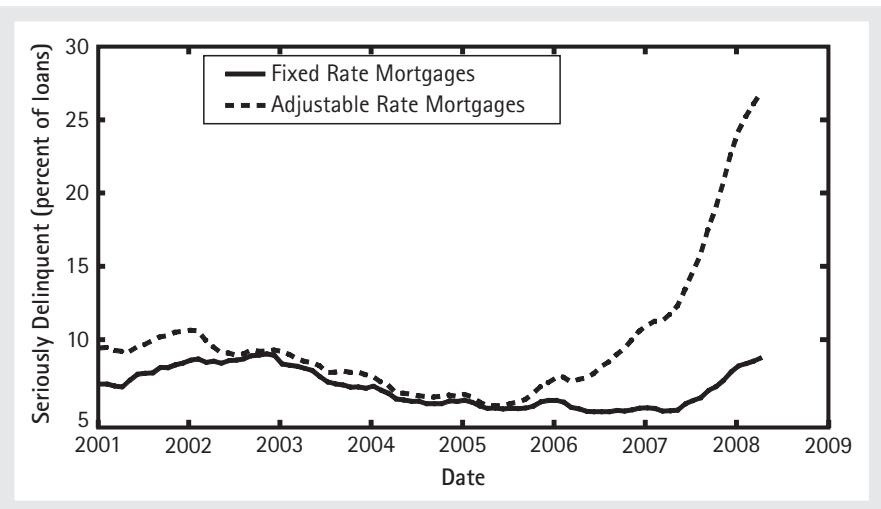

Fig. 23.10. US subprime mortgage delinquency rates

Note: Figure 23.10 shows the percentage of subprime loans seriously delinquent (defined as 90+ days past due or in foreclosure) from Jan. 2001 to Apr. 2008.

Source: First American LoanPerformance.

et al. (2006) find that the lowest-risk loans require very little capital backing—on the order of 20 to 65 basis points. However, higher risk-loans (those with loan-to-value ratios above 90 or severely impaired borrower credit histories) require much more capital, between 1.90 and 7.25 percent, depending on the risk characteristics.

Third, regulatory capital under the older bank capital regime (Basel I) was typically much greater than the prudent economic capital required to hold most residential mortgages. This gave banks a strong incentive to convert whole mortgage loans into securities; the housing-related GSEs known as Fannie Mae and Freddie Mac, for example, had to hold only 45 basis points of capital against credit risk during the regulatory regime that existed prior to the GSE accounting scandals of the mid-2000s. As shown in Figure 23.9, the fraction of mortgage debt outstanding held as whole loans on banks' balance sheets declined significantly from 1980 to 1995, while that securitized or held by the GSEs grew over the same period. Starting in the mid-2000s, private-label mortgage-backed securities issuers began to account for a significant fraction of mortgage debt outstanding. This shift is consistent with the lower capital required on securitized assets.

Related, as one might expect under capital regulation that was broadly insensitive to credit risk, banks also had an incentive to hold riskier mortgages. Alternatively, if they were securitizing a large fraction of their loans, they had an incentive to hold the riskier loans in order to maintain their reputation with secondary market investors for not selling risky loans. Ambrose, LaCour-Little, and Sanders (2005) examined the performance of the loans a major lender securitized as opposed to those that it chose to hold on portfolio and found that, as expected, it chose to keep the riskier loans.

Finally, estimating the correct capital required to back mortgage portfolios requires data on the performance of a wide variety of mortgages under many different circumstances. Banks may not have access to a long enough time series of performance, or performance data on new types of mortgages that they are considering making. Further, under Basel II, regulatory capital has to be estimated using so-called *stress LGDs*; that is, losses experienced during severe economic downturns. Yet, suitable events are relatively rare, and it may be hard to generalize from a narrow set of mortgages that underwent the stress event to a broader set of mortgages that exist at a later time or in a different geographic area.

US MORTGAGE MARKET TURMOIL: 2006–8

Primary and secondary markets for mortgages underwent significant turmoil starting in late 2006. The initial locus of the crisis was the secondary market for

subprime mortgages, but other markets, including broader mortgage markets, were also affected.

Background on subprime mortgages

As we discussed, the 1990s and 2000s saw an increase in the willingness of lenders to extend credit to borrowers with high LTVs and weaker credit histories. The riskiest of these loans were known as 'subprime'. Until the mid-1990s a typical subprime borrower would probably not have been eligible for a mortgage at any interest rate. Even after the initial growth of the subprime market in the late 1990s, such mortgages were traditionally not used to purchase homes. Instead, subprime loans were cash-out refinancings made to homeowners who had experienced a severe financial reversal (such as job loss or uninsured medical expenses), had subsequently missed payments on credit cards and other consumer loans, and who thus had urgent need for substantial cash while also having poor credit histories. Subprime mortgages allowed these to borrowers to tap the equity in their homes via cash-out refinancing.

Table 23.2 shows the distribution of US mortgages as of 2008 Q1. As shown, about half of subprime mortgages carried variable rates, mainly hybrids of the kind described in Table 23.1. (These hybrids accounted for a larger share of originations over the past few years but, because they terminate faster than fixed rate loans, account for a smaller share of outstandings.) As shown in Figure 23.10, serious delinquency rates on variable rate subprime loans, as opposed to fixed rate subprime loans, rose especially quickly and stood at over 25 percent by April 2008.

Table 23.2. Outstanding US mortgages as of 2008: Q1

Loan type	Number (millions)	Percent
All loans	54.7	100
Subprime loans	6.7	12
Fixed rate	3.2	6
Variable rate	3.2	6
Other	0.3	1
Prime loans	42.7	78
Fixed rate	33.5	61
Variable rate	7.7	14
Other	1.5	3
FHA/VA	5.3	10

Note: Table 23.2 gives the estimated distribution of outstanding first-lien mortgages in the US as of 2008: Q1. Distribution based on statistics from the Mortgage Bankers Association; total number of mortgages estimated based on data from the Census and the Survey of Consumer Finances. Near-prime mortgages such as those sold into alt-A securities are split between the prime and subprime categories.

Most subprime loans found their way into mortgage-backed securities. Indeed, Avery, Brevoort, and Canner (2007) find that about half of subprime loans were originated by finance companies not affiliated with a depository institution. These companies typically found it expensive to hold loans on portfolio and instead sold the overwhelming majority of loans they originated. Issuance of securities backed by subprime mortgages boomed in 2005 and 2006.

House price growth peaked in several geographic locations in the US around mid-2006 and then began to decelerate rapidly. This in turn led to an increase in mortgage delinquency rates. Beginning in late 2006, the incidence of early payment defaults, usually defined as defaults within ninety days of origination, sharply increased. While some small fraction of subprime borrowers would normally default within three months, such a rapid rise was unusual. Originators typically agreed to buy back loans that defaulted so early. Faced with a rising number of claims, these companies began to fail, starting with Ownit Mortgage Solutions, which closed its doors in December 2006.

Rapidly rising defaults also led rating agencies to downgrade private-label securities backed by subprime mortgages; thus, investors who had purchased securities with investment-grade ratings in 2006 found themselves holding lower-rated securities in 2007. The rating agencies undertook comprehensive reviews of the models and procedures used to rate MBS. Nonetheless, investors apparently remained skeptical, and appetite for the broad category of MBS dried up. Issuance of private-label MBS generally, and subprime MBS in particular, fell sharply in 2007. The market for private-label MBS remained closed in 2008.

As a result, borrowers whose loans were traditionally sold into the private-label MBS market found credit substantially more expensive and harder to come by. This in turn exacerbated the decline in housing demand in the US.

Government authorities and industry groups have taken a number of steps to ameliorate the problems stemming from the rapid increase in mortgage defaults. The Hope Now Alliance, a consortium of banks, servicers, and investors, has worked to establish standards under which servicers can modify loans held in private-label mortgage-backed securities as well as tracking data on mortgage workouts. To ease the liquidity premium paid by prime borrowers with jumbo loans, the Congress and the Administration temporarily raised the conforming loan limit on loans originated in areas with higher home prices. Regulators and community and industry groups were considering a number of further actions as of mid-2008.

Research on subprime mortgages

Subprime mortgages have attracted considerable attention from researchers. Avery, Brevoort, and Canner (2007) analyze data collected under the Home

Mortgage Disclosure Act (HMDA) to quantify the size of the broad non-prime loan market. As shown in Table 23.2, subprime loans account for only about 13 percent of outstanding mortgages—however, Avery, Brevoort, and Canner show that these mortgages played a considerably larger role in home purchases in 2006. Mayer and Pence (2008) provide the first comprehensive description of the location of subprime mortgages; they also compare the various competing definitions of 'subprime'. They find that subprime mortgages (broadly defined) were more common in economically hard-hit areas, as well as areas with outsized new home construction. Finally, Mayer, Pence, and Sherlund (2008) provide a detailed examination of subprime defaults.

Gerardi, Shapiro, and Willen (2007) recognize that studying mortgage performance can give a misleading picture of the distress of the underlying borrowers. A subprime borrower who successfully refinances poses no risks to the original lender—however, the borrower may subsequently default on the new loan. Thus, they use county courthouse records to construct a dataset of ownership experiences. They find that house prices are the key factor driving ownership outcomes. Demyanyk and Van Hemert (2008) also find that house prices are a key driver of subprime mortgage outcomes—however, they also find that the out-of-sample fit of standard underwriting models began to deteriorate in 2004 and, in real time, ought to have been detectable by 2005. Interestingly, Mian and Sufi (2008) argue that credit expansion itself drove house prices up, at least for a time. They find that geographic areas where more borrowers were turned down for loans experienced greater increases in lending and higher house price appreciation. They argue that improved risk modeling allowed lenders to satisfy the 'latent demand' posed by would-be borrowers who had previously been rejected for a mortgage. Keys et al. (2008), among others, argue that, because many originators were selling mortgages to secondary market investors, they had less incentive rigorously to underwrite loans and greater incentive to increase origination volume. The incentives of the originators were fairly clear at the time; this argument does beg the question of why secondary market investors didn't charge more of a lemon's premium on loans originated by independent mortgage companies.

Gerardi, et al. (2008) use both quantitative models and a close reading of contemporary documents to determine whether market participants underestimated the consequences of a drop in house prices on subprime mortgage performance, or whether they simply discounted the possibility of a widespread drop in house prices. They find that, even using data that did not contain big house price drops, investors should have been able to forecast that falling house prices would lead to rising defaults and decreasing prepayments. Contemporary documents show that market participants thought it highly unlikely that nationwide nominal house prices would fall significantly.

Empirical evidence on subprime defaults and prepayments

Subprime mortgage underwriting standards changed significantly over the past decade, and along several dimensions. We estimate a simple loan-level regression model of subprime mortgage performance to determine which factors appear to best explain the recent deterioration of subprime mortgage credit quality. See also Sherlund, 2008. The outcomes of interest will be whether a loan defaults within twelve months of origination, and, separately, whether it refinances within that time.

We use administrative data on loan characteristics and outcomes sold by First American LoanPerformance. These data cover only loans sold into subprime private-label mortgage-backed securities. The loans used here were originated between 1999 and 2006. Table 23.3 gives sample statistics from the data.

The control variables cover several of the key features discussed in this chapter. The main variables include the log difference in the local house price index in the eighteen months after origination, to proxy for the appreciation of the underlying property. The CTLV measures the all-in combined-loan-to-value ratio at origination;

Table 23.3. Sample statistics on loans used default/prepayment analysis

Variable	Mean	Std. Dev.
Defaults within 12 months	0.0522	0.2225
Refinances within 12 months	0.1695	0.3752
House price appreciation (18 months)	0.1273	0.1092
Combined loan-to-value (CLTV)	82.4705	14.0356
Second lien present	0.1333	0.3399
Is a 228	0.5907	0.4917
Is a 327	0.1327	0.3392
Is a fixed rate mortgage	0.2766	0.4473
Is a refi	0.6773	0.4675
Loan source: retail	0.0656	0.2476
Loan source: wholesale	0.1116	0.3149
Loan source: broker	0.0119	0.1085
Loan source: unknown	0.8109	0.3916
Interest-only term present (FRM)	0.0039	0.0622
Interest-only term present (ARM)	0.0864	0.2810
Mortgage features balloon payment	0.0371	0.1890
Non-owner occupier	0.0653	0.2471
Full documentation	0.6882	0.4632
Low documentation	0.3033	0.4597
Borrower's FICO score at origination	610.6717	60.3626
Prepayment penalty present.	0.7612	0.4264
Initial rate	8.0662	1.5863
Number of observations		5.447,455

Note: Table 23.3 gives statistics on a sample of subprime first lien mortgages originated in the US between 1999 and 2006 and sold into secondary market pools; data are published by First American LoanPerformance.

we include a separate indicator for whether this leverage was achieved using a second lien (e.g., an 80 percent LTV loan and a 20 percent LTV 'piggyback' loan as opposed to a single 100 percent LTV loan). The three main mortgage types considered are fixed rate mortgages and so-called '2/28' and '3/27' hybrids. These hybrid adjustable rate mortgages carry a fairly low initial rate for two or three years, after which they adjust to the six-month LIBOR with an extremely high margin (more than 500 basis points). As described above, payments on these mortgages can increase dramatically after their first rate reset. The regression also includes the purpose of the loan (i.e., to refinance an existing loan or to purchase a home). A series of indicator variables flag for the entity that originated the loan (if known): a retail branch, via a wholesale relationship, or through a broker; for about 80 percent of the loans, the origination channel is not known. We also include a series of indicators to flag non-standard amortization schedules, with special attention to whether these were attached to fixed or adjustable rate loans. Non-owner occupiers are expected to be especially ruthless in their default behavior because they do not face eviction. We include indicators for loans originated with full verification of income and assets, and those in which verification was incomplete; the excluded category are those loans in which there was no verification of income or assets. We include a popular measure of the borrower's credit history, the so-called 'FICO score', and a flag for whether or not the loan carried a prepayment penalty.

Table 23.4 gives the results (note that with about 5.5 million observations, all estimated coefficients are highly statistically significant, so standard errors are suppressed). Note that the data available do not include all of the information used in the underwriting decision. Thus, these coefficients cannot be taken as truly estimating the causal relationship between the right-hand side variables and the outcomes of interest.

In the default model, the first three variables summarize how much equity the borrower has in the property—not surprisingly, the less equity, the greater the probability of default. The teaser-rate hybrid ARMs known as 2/28s and 3/27s have higher default rates than fixed rate loans (the excluded category), perhaps because fixed rate loans have a longer expected life, so lenders exercised greater care in making them. Borrowers using 2/28s and 3/27s may have had a shorter planning horizon and hence a more speculative motive for borrowing. Loans originated as refinancings had lower default rates than loans originated as purchases (the excluded category). Loans whose origination channel is unknown are likelier to default than others. The presence of some kind of interest-only feature apparently led to decreased default rates for both ARMs and FRMs, while the presence of a balloon increased default rates. As expected, non-owner occupiers defaulted at greater rates. Loans with more complete documentation defaulted less than loans with no documentation. Borrowers with lower credit scores were likelier to default.

In the refinancing model, almost all the coefficients have the opposite sign as in the default model. In part, this is the competing hazard effect: anything that makes a

Table 23.4. Probit estimation of defaults and prepayments

Variable	Marginal effects	
	Default within 12 months	Refinance within 12 months
House price appreciation (18 months)	−0.1590	0.3552
Combined loan-to-value (CLTV)	0.0008	−0.0014
Second lien present	0.0091	−0.0040
Is a 2/28	0.0123	0.0892
Is a 3/27	0.0131	0.0726
Is a refi	−0.0201	0.0429
Loan source: Retail	0.0079	0.0210
Loan Source: Wholesale	0.0011	0.0066
Loan Source: Broker	0.0032	0.0071
Interest-only term present (FRM)	−0.0084	0.0831
Interest-only term present (ARM)	−0.0083	0.0400
Mortgage features balloon payment	0.0137	0.0548
Non-owner occupier	0.0249	−0.0004
Full documentation	−0.0308	−0.0225
Low documentation	−0.0086	0.0143
Borrower's FICO score at origination	−0.0005	−0.0002
Prepayment penalty present	0.0026	−0.0982
Pseudo R2	0.0748	0.0447
Observed probability	0.0523	0.1698
Predicted probability (at variable means)	0.0398	0.1589

Notes: Table 23.4 gives estimated marginal effects from a probit regression of indicator variables for loan-level default and prepayment within a year of origination. All coefficients are statistically different from zero at all measurable levels of confidence, so statistics on inference are suppressed to save space. Marginal effects give the increase in the probability of the given outcome from a one standard deviation increase in the RHS variables holding other variables at their means; for an indicator variable, marginal effect is computed by switching the variable from 0 to 1.

borrower more likely to default should make him less likely to refinance. There are a few interesting exceptions. Hybrid mortgages (2/28s and 3/27s) refinance *and* default more than fixed rate mortgages. Presumably, borrowers are refinancing ahead of the major increase in payments built into these mortgages. The same logic probably governs the positive effect of a balloon feature on both default and prepayment probability.

These results can shed light on the rapid rise in defaults on subprime mortgages in 2006–7. Table 23.5 shows the outcomes of interest as well as the average eighteen-month house price appreciation; average CLTV; percent of loans with complete documentation; and average 'FICO score' by year of origination. As shown, mortgages originated in later years were much more likely to default than mortgages originated in earlier years. These loans also experienced radically lower house price appreciation, had higher LTVs, and were less likely to have complete documentation. Note that borrower credit scores actually rose over time. Comparing these changes over time with the marginal effects shown in Table 23.4 suggests that the

Table 23.5. Subprime outcomes and characteristics 1999–2006

Year	Default	Refinance	House price appreciation	CLTV	Fully documented	FICO Score
1999	0.06	0.09	0.11	76.59	0.72	593
2000	0.08	0.11	0.10	76.77	0.78	584
2001	0.06	0.11	0.11	79.10	0.78	598
2002	0.04	0.14	0.14	80.04	0.72	605
2003	0.03	0.18	0.19	81.96	0.70	614
2004	0.04	0.20	0.18	83.41	0.68	614
2005	0.06	0.18	0.08	84.65	0.65	616
2006	0.11	0.20	0.00	85.15	0.64	612

Note: Table 23.5 gives sample means of indicated variables by year of loan origination. Default and refinance outcomes are measured against a twelve-month window, as before.

sharp slowdown in house price appreciation was the single most important factor in explaining the rise in delinquency rates, although the willingness of lenders to make high CLTV loans and loans with incomplete documentation also played a role.

Conclusion

Mortgage lending is one of the oldest forms of intermediation; it is often the largest credit transaction undertaken by a household, and is an important part of the financial services sector. Mortgage banking comprises three functions: origination, funding, and servicing. Advances in technology and financial market developments have allowed institutions to split these three parts. Mortgages can now be originated by small-scale operators using web-based underwriting applications, sold to financial institutions who in turn package the loans and sell them to investors, and serviced by specialty operations that take advantage of economies of scale. This unbundling of mortgage banking has highlighted the different decision problems faced by borrowers, mortgage funders, and regulators. Many of the tensions inherent in the new mortgage business model were highlighted by the recent turmoil in the US subprime market.

Residential mortgages are collateralized obligations of the household sector. They tend to carry longer maturities and lower interest rates than unsecured borrowing. However, the advantages of mortgages over other forms of borrowing depend on the ability of lenders to seize the underlying collateral. When originating loans, lenders must consider the value of the underlying asset, the amount of equity used by the borrower, the borrower's income and assets, and the borrower's history of making timely payments on other debt obligations.

Households face two types of decisions regarding mortgages: decisions made before origination regarding the characteristics of the household's preferred loan, and decisions made after origination regarding whether to refinance or default. Before origination, households face a menu of choices that varies across countries. Broadly speaking, households must decide on how much to borrow, how much interest rate risk to accept, and what kind of amortization schedule to use.

Once in a mortgage, households may be able to refinance the mortgage, or they may decide, or be forced by circumstances, to default on the mortgage. Economists generally view these choices as options embedded in the mortgage. However, pure option-theoretic models have difficulty matching borrowers' actually decisions, especially the decision to default, suggesting that transactions costs play a big role in household decision making.

Of course, any options embedded in a mortgage and available to the household were placed there by the lender, who must price these options and determine appropriate compensation. A large and diverse literature has grown up around the problem of accurately modeling the household's refinancing decision. The more sensitive the borrower is to changes in the spot interest rate, the more care the ultimate holder of a leveraged portfolio of mortgages must take in matching the duration of the mortgage assets and the liabilities used to fund them. Sudden changes in interest rates can decrease the net value of the portfolio.

Bank capital regulators face the problem of matching the regulatory capital on mortgages to their economic capital. Indications are that capital charges under the older Basel I scheme are higher than warranted by the credit risk on most mortgages. This provides an incentive for regulated banks to move some mortgages off balance sheet.

Finally, widespread financial disruptions of 2007–8 had their initial locus in the US subprime mortgage market. These loans were made to borrowers with poor credit histories, very high LTVs, or some combination of both. Over time, more loans with high LTVs were made, as were loans with incomplete verification of income and assets. Empirical evidence suggests that this deterioration in underwriting standards, coupled with the rapid deceleration of US house prices that began in 2006, together led to the rapid rise in subprime defaults from 2006 to 2008.

References

Ambrose, B. W., C. A. Capone, Jr., and Y. Deng (2001). 'Optimal Put Exercise: An Empirical Examination of Conditions for Mortgage Foreclosure', *Journal of Real Estate Finance and Economics,* 23: 213–34.

—— M. LaCour-Little, and A. B. Sanders (2005). 'Does Regulatory Capital Arbitrage, Reputation, or Asymmetric Information Drive Securitization?', *Journal of Financial Services Research,* 28: 113–33.

Amromin, G., J. Huang, and C. Sialm (2007). 'The Tradeoff between Mortgage Prepayments and Tax-Deferred Retirement Savings', *Journal of Public Economics,* 91: 2014–40.

Avery, R., K. Brevoort, and G. Canner (2007). 'The 2006 HMDA Data', *Federal Reserve Bulletin,* 93: A73–A109.

Bank for International Settlements (2006). 'Housing Finance in the Global Financial Market', Committee for the Global Financial System Paper 26, http://www.bis.org/publ/cgfs26.pdf.

Barakova, I., R. Bostic, P. Calem, and S. Wachter (2003). 'Does Credit Quality Matter for Homeownership?', *Journal of Housing Economics,* 12: 318–36.

Ben-David, I. (2008). 'Manipulation of Collateral Values by Borrowers and Intermediaries', Manuscript, Graduate School of Business, University of Chicago.

Brueckner, J. (1993). 'Why Do We Have ARMs?', *American Real Estate and Urban Economic Associations Journal,* 21: 333–45.

—— (1994). 'Borrower Mobility, Adverse Selection and Mortgage Points', *Journal of Financial Intermediation,* 3: 416–41.

—— and J. R. Follain (1988). 'The Rise and Fall of the ARM: An Econometric Analysis of Mortgage Choice', *Review of Economics and Statistics,* 70: 93–102.

Bucks, B., A. Kennickell, and K. Moore (2006). 'Recent Changes in U.S. Family Finances: Evidence from the 2001 and 2004 Survey of Consumer Finances', *Federal Reserve Bulletin,* 92: A1–A38.

—— and K. Pence (2006). 'Do Homeowners Know Their House Values and Mortgage Terms?', Finance and Economics Discussion Series 2006–03, Board of Governors, Washington, DC. http://www.federalreserve.gov/Pubs/feds/2006/200603/index.html.

Calem, P. S. and M. LaCour-Little (2004). 'Risk-Based Capital Requirements for Mortgage Loans', *Journal of Banking & Finance,* 28: 647–72.

Campbell, J. Y. and J. Cocco (2003). 'Household Risk Management and Optimal Mortgage Choice', *Quarterly Journal of Economics,* 118: 1449–94.

Canner, G., K. Dynan, and W. Passmore (2002). 'Mortgage Refinancing in 2001 and early 2002', *Federal Reserve Bulletin,* 88: 469–90. http://www.federalreserve.gov/pubs/bulletin/2002/1202lead.pdf.

Capone, Jr., C. A. (1996). *Providing Alternatives to Mortgage Foreclosure: A Report to Congress.* Washington, DC: US Department of Housing and Urban Development.

Ciochetti, B. A. (1997). 'Loss Characteristics of Commercial Mortgage Foreclosure', *Real Estate Finance,* 14: 53–69.

Clauretie, T. M. (1989). 'State Foreclosure Laws, Risk Shifting, and the PMI Industry', *Journal of Risk and Insurance,* 56: 544–54.

—— and T. Herzog (1990). 'The Effect of State Foreclosure Laws on Loan Losses: Evidence from the Mortgage Insurance Industry', *Journal of Money, Credit and Banking,* 22: 221–33.

Cordell, L., K. Dynan, N. Liang, and E. Mauskopf (2008). 'The Incentives of Mortgage Servicers: Myths and Realities'. Finance and Economics Discussion Series 2008-46, Board of Governors, Washington, DC. http://www.federalreserve.gov/pubs/feds/2008/200846/200846abs.html.

Del Negro, M. and C. Otrok (2005). 'Monetary Policy and the House Price Boom across U.S. States', Federal Reserve Bank of Atlanta Working Paper 2005–24.

Demyanyk, Y. and O. van Hemert (2008). 'Understanding the Subprime Mortgage Crisis', Manuscript, Stern School of Business, New York University.

Deng, Y., J. M. Quigley, and R. van Order (2000). 'Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options', *Econometrica,* 68: 275–307.

Duffie, D. (1992). *Dynamic Asset Pricing Theory,* Princeton, NJ: Princeton University Press.

Dunn, K. B. and C. S. Spatt (1985). 'An Analysis of Mortgage Contracting: Prepayment Penalties and the Due-on-Sale Clause', *Journal of Finance,* 40: 293–308.

Edelberg, W. (2006). 'Risk-Based Pricing of Interest Rates for Consumer Loans', *Journal of Monetary Economics,* 53: 2283–98.

Eggert, K. (2007). 'Comment: What Prevents Loan Modifications?', *Housing Policy Debate,* 18: 279–97.

Elliehausen, G., M. E. Staten, and J. Steinbuks (2008). 'The Effect of Prepayment Penalties on the Pricing of Subprime Mortgages', *Journal of Economics & Business,* 60: 33–46.

Fabozzi, F. J. and F. Modigliani (1992). *Mortgage and Mortgage-Backed Securities Markets,* Boston, MA: Harvard Business School Press.

Foote, C., K. Gerardi, and P. Willen (2008). 'Negative Equity and Foreclosures: Theory and Evidence', Federal Reserve Bank of Boston Public Policy Discussion Paper 08-3. http://www.bos.frb.org/economic/ppdp/2008/ppdp0803.pdf.

Fu, Q., M. LaCour-Little, and K. Vandell (1997). 'Retiring Early: An Empirical Analysis of the Mortgage Curtailment Decision', University of Wisconsin Center for Urban Land Economic Research Working Paper 97–09.

Gerardi, K., A. Lehnert, S. Sherlund, P. Willen (2008). 'Making Sense of the Subprime Crisis', *Brookings Papers on Economic Activity,* 2008: 69–159.

——A. H. Shapiro, and P. Willen (2008). 'Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures', Federal Reserve Bank of Boston Working Paper 07-15. http://www.bos.frb.org/economic/wp/wp2007/wp0715.pdf.

Graham, J. R. (2000). 'How Big Are the Tax Benefits of debt?', *The Journal of Finance,* 55: 1901–42.

Hancock, D., A. Lehnert, W. Passmore, and S. M. Sherlund (2006). 'The Competitive Effects of Risk-Based Capital Regulation: An Example from U.S. Mortgage Markets', Finance and Economics Discussion Series 2006–46, Board of Governors of the Federal Reserve System, Washington, DC. http://www.federalreserve.gov/Pubs/Feds/2006/200646/200646pap.pdf.

Haurin, D. R., C. E. Herbert, and S. S. Rosenthal (2007). 'Homeownership Gaps among Low-Income and Minority Households', *Cityscape,* 9: 5–52.

Hurst, E. and F. Stafford (2004). 'Home is Where the Equity Is: Mortgage Refinancing and Household Consumption', *Journal of Money, Credit, and Banking,* 36: 985–1014.

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig (2008). 'Did Securitization Lead to Lax Screening? Evidence from Subprime Loans', Manuscript, Graduate School of Business, University of Chicago.

LaCour-Little, M. and S. Malpezzi (2003). 'Appraisal Quality and Residential Mortgage Default: Evidence from Alaska', *Journal of Real Estate Finance and Economics,* 27: 211–33.

—— and J. Yang (2008). 'Pay Me Now or Pay Me Later: Alternative Mortgage Products and Housing Consumption', Manuscript, Real Estate & Land Use Institute, College of Business and Economics, California State University, Fullerton.

Lamont, O. and J. Stein (1999). 'Leverage and House-Price Dynamics in U.S. Cities'. *RAND Journal of Economics,* 30: 498–514.

LeRoy, S. F. (1996). 'Mortgage Valuation under Optimal Prepayment', *Review of Financial Studies,* 9: 817–44.

Leventis, A. (2006). 'Removing Appraisal Bias from a Repeat-Transactions House Price Index: A Basic Approach', Office of Federal Housing Enterprise Oversight (OFHEO) Working Paper 06-1. http://ssrn.com/abstract=1084984.

Mayer, C. and K. Pence (2008). 'Subprime Mortgages: What, Where, and to Whom?' Manuscript, Federal Reserve Board, Washington, DC.

——— and S. M. Sherlund (2008). 'The Rise in Mortgage Defaults: Facts and Myths', Manuscript, Federal Reserve Board, Washington, DC.

Mian, A. R. and A. Sufi (2008). 'The Consequences of Mortgage Credit Expansion: Evidence from the 2007 Mortgage Default Crisis', Manuscript, Graduate School of Business, University of Chicago.

Miles, D. (2004). *The UK Mortgage Market: Taking a Longer-Term View,* Norwich, UK: HMSO.

Pence, K. (2006). 'Foreclosing on Opportunity: State Laws and Mortgage Credit', *Review of Economics and Statistics,* 88: 177–82.

Perli, R. and B. Sack (2003). 'Does Mortgage Hedging Amplify Movements in Long-Term Interest Rates?' *The Journal of Fixed Income,* 13: 7–17.

Piskorski, T. and A. Tchistyi (2007). 'Optimal Mortgage Design', Manuscript, Graduate School of Business, Columbia University. http://www1.gsb.columbia.edu/mygsb/faculty/research/pubfiles/2686/Optimal%20Mortgage%20Design%2Epdf.

Quigley, J. and R. Van Order (1995). 'Explicit Tests of Contingent Claims Models of Mortgage Default', *Journal of Real Estate Finance and Economics,* 11: 99–117.

Rothberg, J. P., F. E. Nothaft, and S. A. Gabriel (1989). 'On the Determinants of Yield Spreads between Mortgage Pass-Through and Treasury Securities', *Journal of Real Estate Finance and Economics,* 2: 301–15.

Schwartz, E. S. and W. N. Torous (1989). Prepayment and the Valuation of Mortgage-Backed Securities, *Journal of Finance,* 44: 375–92.

Sherlund, S. (2008). 'The Past, Present, and Future of Subprime Mortgages'. Finance and Economics Discussion Series 2008-63, Federal Reserve Board, Washington, DC. http://www.federalreserve.gov/pubs/feds/2008/200863/200863abs.html.

Shilling, J. D., U. S. Dhillon, and C. F. Sirmans (1987). 'Choosing Between Fixed and Adjustable Rate Mortgages', *Journal of Money, Credit, and Banking,* 19: 260–7.

Stanton, R. and N. Wallace (1999). 'Anatomy of an ARM: The Interest Rate Risk of Adjustable Rate Mortgages'. *Journal of Real Estate Finance and Economics,* 19: 49–67.

Stegman, M. A., R. G. Quercia, J. Ratcliffe, L. Ding, and W. R. Davis (2007). 'Preventive Servicing is Good for Business and Affordable Homeownership Policy', *Housing Policy Debate,* 18: 243–78.

Stein, J. (1995). 'Prices and Trading Volume in the Housing Market: A Model with Down-payment Effects', *Quarterly Journal of Economics,* 110: 379–406.

CHAPTER 24

SECURITIZATION

INSTRUMENTS AND IMPLICATIONS

DAVID MARQUES IBANEZ
MARTIN SCHEICHER[1]

Introduction

Prior to the 2007–8 credit risk turmoil, banks securitized an increasingly wide range of assets. Securitization can be defined as the issuance of claims backed by a pool of default-risky instruments where the new claims frequently have varying exposures to the underlying pool of collateral. Initially the most commonly securitized assets were mortgage loans, while in the run up to the recent credit crisis, more sophisticated forms of securitization were developed. As a result, banks and other corporations were able to securitize in recent years large portions of their credit book. Due to their sheer size and relative importance, this chapter focuses on securitization originated by banks.

These developments in securitization activity have produced significant changes both in the financial structure of most developed countries and in the role of banks

[1] The opinions expressed in this chapter are those of the authors only and do not involve the responsibility of the institutions to which they are affiliated.

therein. In this respect, the increase in securitization activity has modified the functioning of credit markets, reducing the fundamental role of liquidity transformation traditionally performed by financial intermediaries. The changing role of banks from 'originate and hold' to 'originate, repackage, and sell' has also had implications for the incentives of banks to grant credit, take on more risks, and react to monetary policy changes.

Over the past decade and prior to the credit risk turmoil, securitization and more broadly credit risk transfer techniques expanded tremendously. While securitization in a narrow view has been used as a technique in the US for more than fifty years, the decade prior to the credit turmoil coincided with spectacular increases in the amount of securitization activity, in the number countries using these techniques, and in the development of new credit risk transfer instruments. Regarding the latter, instruments grew dramatically both in number and complexity. This revolution in credit risk transfer techniques can be traced back to a number of concurrent factors, such as financial market globalization, improvements in information technology, and pricing models as well as the movement toward a more market-based financial system. The large increases in the use of securitization activity formed part of a wider trend of financial innovation toward the commoditization and trading of credit risk. This trend also encompassed the development of credit derivatives and the use of securitization techniques, often combined with more traditional forms of transferring credit risk such as the syndicated loan market. Together, all these developments helped to make credit risk more tradable.

At the same time, the recent credit crisis brought to the fore some significant features of the market for credit risk transfer, which can impair market functioning in times of strain. These relate to various incentive problems, investors' overreliance on credit ratings, increased levels of complexity and opacity, and related valuation difficulties. The credit crisis was initially triggered by strongly rising delinquencies in US subprime mortgage markets. Market participants then became increasingly concerned about the valuation of all credit risk transfer instruments in their portfolio as well as to the possible positions of their counterparts. At this juncture, mark-to-market and mark-to-model valuations as well as the risk assessments of rating agencies were all called into question, leading to a collapse in primary credit markets.[2] The corresponding general repricing of credit risk manifested itself in rising credit spreads in many segments of the credit market and a sharp decline in new issuance.

[2] In a mark-to-market valuation, pricing is based on observed market prices whereas mark-to-model valuations rely on the implementation of a theoretical pricing model.

The dynamic growth of securitization occurred against the backdrop of other changes in the financial system. In this respect, after the collapse of the new-economy stock markets, investors started to diversify more significantly their portfolios into credit markets thereby encouraging the development of new credit products including corporate bonds and securitization. This was also fuelled by strong demand from institutional investors such as mutual funds, pension funds, and insurance corporations. Rapid innovations in credit risk modeling encouraged by the Basel II process also played a major role promoting the development of securitization activity.

This chapter starts by introducing some of the main instruments used in the securitization markets, which are often utilized as building blocks to construct more complex instruments. It then briefly reviews the spectacular increase in securitization that occurred globally and considers the main originators' motives for using those instruments. The last section is more tentative and in light of recent evidence discusses some potential effects of securitization on how banks operate, on the broad financial system, on risk taking, and on the transmission mechanism of monetary policy. Throughout the chapter, we consider the effects of the recent crisis on credit markets and dwell on how it has affected securitization and more broadly the markets for credit risk transfers. We illustrate the repricing both for corporate credit risk as well as for US subprime 'asset-backed securities' (ABS).

Credit risk transfer: Main instruments

This section considers the main securitization instruments. It starts by analyzing traditional 'granular' instruments with a special emphasis on 'mortgage-backed securities' (MBS) and shows how they differ from an important alternative tool of market funding for banks: The covered bond market. It then considers more complex securities used in synthetic securitization. Namely, it describes 'credit default swaps' (CDSs) and 'collateralized debt obligations' (CDOs) as well as other contracts often constructed as a combination of these two instruments. It then provides a picture of events in the credit market turmoil by analyzing developments in prices of CDS index tranches. These developments are then linked to difficulties in pricing and valuing some of the more complex instruments. Against this background, the chapter illustrates the spectacular increase in securitization volumes in recent years globally and shows the decline in issuance as a result of the recent turmoil in credit markets.

ABS, MBS, AND OTHER FUNDED INSTRUMENTS

While securitization can be traced back to the 1930s (when the Federal National Mortgage Association was created to buy and sell federally insured mortgages), the modern foundations of securitization originated from developments in the residential mortgage market in the 1970s (Kendall and Fishman, 1996). Namely, the market for asset-backed securities developed during the early 1970s by means of government-sponsored agencies (such as the Federal National Mortgage Association, 'Fannie Mae', and the Federal Home Loan Mortgage Corporation, 'Freddie Mac') that enhanced mortgage loan liquidity by issuing and guaranteeing, but not originating, MBS.

Traditional securitization can be defined as the pooling of credit risky assets, such as residential mortgage loans, and their subsequent sale to a special purpose vehicle (SPV) which then issues securities (see, e.g., Watson and Carter, 2006; Kothari, 2005; or Kendall and Fishman, 1996). These securities are usually fixed income instruments—ABS—which are then sold to investors where the principal and interest depend on the cash flows produced by the pool of underlying financial assets. Figure 24.1 shows a stylized version of this process.

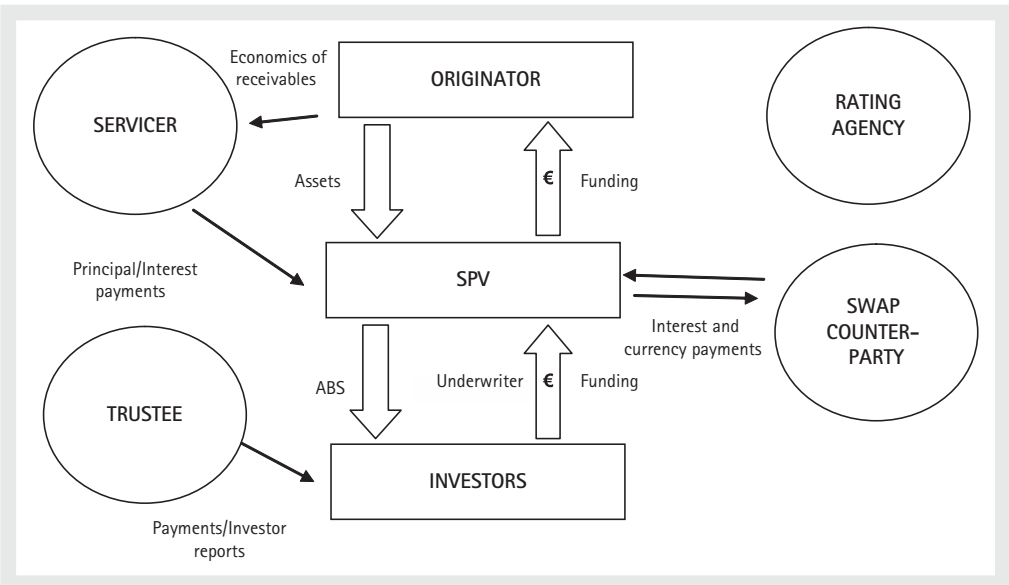

Fig. 24.1. The securitization process

Source: Amended version from *European Securitisation: A Resource Guide*, European Securitisation Forum (2006).

The SPV usually acquires the underlying assets from the originator in what is known as a 'true sale'. For investors in securitization instruments this process helps to guarantee the separation or 'remoteness' of the expected cash flows of the underlying assets from the solvency of the originator. The SPV usually does not have any other function apart from issuing the securities and owning the assets underlying these securities, which reduces incentives for another party to place the SPV into insolvency (see Souleles and Gorton, 2006). The cash received from the investors who purchase the securities issued by the SPV is then passed back on to the originator via the SPV. The SPV also appoints a servicer to collect interest and principal payments on the underlying loans. Two other important parties to the transaction are the swap counterparty (normally involved to hedge the interest rate and currency risks on the pool), and the trustee who ensures that money is transferred from the servicer to the SPV and that investors are paid in accordance with the promised priority. Despite the seeming complexity of the securitization process, the key underlying concept is that if the originator goes bankrupt, there is no recourse to the collateral held by the SPV and the servicer ensures that payments on the collateral continue to be made and investors still receive their interest and principal. The credit quality of the securities issued by the SPV is thus delinked from the solvency of the originator—for example, the bank. The recent credit turmoil, however, has exposed that although the SPVs are separate legal entities from the financial intermediaries that sponsor them, financial intermediaries often have exposures to them from liquidity enhancements and various forms of retained interest (see Gorton and Souleles, 2006 and below).

Most segments of the markets for securitized products have consisted of rated instruments. Hence, rating-agency analysis of ABS or MBS is a crucial part of banks' sales of ABS to investors (a more detailed discussion of valuation issues and ratings is given below). Although the credit quality of individual loans in the underlying pool of assets may be rather low, the credit quality (and therefore the rating) of the overall portfolio can be boosted substantially by pooling the portfolio of credit-risky assets. This credit risk of the portfolio could also be enhanced by using a variety of credit enhancement techniques such as third-party guarantees, overcollateralization, and 'excess spread'. Regarding the latter, excess spreads are credit enhancement technique for originators that represents the 'cash left over after other expenses' for the purpose of credit enhancement. Consequently, excess spread can also be understood as the inherent rate of return in the securitized portfolio over (i) the expenses of the transaction; (ii) the senior servicing fees; and (iii) the rate of return offered to investors (see Kothari, 2006). A key feature of securitization is the slicing of the liabilities of the SPV into different 'tranches' (see DeMarzo, 2005). This tranching is also helpful to cater for investor preferences in terms of maturity or rating.

In the simplest transaction, the securities issued by the SPV would be broken down into three 'tranches' with quite different risk–return profile: the senior

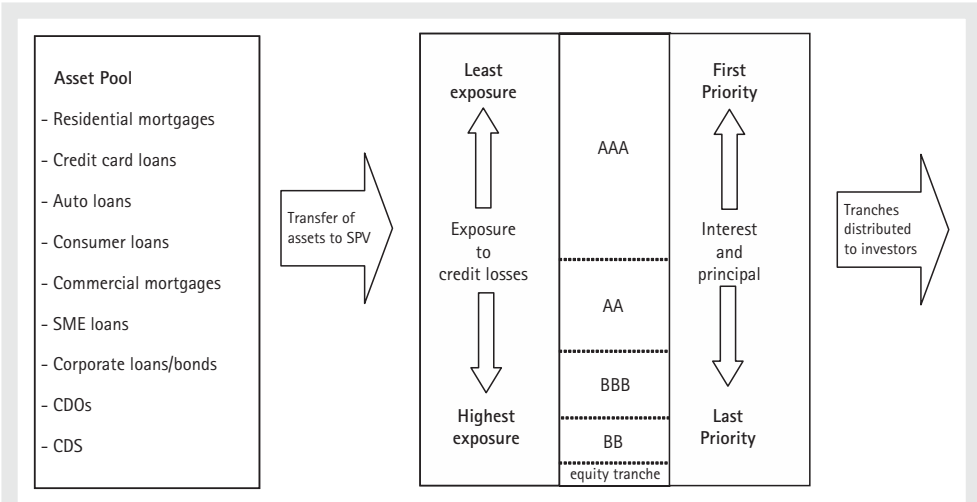

Fig. 24.2. Capital structure and prioritization

tranche, the mezzanine tranche (which are both rated), and the unrated equity tranche (see Figure 24.2). In practice, the number of tranches is normally much higher than three and the senior tranche can be broken into further 'sub-tranches', which often have the same credit rating, but different maturity dates. All tranches are backed by the same pool of credit-risky assets but, if some of the underlying asset defaults, there is a 'cascade' of payments such that the equity tranche is the first to suffer losses, followed by the mezzanine tranche, and lastly the senior tranche. In parallel, all payouts coming from the pool first go to the holders of the senior notes, then the mezzanine tranche and only then would any residual payouts be transferred to equity holders. Thus, the tranche structure in many ways resembles the capital structure of a firm.

In order to signal the quality of the securitized assets and align its interests with those of investors, the originator of the assets may retain part of the equity tranche on its balance sheet. As it would bear most of the risks, the originator would in principle maintain a stronger incentive to continue monitoring the credit quality of the underlying assets (see Innes, 1990; DeMarzo and Duffie, 1999; and DeMarzo, 2005). In recent years, however, the equity tranches have often been sold off (e.g., to hedge funds).

Traditionally, the majority of securitized assets in ABS pools have been large numbers of small-sized, relatively homogenous consumer-related assets, such as prime residential mortgage loans. These assets are particularly appropriate for securitization because the information asymmetries (or the different degrees of knowledge) between originating banks and outside investors regarding the quality

of the underlying claims are usually rather low. Furthermore, pooling large amounts of homogenous, small assets helps to reduce idiosyncratic risks—that is, risks related to individual underlying assets. A prominent example is the securitization of mortgage loans (Lehnert, 2008). Similar structures include the securitization of student loans, credit card payments, car loans, and other securities backed by a large number of small (also termed 'granular') assets. At the same time, the underlying portfolio remains subject to macroeconomic risks including, for instance, declines in housing prices. Those systematic risks can have a strong impact on the value of the securities as illustrated by recent developments in US subprime mortgage-backed securities.

Another major source of market funding for banks, which shares similarities with MBS, is the covered bond market. This market originated in late eighteenth century in Denmark and Germany as a mortgage funding mechanism (Golin, 2006) and, similar to securitization, it is a funding technique that creates negotiable securities by pooling less liquid assets (Smallman, 2006). The assets included in the pool are recorded and registered so that they could be easily identified. In case of insolvency of the issuing bank, covered bond investors have a preferred claim on the assets of the pool.

Despite the similarities, there are significant differences between covered bonds and MBS. A major difference is that covered bonds offer a dual nature of protection as payment is backed by both the originator and the pool of underlying collateral. In addition, the assets supporting covered bonds remain on the originator's balance sheet rather than in a SPV. Another significant difference is that in the case of covered bonds eligible assets are specifically defined by law and are usually substitutable, so the asset mix of the pool could vary over time (Cross, 2004; and Packer, Stever, and Upper, 2007). Covered bonds generally attract a different investor base from ABS because in many countries they have the additional protection of a special legal framework and have a higher level of market liquidity. ABS, on the other hand, normally have amortizing structures, such that the principal is paid back gradually over time and the maturity date of the security is often not known in advance.

The market for MBS has traditionally been considered as a relatively stable market. It was, however, a segment of this market where the initial focus of the turmoil in credit markets in 2007 took place. Indeed, the credit crisis started to be visible on the secondary market for subprime mortgages in the US, but it quickly expanded to the broader mortgage markets and to credit transfer markets in other countries (Lehnert, 2008, see also below). In particular, the problems started in the so-called 'subprime' mortgages market in which borrowers with poor credit histories were given very high loan-to-value mortgages. While significant macroeconomic factors—including the search for yield in credit markets—seem to have been at work contributing to problems on the subprime mortgage markets (Shiller, 2008), there were a number of microeconomic frictions in mortgage markets contributing to the

turmoil. Based on economics of information, Ashcraft and Schuermann (2008) focus on the seven key informational frictions which arise in the securitization market. Foremost among these frictions are the complexity of mortgage products offered to subprime borrowers, which were subject to misunderstanding. Other major frictions include the lower incentives from managers and arrangers to conduct their own due diligence and excessive incentives from originators to lower their lending standards. In addition, credit ratings also provided ratings to subprime MBS with considerable errors (see also Brunnermeier, 2008).

CDOs and CDSs

Over the last few years, there has been a significant trend toward the creation of instruments backed by fewer but larger and more heterogeneous assets than in the traditional 'granular' securitization described above. The assets backing the new securities—CDOs—could include high-yield bonds, leveraged loans, or mezzanine tranches of other ABS transactions and they often combine some of the techniques of traditional securitization with recent innovations in credit risk management. CDOs aim to create value by attracting liquidity toward credit risk in asset classes that, on their own, would be too illiquid or too complex for some investors to consider.

As well as using the 'true sale' cash method that is characteristic of traditional securitization, banks arranging CDO transactions often use CDSs to transfer the credit risk on the underlying pool of assets, which is often termed 'synthetic securitization'. CDSs are the most commonly traded credit derivatives. They transfer the risk that a certain individual entity defaults from the 'protection buyer' to the 'protection seller' in exchange for the payment of a premium. Theoretically, the CDS premium equals the spread over LIBOR on a par-floating rate note (Duffie, 1999).

In synthetic securitizations, the transactions are highly flexible in terms of the asset mix and risk–return characteristics, enabling investors to choose 'tailor-made' CDOs to suit their needs. The underlying assets remain on the balance sheet of the originator or arranger, while the SPV holds a pool of CDSs that reference the assets.

In a synthetic securitization, the CDS generates a premium payment from the originator or arranger to the SPV. In the event that any of the underlying assets default, the SPV would be responsible for any losses. On the liability side, the SPV still issues fixed income securities to investors which can either be 'funded' or 'unfunded'. In funded synthetic securitizations, investors pay for the notes in cash which is then invested by the SPV in high-quality assets such as government bonds.

In unfunded transactions, the investors do not put any cash upfront, which means that the arranging bank risks the investor failing to provide compensation if the underlying assets default. Most synthetic transactions tend to be partially funded, with the super senior tranche being unfunded, and the other senior and subordinated tranches being funded.

CDOs have been constructed with a variety of assets:

- Corporate bonds = collateralized bond obligations (CBOs)
- Corporate/leveraged loans = collateralized loan obligations (CLOs)
- BBB ABS = mezzanine CDO
- Tranches of other CDOs = CDO squared.

A crucial distinction is between 'static' and 'dynamic' CDOs. In the latter case, the manager sells or buys assets depending on market conditions. Another usual distinction is between 'arbitrage' CDOs and 'balance sheet' CDOs. In the former, managers aim to profit from expected differential between assets and liabilities in the SPV while balance sheet deals are driven by regulatory or economic capital management considerations.

Other instruments

Other important instruments include collateralized mortgage obligations (CMOs) and asset-backed commercial paper (ABCP). The former, which has seen particular use in the US, is more oriented toward the allocation of the prepayment risk in mortgages, in contrast to CDOs, which mainly allocate credit risk in the collateral pool.

ABCPs show a number of similarities to 'traditional' ABS as they use a funding structure to issue commercial paper. Typically an ABCP uses short-term debt (with maturities starting from one day to several months) to finance a pool of credit assets such as trade receivables, corporate loans, mortgage loans, CDO tranches, or other credit assets sourced from the market, such as US subprime mortgages. As the assets in the collateral pool can have maturities of several years whereas the ABCP's liabilities are only short term, such structures typically have a sizable maturity mismatch. In order to achieve high ratings by major rating agencies, the risks arising from the mismatch in the maturities was typically mitigated by liquidity guarantees provided by the sponsoring banks. In August 2007, uncertainty and a lack of markets essential for valuing structured credit products meant that liquidity in the ABCP market dried up, forcing sponsoring banks to honor committed liquidity lines. Given the size of the ABCP market (which is estimated at around $1.4 trillion, see Fitch, 2007) and market-wide credit concerns, this was one of the main amplifying factors of the repricing of credit risk.

Considerations on instruments in light of the credit turmoil

As indicated and partly due to misaligned incentives, there has been a dramatic growth in the complexity of securitization instruments. In this respect a number of authors have emphasized the value of 'standard securities' for the design of securities—namely, those securities for which investors have overcome the fixed cost of understanding the security design (see Gale, 1992). Market participants (American Securitization Forum, et al., 2008), also suggest that there has been an excessive broadening of securitization toward products significantly more complex than their earlier counterparts that are inherently difficult to understand. For instance, some of the ABS or CDOs issued before the turmoil frequently were themselves backed by structured securities. This resulted in so-called 'two-layer' or 'double-leveraged' securitizations in which structured products are used to fund other structured products. These products are, however, extremely difficult to value in normal times, let alone in periods of turmoil and they exceeded the analytical capabilities of even the most sophisticated investors (see Duffie, 2007).[3] After the start of the recent credit turmoil there has been a re-evaluation on the ability of agents to model credit risk, particularly for the more complex products. As a result, in the few securitization securities issued after the start of the turmoil the primary issuance of complex products also almost disappeared and there has been a return to simplicity, or 'back to basis' in terms of the characteristics of the securitization products, which is expected to continue in the near future (Aluwalia, Doctor, and Davies, 2008; and Wheeler et al., 2008).

Trading of CDOs

For a large fraction of CDOs there is no active trading and valuation needs to rely on model estimates rather than on market prices. This segment of the CDO market mostly consists of tailor-made instruments and it has recently been the source of sizable losses for many market participants. These 'bespoke' securities are frequently sold in private transactions where an institutional investor (e.g., an insurer) can choose the underlying credit portfolio or the structure of cash flows. The specific features in these transactions limit the development of an active secondary market and investors tend to hold these securities until maturity.[4]

[3] See below and Duffie (2007).

[4] Huddart and Picone (2007) describe how banks use CDS index tranche data to price synthetic CDOs.

The launch of harmonized CDS indices provided a major step in the evolution of the credit risk transfer market. In June 2004, a new family of indices was introduced—namely, iTraxx in Europe and Asia and CDX in North America. This harmonization has led to generally accepted benchmarks for the credit market, therefore increasing market transparency and market liquidity. The indices are divided into several subgroups, ranging from sector categories to a high-yield segment. In the investment-grade corporate segment, the indices contain the equally weighted CDS premiums of the 125 most liquid firms. Selection of index constituents is based on a semi-annual poll of the main CDS dealers, which then leads to an update of the index composition in March and September of each year (JPMorgan, 2006). The CDS premium on the index represents the price of credit protection on the entire pool of firms—that is, a portfolio credit default swap covering all 125 firms in the index.

Index-based CDOs, also known as CDS index tranches, can be seen as the 'tip of the iceberg' of the CDO market segment. Compared to many other credit instruments, trading in CDS index tranches is quite active: in 2006, trading in CDS index tranches amounted to $1,736 billion (Bank for International Settlements, 2007).

Given the iTraxx/CDX index composition, the corresponding CDO structure comprises instruments with varying degrees of exposure to the joint-loss distribution of the 125 underlying firms (Calamaro, et al., 2004). As in bespoke CDOs, the tranches provide claims to the cash flows of the iTraxx CDS portfolio and serve as protection for a certain range of defaults in the portfolio. As in all securitization structures, the equity tranche serves as the first level of protection against any defaults among the firms in the index. The following levels of default protection are provided by mezzanine and by senior tranches, where investors' exposure to default risk in the portfolio is smaller than in the equity tranche. Specifically, the six iTraxx main index tranches are equity (range from 0 percent to 3 percent of the joint loss distribution), low mezzanine (3 percent to 6 percent), mid-mezzanine (6 percent to 9 percent), high mezzanine (9 percent to 12 percent), senior (12 percent to 22 percent), and super senior (22 percent to 100 percent). Tranche trading takes place in the OTC market among banks and brokers. Because the instruments are constructed as synthetic single-tranche CDOs investors can buy or sell all tranches individually.

Market pricing of structured finance instruments

Given the active trading in CDS index tranches, their prices provide a unique picture of events after the subprime turmoil started. Two snapshots of the iTraxx tranche premiums for 29 January 2008 and 23 January 2007 are shown in Table 24.1. As might be expected, there are large differences in individual tranche premiums

Table 24.1. Tranche premiums for iTraxx Europe Main 5Y

Instrument	Loss segment %	Rating	Premium (23 Jan. 2007)	Premium (29 Jan. 2008)	Premium (27 Nov. 2008)
CDS index	0–100	A-BBB	23	70.5	170
Equity	0–3	NA	750	1,243.8	1880
Junior mezzanine	3–6	BBB	40	294	1058.5
Mezzanine	6–9	AAA	12	188	545.5
Senior 1	9–12	AAA	6	117.5	299
Senior 2	12–22	AAA	2.25	59.5	114.5
Super senior	22–100	AAA	0.95	20	62.25

Notes: Table 24.1 reports the CDS premia for the iTraxx Europe Main five-year investment grade index and the corresponding tranches. The rating estimates are taken from Calamaro, et al. (2004).

Source: JPMorgan.

due to the differences in their inherent sensitivity to portfolio credit risk. The tranche providing exposure to the 12 percent to 22 percent segment of the loss distribution paid 59.5 basis points annually on 29 January 2008; the 9–12 percent tranche paid 117 basis points; and the equity tranche 1,243 basis points. Thus, for taking on the first loss piece of the capital structure of the default insurance for the iTraxx portfolio the equity holder would be compensated with an expected annual payment of around 12.5 percent of his notional amount.

Another perspective on the capital structure is that the CDS index portfolio with an annual premium of around 70 basis points generates six new instruments, with premiums ranging from 19.5 basis points (22 percent to 100 percent tranche) to 1,243 basis points (0 percent to 3 percent tranche). This variety of payoffs illustrates how CDOs extend the range of available fixed income products by offering a broad range of risk–return profiles. However, the new instruments also have risk profiles which differ from those of many other instruments. In particular, senior tranches are exposed to sizable 'tail risk'—that is, the risk of very infrequent but catastrophic losses. As Coval, et al. (2007) show, tail risk is a significant factor in the theoretical valuation of CDX tranches already before the start of the credit market turmoil. Coval, et al. (2007) also argue that tranche investors were not aware of the extent of their exposure to tail risk.

After credit traders started their reassessment of the pricing of credit risk in the summer of 2007, investment grade premiums jumped upwards over a short period of time, leading to large mark-to-market losses. All tranche premiums widened significantly, although the degree of change differed across the capital structure. Table 24.1 shows that from 23 January 2007 to 29 January 2008 the equity tranche premium rose from 750 basis points to 1,243 basis points, whereas the premium on

the 12 percent to 22 percent tranche rose spectacularly from 2.25 basis points to around 60 basis points. A similarly sharp increase is also observed for the 22 percent to 100 percent tranche where the premium went up from around 1 basis point to around 20 basis points. This latter case shows the intensity of the repricing of the super-senior tranches, which were perceived to be almost free of default risk before August 2007. Furthermore, the premium of 1 basis point for the 22 percent to 100 percent tranche also explains the popularity of 'leveraged super senior' trading strategies where high expected returns were not generated by investing in risky assets but rather by taking a supposedly low-risk tranche and leveraging it up to obtain higher returns. This repricing gained even more ferocity from January 2008 to November 2008, driving, for example, the iTraxx Main index up to a level of 170 basis points.

Overall, the movements in tranche premiums imply that tranche investors became seriously concerned about losses hitting even the higher components of the capital structure of the iTraxx index tranches. Hence, the pattern of price changes in the less risky parts of the CDO capital structure over the last year can be interpreted as a reassessment of the weight of large, low-probability loss events.

In addition to corporate credit risk, there is also a market for trading US subprime ABS. The ABX.HE indices, which are based on CDSs written on US home equity loan (HEL) MBS, track the price of credit default insurance on a basket of such deals. Since the start of the recent credit crisis in the summer of 2007, the ABX index family provided a widely followed 'barometer' of the collapsing valuations in the US subprime mortgage market, which have been at the core of observed credit market developments. In addition, and despite some shortcomings, ABX price information appears to have been widely used by banks and other investors as a tool for hedging and for gauging valuation effects on subprime mortgage portfolios more generally (see Fender and Scheicher, 2008, for more details).

The ABX family of indices, which started trading on 19 January 2006, consists of a series of equally weighted, static portfolios of CDSs referencing twenty subprime MBS transactions. These contracts, which allow investors to buy and sell protection against the default risk of subprime mortgages, had seen particularly strong growth due to their inclusion in synthetic collateralized debt obligations. The mechanics of the ABX indices, which are offered for trading by a consortium of major credit derivatives dealers, are determined by vintage and credit rating considerations. New on-the-run ABX series are introduced every six months, and each of these index vintages references twenty completely new subprime MBS deals issued during a six-month period prior to index initiation. Trade documentation excludes any form of physical settlement, thus decoupling ABX trading from the availability of the underlying cash instruments. This has aided market development, supporting the adoption of ABX index contracts as a tool for trading and hedging.

Each index vintage consists of five sub-indices, each referencing exposures to the same twenty underlying subprime mortgage securitizations, though at different levels of the liability structure. The ABX 06-1 AAA index, for example, represents tranches with an original rating of AAA from a pool of MBS originated in the latter half of 2005. The other sub-indices, in turn, are backed by tranches of the same securitizations at the AA, A, BBB, and BBB– levels of credit quality. Underlying MBS are selected on the basis of set criteria, targeting large and liquid structures with at least $500 million of deal size at issuance. Once created, index composition remains static, implying that underlying credit quality can migrate to ratings that are lower than indicated by the index name. Trading is conducted in price terms, where prices are quoted in percent of par for each individual index of a given vintage.

The evolution of ABX prices clearly shows the steep decline in prices since June 2007, following an initial price correction early in 2007. Figure 24.3 shows the time series for the three vintages, 2006-1, 2006-2, and 2007-1. Tranches AAA were quoted close to par in June 2007, whereas they were quoted at around 93, 87, and 75, respectively, at the end of December 2007. By the end of November 2008, valuations had deteriorated further, illustrating how the market differentiates between the vintages. Hence, the 2007-1 AAA tranche (which had been downgraded to a lower rating by then) was now priced only at 35, implying a loss of more than 60 percent on the notional.

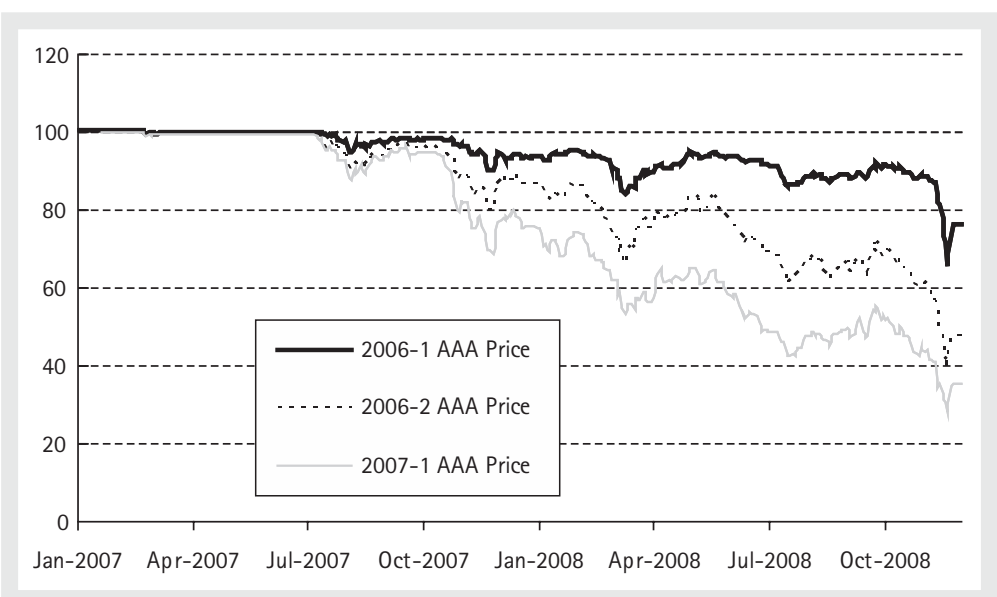

Fig. 24.3. Prices of AAA of ABX index by year vintage

Source: Bloomberg.

Valuation of ABS and CDOs

Even before the credit crisis, market participants faced sizable challenges in the valuation of their CDO positions (a general discussion of the credit crisis is offered by Brunnermeier, 2008). As Duffie argues, 'even specialists in collateralized debt obligations are currently ill equipped to measure the risks and fair valuation of tranches that are sensitive to default correlation' (2007: 4).

In particular, two issues made the CDO valuation more complex than the pricing of many other financial instruments. First, for most CDOs there is no active trading. Typically, an investment bank sells the tranches in private transactions to an insurance firm, hedge fund, or pension fund. As these transactions represent 'tailor-made' instruments, investors usually hold these securities in their books until maturity, making secondary trading quite illiquid.

Second, the theoretical valuation of CDOs is particularly complex as it requires accurate and up-to-date estimation of the co-movement of defaults among the entities in the credit portfolio backing the CDO. Tranche premiums are very sensitive to the default correlation between the firms in the portfolio because this correlation directly influences the distribution of risk in the capital structure. In particular, tranche premiums depend on the joint loss distribution of the underlying portfolio and, given all other parameters, the default correlation determines the shape of this distribution. As default correlation changes, the corresponding movement in the shape of the joint loss distribution is directly transmitted to the relative allocation of portfolio credit risk between equity, mezzanine, and senior tranches. However, estimation of the credit risk correlations poses significant challenges both from a data availability as well as from a modeling perspectives (see Duffie and Garleanu, 2001; or Gibson, 2004). For example, the pricing of a typical CDO based on 100 corporate loans would require estimation of the default co-movement of 100 firms.

Among market participants, CDO valuation frequently relies on the asymptotic single-factor model of credit risk (see Andersen and Sidenius, 2006; or Isla and Willemann, 2007). The single-factor credit portfolio model represents a parsimonious extension of the univariate Merton (1974) model to a multivariate context. In this approach, firms can default due to deterioration in the systematic factor or due to idiosyncratic—that is, firm-specific shocks. The correlation of a firm's asset value with the systematic factor determines the weight of the systematic and idiosyncratic components.

The empirical papers available so far focus on the performance of CDO valuation models for pricing index tranches. In most cases, their sample periods do not capture the repricing since summer 2007. Longstaff and Rajan (2008) find that a three-factor portfolio credit model explains virtually all of the time series and cross-sectional variation in CDX tranche premiums. Bhansali, et al. (2008) use a

more-simplified specification of the same model to study the turmoil period. They find that the subprime turmoil has caused more than twice the level of systemic risk of the May 2005 downgrade of GM and Ford. Tarashev and Zhu (2007) document a large correlation risk premium in CDX tranche prices. Coval, et al. (2007) apply fundamental asset-pricing theory to price CDX tranches. Feldhuetter (2007) implements intensity-based models, finding that pricing performance differs across CDX tranches while Eckner (2007) and Azizpour and Giesecke (2008) decompose the risks priced in CDX tranches.

Scheicher (2008) finds that pricing of CDX and iTraxx tranches differs although the specifications of the two contracts are very similar. Since July 2007, tranche investors appear to have repriced CDX contracts to a larger extent than iTraxx contracts. Credit risk and liquidity factors are priced in almost all tranches with liquidity risk playing a larger role since the start of the turmoil.

Owing to the weaknesses of mark-to-market and mark-to-model valuations, many investors overly relied on rating agencies for their risk assessment. From the beginning, CDOs and ABS have mostly been 'rated' instruments. Ratings frameworks are based on the likelihood of default or the expected loss. Credit ratings can be seen as providing an unconditional risk assessment—that is, a 'cycle-neutral' or 'through-the-cycle' view. A comprehensive description of a typical ABS valuation model as it is used by rating agencies is provided in Ashcraft and Schuermann, 2008).

In most cases, the rating process can be split into two steps: (1) estimation of a loss distribution, and (2) simulation of the cash flows. Given this loss distribution, the required amount of credit enhancement (such as the size of the non-rated first-loss piece) necessary for a tranche to obtain a certain credit rating can be calculated. Credit enhancement, then, is simply the amount of loss on underlying collateral that can be absorbed before a certain tranche with, for example, an investment grade rating has to take on losses from the pool.

The mean of the loss distribution is measured through the construction of a baseline frequency of foreclosure and loss severity for each loan that depends on the characteristics of the loan and local area economic conditions. The distribution of losses is calculated by estimating the sensitivity of losses to local area economic conditions for each mortgage loan, and then simulating future paths of local area economic conditions. Based on this first step, a Monte Carlo exercise is conducted. The second part of the rating process involves simulating the cash flows of the pool in order to determine how much credit excess spread will receive toward meeting the required credit enhancement.

A crucial restriction of rating agencies' risk measures is that they only focus on some components of overall risk. In particular, their focus on expected loss means that the tail of the loss distribution is ignored. Furthermore, the risks from sharp market movements or from declining market liquidity are not captured (Fender and Mitchell, 2005). The instruments are also influenced by the conflicts of interest,

which agencies face in their work. Furthermore, models by rating agencies are naturally also exposed to 'model risk'.

As the drawbacks of the rating agency models became widely known, investors lost confidence in CDO valuations in general. Eventually, investors' attempts to reduce their CDO exposures brought market activity to an almost complete standstill.

Developments in securitization volumes

In recent years, and prior to the 2007–8 credit crisis, the growth in securitization products has been exponential and expanded also outside the US, recording strong growth rates in Asia and Europe (see European Central Bank, 2008a; and Lejot, Arner, and Schou-Zibell, 2008). Figure 24.4 shows how securitization grew in Europe and the US from around $550 billion in 2000 to more than $2,800 billion in 2006. On the synthetic side, figures from the Bank for International Settlements (2008) showed that the notional amounts of CDS expanded tremendously in

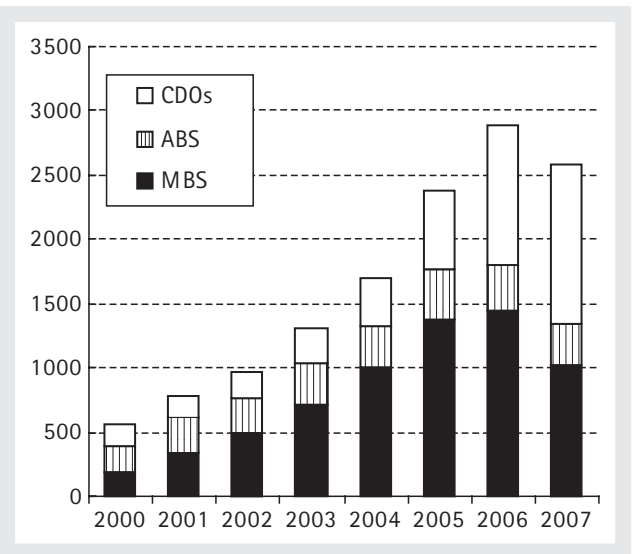

Fig. 24.4. European and US securitization activity (in billions of US dollars)

Notes: CDOs are collateralized debt obligations; ABS, asset-backed securities including auto, credit card, etc. and excluding MBS; and MBS which are mortgage-backed securities, but exclude US agency MBS.

Source: International Monetary Fund, JPMorgan Chase and European Securitization Forum.

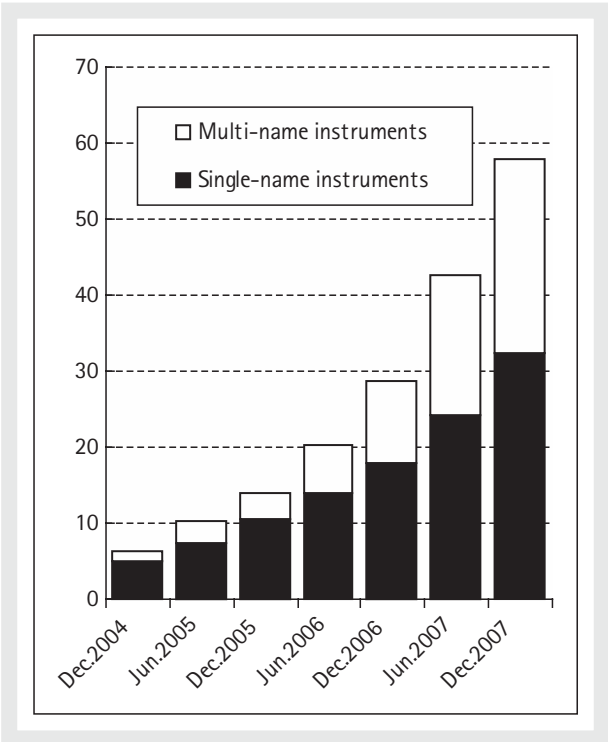

Fig. 24.5. Credit default swaps globally (notional amounts in trillions of US dollars)

Source: JPMorgan Chase, Bank for International Settlements.

recent years and reached around $58 trillion at the end of December 2007 (Figure 24.5).

The increased demand from investors, financial globalization, and technological and financial innovation seem all to have contributed to this growth in securitization activity. First, the demand for ABS has grown rapidly from institutional investors more willing and able to invest in credit risk. Second, technological advancements have been instrumental for the development of securitization globally via the dramatic improvements in storage, processing, and pricing of financial data. Third, globalization and regional financial integration have given an additional impulse to these global trends. For instance, in Europe, the introduction of the euro has provided a strong impulse to the securitization market in Europe (European Central Bank, 2008). The disappearance of exchange rate risk among euro area countries, the increase in financial integration and a more marked-based financial system all contributed to enhancing the liquidity and size of the securitization market. At the same time, the securitization activity in the euro area remains relatively small when compared to the US or the UK (Figure 24.6).

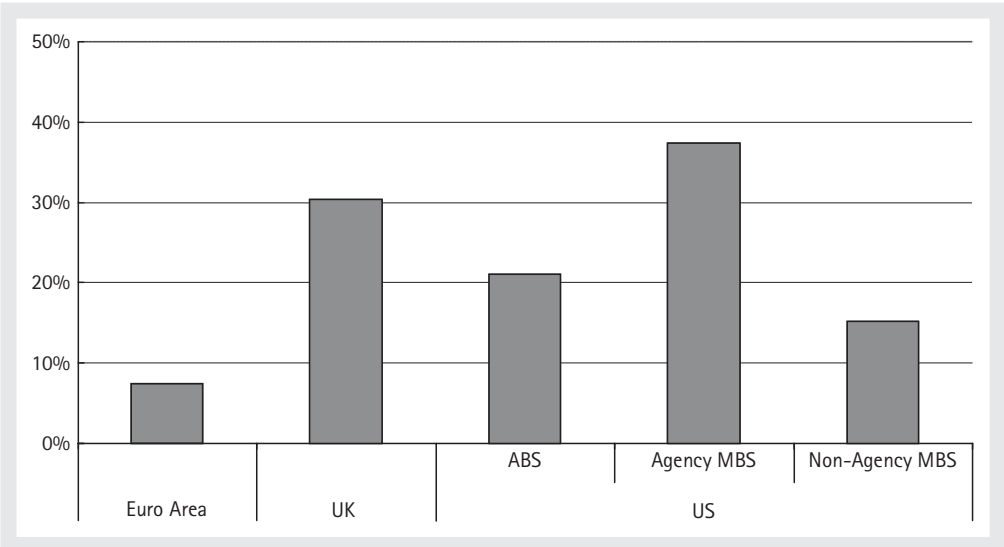

Fig. 24.6. Outstanding values of securitization by geographical area (as a percentage of GDP in 2008: QIII)

Source: European Securitization Forum, Bond Market Association, Eurostat, and International Monetary Fund.

The recent credit crisis is having a massive negative impact on the securitization markets with a large decline in activity in 2008 (see Joint Forum, 2008). In this respect, while securitization activity in primary markets has remained robust in some countries, most of this securitization seems to be retained within originators' balance sheets. In other words, the public market for securitization has been very slim and almost ground to a halt in 2008—as indicated, for instance, by issuance developments in the US (Figure 24.7). In terms of asset classes, some markets such as credit card and consumer loans have shown some signs of activity, particularly in the US, but have also been also dramatically affected by the credit crisis. In the secondary market, also, very illiquid trading conditions have been lingering on after the turmoil started. This is likely to be also the case in the near future owing to a dislocated investor base, which recently experienced very heavy losses, the high level of uncertainty, and an excess of pre-turmoil supply (Aluwalia, Doctor, and Davies, 2008; and Wheeler, et al., 2008).

Owing to its potential benefits for originators and investors, a more active securitization market is likely to reappear. The comeback of a robust securitization market is expected, however, in a very different form from the pre-turmoil period—a strong reduction in the level of complexity and leverage of the instruments issued is expected, while a higher level of transparency and more aligned incentives are warranted.

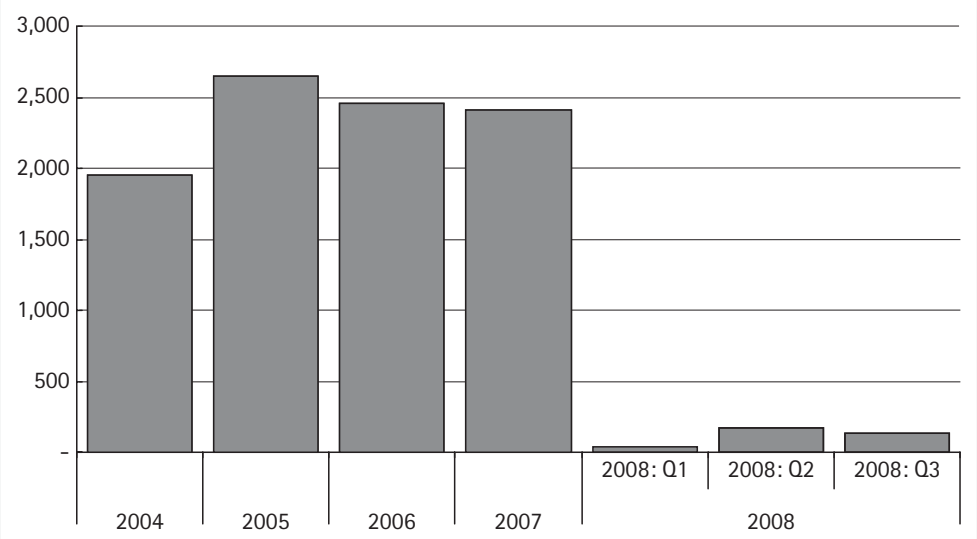

Fig. 24.7. Securitization issuance in the US (in billions of euros)

Source: American and European Securitization Forum.

Motivations for securitization

For originators, a significant motivation for the use of funded securitization is to raise funding and/or reduce financing costs (Fabozzi, David, and Choudhry, 2006). A typical example is a bank with a large portfolio of mortgages loans which can be used as collateral to raise financing, and which could, in turn be used to increase lending. At the same time, securitization allows originators to diversify their sources of financing.

Securitization is also a tool for issuers to tap into additional segments of financial markets, as investors on ABS are often different from covered bonds investors or banks' depositors. This is partly due to the features of the securitization instruments that allow catering to a larger extent than alternative instruments for investors' preferences. In particular, the pooling of assets gives investors diversification benefits which cannot be obtained via outright loan sales or syndication of loans. Likewise, the tranching of securities into instruments with different maturities and credit risk profiles allows for a higher specificity of the assets' profile than covered bonds. The use of credit derivatives also allows investors to diversify into segments of credit markets, which were not feasible in the past. For instance, it has allowed investors to broaden their risk exposure to firms in industries and countries in which it had not been possible in the past owing to market imperfections

that made the trading of loans illiquid (DeMarzo, 2005). As a result, synthetic securitization has played a beneficial role in fostering 'complete markets'. In addition, since the assets can stay on the originator's balance sheet, the legal and administrative costs are significantly lower than those involved in an outright sale.

Via true-sale securitization, originators are able to transfer credit risk off their balance sheet through to the markets instead of having to hold it until maturity. As a result, securitization has often allowed banks to lower their regulatory needs for costly equity capital charges related to loans on the balance sheet, thereby reducing the overall cost of financing (Watson and Carter, 2006). The use of credit derivatives also allowed banks to increase the use of scarce resource 'capital' by means of regulatory risk mitigation whereby banks improve their management of regulatory capital (Fabozzi and Choudhry, 2004).

Securitization could also be used for general risk management purposes, as the capital relief and new funding may be employed to modify the risk profile of an institution—for instance by diversifying the loan portfolio geographically or by sector. This latter motivation is even stronger for the case of credit derivatives (Jeffrey, 2006). They allow banks to optimize the credit risk portfolio management for banks and reduce the degree of concentration. That is, through the use of credit derivatives banks can apply portfolio-optimization techniques to their loan books and to target a certain credit risk profile with relatively low transaction costs (Duffie, 2007).

A related but more specific motivation is the use of credit derivatives to manage counterparty risk. This is usually undertaken to reduce the risk of a default of a—normally large or very active—counterparty. At the same time, the hedging vis-à-vis a specific client, allows banks to exploit certain lines of credit on which it is valuable to continue providing credit without increasing excessively their exposure to an individual borrower. This would be the case of a business strategy in which a bank aims to continue exploiting a profitable relationship (in which value is generated via superior information or the cross-selling of other banking products) without incurring excessive counterparty risks.

Implications

Banks' change of business model

Securitization and financial innovation in credit markets have changed dramatically the financial structure and the role of banks therein. One of the main effects of developments in true sale and, in particular, synthetic securitization, is that large amounts of credit which were traditionally illiquid have now become available

outside the banking sector. In a sense, while the origination of loans remains largely local, securitization can make loan-funding global, making it tradable and available to investors. As a result, banks have maintained and, probably, enhanced their role as originators of credit while progressively losing importance as primary holders of illiquid assets.[5]

The development of securitization can be placed in the wider context of the institutionalization of savings (Kothari, 2006). The latter refers to the relative change in saving patterns in which the growth in market products such as money market funds and mutual funds has increased progressively. On the banking side, there has also been a significant growth in market funding as loan demand often outstripped available deposits and banks used securitization as means of funding. In this respect, securitization adds to the process of disintermediation by shifting more assets into capital markets. This process runs in parallel with the increased demand derived from the growth of non-bank institutional investors such as mutual funds, insurance corporations, and pension and sovereign funds (see European Central Bank, 2007).

Securitization and financial stability

In principle, a key consequence of securitization relates to its effect on risk sharing between banks and markets and the implications for systemic risk. Some of the benefits of securitization are as follows: securitization and credit risk market instruments smooth out the risk among many investors as credit risk can be more easily transferred and potentially widely transferred across the financial system. Even if the total risk remains within the banking sector, securitization allows banks to hold less risk simply due to diversification and more tradeability. The transfer of credit risk can produce a more efficient use of banks' capital and a reduction in the cost of raising capital for loan intermediation, leading, in turn, to a lower cost of credit (Duffie, 2007). Furthermore, market pricing provides valuable information for banks as well as investors and public authorities.

At the same time, as the 2007–8 credit turmoil has illustrated, securitization could lead to financial instability if it produces an increase in the risk of the occurrence of banking crises. Wagner (2005) shows that in theory increased portfolio diversification can augment the probability of a liquidity crisis. This is because higher diversification induces banks to reduce the amount of risky assets they hold and the amount of risky assets might increase. In this respect, Morrison (2003) and Duffee and Zhou (2001) develop theoretical models which show that a market for credit derivatives can lead to a reduction of welfare. Building on Allen

[5] Chiesa (2004) shows that credit risk transfer improves efficiency by allowing banks to use their comparative advantage in evaluating credit risk.

and Gale (2004), Allen and Carletti (2006) show that credit risk transfer could produce a reduction of welfare through the creation of contagion in others. Shin (2008) argues that securitization has proven to be deleterious from a financial stability standpoint. This is because it allows banks to overextend their balance sheet (for a given level of bank capital) and lower their credit standards.

From an investor's perspective, under normal conditions securitization increases liquidity in credit markets. It offers more instruments for trading, hedging, and also an improved menu and supply of investments in different credit categories, thereby completing credit markets. At the same time, due to agency problems between fund managers and final investors, credit risk transfer could also produce incentives for banks to take on excessive risks (Rajan, 2005).

Overall, there are clear benefits from securitization. At the same time—as suggested by a number of studies, and more vividly demonstrated by the recent credit crisis—there could be significant financial stability implications derived from the spectacular development of the market for credit risk transfers in recent years. As result of these dangers to financial stability a number of initiatives have recently been undertaken to strengthen weaknesses that gave rise to problems attached to securitization and more broadly to the 'originate and distribute' model. In October 2007, the G7 ministers and Central Bank Governors asked the Financial Stability Forum (FSF) for a set of recommendations to strengthen the financial system in light of the risks posed by the recent turmoil (see Financial Stability Forum, 2008; and Joint Forum, 2008). The FSF identified a number of issues that deserved to be strengthened such as:

- Misaligned incentives along the securitization line. In this respect, there was, for instance, evidence of poor underwriting standards coupled with poor investor diligence. Originators, arrangers, and managers did not provide sufficient information on the quality and performance of underlying assets.
- Lack of transparency about the risks underlying securitized products, including the quality and correlation of the underlying assets.
- Poor management of the risks associated with the securitization business, including liquidity risk and credit lines. There were shortcomings in firms' risk management tools and models, which often severely underestimated default and liquidity risks—for instance, in the subprime segment where macroeconomic scenarios regarding projections of house prices were far too optimistic.

The FSF proposes concrete actions in a number of areas including:

- Strengthened prudential oversight of capital, liquidity, and risk management. Among other proposals, this would include strengthening capital charges for (i) complex structured credit products; (ii) liquidity facilities to off-balance sheet conduits; and (iii) default and event risk in the trading books.
- Enhancing transparency and valuation: more-helpful risk disclosures (including off-balance sheet vehicles and expanded information about securitized products).

- Changes in the role and uses of credit ratings (including an improvement of the rating process and different ratings on structured credit products from those on bonds).

Impact on credit and the transmission mechanism of monetary policy

From a macroeconomic perspective, securitization is bringing about strong changes in credit markets thereby altering loan dynamics.[6] One of the expected consequences of securitization from a macroeconomic perspective is an overall increase in the aggregate supply of loans. This is due to the characteristics of securitization activity that completes credit markets allowing for a larger mobilization of funds. From a microeconomic standpoint, by being able to securitize part of their assets, banks have access to additional funding. The latter can be used, in turn, to grant additional loans. Furthermore, by fully removing loans from their balance sheet, banks have often been able to obtain regulatory capital relief, which could also be used to expand the supply of loans.

Evidence from the US suggests that the expansion in the supply of credit in recent years was partly driven by securitization. By allowing mortgage originators to distribute credit risk, securitization has increased the amount of lending available to riskier borrowers. Controlling for other factors, Mian and Sufi (2008) show that credit growth is higher in those areas experiencing larger increases in securitization activity. Loutskina (2005) also shows that securitization reduces banks' holdings of liquid securities and increases banks' portfolios. While still tentative, these findings seem to be consistent with European evidence whereby banks more active in the securitization market also seem to supply more loans (Altunbas, Gambacorta, and Marqués Ibáñez, 2007).

In addition, securitization may have altered the monitoring function performed by banks (Diamond, 1984; and Holmström and Tirole, 1997). By moving instruments from banks' balance sheets to the markets there could be fewer incentives for financial intermediaries to screen borrowers (Gorton and Pennacchi, 1995). This is consistent with recent evidence from the US suggesting that lending standards declined more in areas with high securitization rates (Keys, et al., 2008; and Dell'Ariccia, Igan, and Laeven, 2008). In the long term, the change in incentives would be expected to lead to higher default rates on bank loans.

These changes in the role of banks are also expected to have a bearing on the transmission mechanism of monetary policy via the so-called credit channel of monetary policy and more particularly on the bank lending channel or 'narrow'

[6] For an overview of the role of credit and the banking sector from a monetary policy perspective, see Stiglitz and Greenwald (2003).

credit channel. Put simply, the 'credit channel' focuses on how financial imperfections in credit markets (such as asymmetries of information between borrowers and lenders) would affect the supply and conditions of credit. In relation to the 'broad' credit channel, the enhanced liquidity and more continuous pricing of credit market products, as well as the parallel move to fair-value accounting standards offered by credit risk transfer instruments, may have accentuated the sensitivity to changes in monetary policy on the external finance premium that borrowers face when trying to raise financing.

At the same time, the advent of structured credit products has provided the markets with a range of new tools to assess the creditworthiness of borrowers. This increase in credit market information may contribute to compressing the overall external finance premium and hence reducing the effectiveness of the broad credit channel. Hence, a priori, the net effect of financial innovation on the balance sheet channel is somewhat ambiguous (European Central Bank, 2008a). The bank lending channel focuses on how financial imperfections *within* the banking system affect the transmission mechanism of monetary policy (Bernanke, 2007). By providing banks with new funding sources more dependent on market conditions and competing credit markets, changes in securitization activity are also likely to affect the transmission mechanism of monetary policy. According to some recent tentative empirical evidence for the US jumbo mortgages market (Loutskina and Strahan, 2006), and for securitization activity in the euro area (Altunbas, Gambacorta, and Marqués Ibáñez, 2007), the increase in securitization is expected to diminish the impact of monetary policy changes on banks' loan supply, although this effect seems to depend on the economic cycle.

In this respect, by making banks more dependent on market funding (and also owing to the opacity of many of the instruments), securitization could tighten the connection between banks' funding and financial markets. As a result, banks' incentives and abilities to lend are therefore expected to depend on financial market conditions to a larger extent than in the past, when banks were overwhelmingly funded via bank deposits. This is mainly because deposits tend to have more stable remuneration and are, by definition, less dependent on financial market conditions than tradable instruments. Overall, under more extreme circumstances, securitization could have a significant impact on the banking sector's ability to grant credit (European Central Bank, 2008b).

Impact on risk taking

Securitization activity is likely to affect banks' incentives toward risk taking, which could, in turn, have significant financial stability implications (Rajan, 2006). The impact of securitization activity on bank risk taking is, however, multifaceted. On the one hand, securitization and credit risk transfer mechanisms allow banks to

shift their risks outside their balance sheet as well as to achieve portfolio and funding diversification more easily (European Central Bank, 2008b). On the other hand, securitization could also lead banks to take on additional risks whether by using the funding obtained from securitization to grant riskier credit or simply by acquiring credit risk more easily on the market.

At the same time, incentives for banks' risk taking might have been changing in recent years owing to a number of factors. Foremost among those factors would be the impact of securitization and other forms of financial innovation. In addition, changes in bank competition owing to de-regulation and prudential re-regulation (such as Basel II), increased pressures from shareholders to create market value—and a greater reliance on market sources of funding have also been important. Overall then, bank risk and securitization considerations need to be carefully modeled when considering the possible effects on the supply of bank loans.

Borio and Zhu (2007) argue that financial innovation in parallel with changes to the capital regulatory framework (Basel II) have enhanced the importance of the perceptions, pricing, and management of risk on the behavior of banks and other financial intermediaries (European Central Bank, 2008b). Rajan (2005) stresses that more market-based pricing exacerbates the incentive structures driving banks and institutional investors which could (under extreme circumstances) lead to excessive risk taking behavior. In light of the recent banking turmoil, securitization is likely to have had an impact on risk taking behavior—at least for some types of banks. At the same time, evidence on the link between securitization and bank risk taking remains tentative. Jiangli, Pritsker, and Raupach (2008) use US data from bank holding companies and find that banks active in the securitization market tend to have lower solvency risk and higher profitability than banks not active in the securitization market. In contrast, Franke and Krahnen (2005) found an increase in banks' risk (measured as their betas) surrounding the announcement of a CDO issue, while Bannier and Hänsel (2007) suggest that securitization should not be taken as a consequence of banks' appetite for risk but mainly as a risk-transferring tool. Shin (2008) argues that securitization actually increases banks' tolerance to risk for a given level of capital. Overall, then, tentative evidence suggest that the impact of securitization activity on bank risk taking does probably exist—but it also seems to be multifaceted, and further research is warranted to shed light on the recent credit crisis.

Conclusions

The spectacular increase in securitization activity and more broadly in credit risk transfer instruments over the last decade has altered radically the functioning of

credit markets. It has transformed the traditional role of banks as providers and monitors of credit. As a result of securitization, large parts of banks' portfolios no longer need to be held until maturity in banks' balance sheets, opening new segments of the credit spectrum to investors. This chapter describes the main building blocks and instruments of the securitization market. It shows how the process toward the commoditization of credit risk has evolved from 'granular'-funded mortgage-backed securities toward more complex structures. The latter are often unfunded and allow market participants to trade credit risk relatively easily, reaching parts of the credit spectrum which were not feasible in the past. Research on the implications of securitization activity, however, remains at its infancy and has not kept track with financial innovation. Existing evidence suggests that securitization activity is expected to alter risk sharing mechanisms within the financial system. It is also having an impact on banks' incentives to lend and manage credit risk. As a result, securitization activity is probably altering both loan dynamics and the impact of interest rate changes on the supply of credit.

References

Acharya, V. and Johnson, T. (2007). 'Insider Trading in Credit Derivatives', *Journal of Financial Economics* 84: 110–41.

Allen, F. and Carletti, E. (2006). 'Credit Risk Transfer and Contagion', *Journal of Monetary Economics*, 53: 89–111.

——and Gale, D. (2004), 'Financial Intermediaries and Markets' *Econometrica*, 72, 1023–61.

—— —— (1994). 'Systemic Risk and Regulation', in M. Carey and R. M. Stulz (eds.), *The Risks of Financial Institutions*, National Bureau of Economic Research Conference Report, 2006, Chicago: University of Chicago Press.

Altunbas, Y., Gambacorta, L., and Marqués Ibáñez, D. (2007). 'Securitisation and the Bank Lending Channel, European Central Bank Working Paper 838.

Aluwalia, R., Doctor, S., and Davies, G. (2008). 'What Is the Future of European Structured Products?', European Credit JPMorgan, 12 November 2008.

Amato, J. and Remolona, E. (2003). 'The Credit Spread Puzzle', *Bank for International Settlements Quarterly Review*, December, 51–63.

American, European, and Australian Securitization Forum and the Securities Industry and Financial Markets Association (SIFMA) (2008). 'Restoring Confidence in the Securitization Markets', Global Joint Initiative to Restore Confidence in the Securitisation Markets, 3 December 2008.

Andersen, L. and Sidenius, J. (2005). 'CDO Pricing with Factor Models: Survey and Comments', *Journal of Credit Risk* 1: 71–88.

Ashcraft, A. and Santos, J. (2007). 'Has the CDS Market Lowered the Cost of Corporate Debt?', Federal Reserve Bank of New York Working Paper.

Ashcraft, A. and Schuermann, T. (2008). 'Understanding the Securitization of Subprime Mortgage Credit', Federal Reserve Bank New York Staff Report 318.

Azizpour, S. and Giesecke, K. (2008). 'Premia for Correlated Default Risk', Working Paper, Stanford University.

Bank for International Settlements (2004). 74th Annual Report.

—— (2007). 77th Annual Report.

Bannier, C. E. and Hänsel, D. N. (2006). 'Determinants of Banks' Engagement in Loan Securitisation', Goethe University Working Paper 171.

Batchvarov, A. (2006). 'How can Securitisation Benefit Your Company', in P. Jeffrey (ed.), *A Practitioners Guide to Securitisation*, London: City and Financial Publishing.

Basel Committee on Banking Supervision (2006). 'Basel II: International convergence of Capital Measurement and Capital Standards: A Revised Framework', Basel: BIS.

Bernanke, B. (2007). 'The Financial Accelerator and the Credit Channel', Speech at the Conference entitled 'The Credit Channel of Monetary Policy in the Twenty-First Century', Federal Bank of Atlanta, Georgia, 15 June 2007.

Bhansali, V., Gingrich, R., and Longstaff, F. (2008). 'Systemic Credit Risk: What Is the Market Telling us?', University of California, Los Angeles (UCLA) Working Paper.

Blanco, R., Brennan, S., and Marsh, I. W. (2005). 'An Empirical Analysis of the Dynamic Relationship between Investment-Grade Bonds and Credit Default Swaps', *Journal of Finance* 60: 2255–81.

Brunnemeier, M. (2008). 'Deciphering the 2007–08 Liquidity and Credit Crunch', *Journal of Economic Perspectives*, 23: 77–100.

Calamaro, J. P., Nassar, T., Thakkar, K., and Tierney, J. (2004). 'Trading Tranched Index Products: The First Steps', Deutsche Bank, Global Markets Research, Quantitative Credit Strategy.

Chen, L., Lesmond, D., and Wei, J. (2007). 'Corporate Yield Spreads and Bond Liquidity', *Journal of Finance* 62: 119–49.

Committee on the Global Financial System (2005). 'The Role of Ratings in Structured Finance: Issues and implications', CGFS Publications 23, January.

Coval, J., Jurek, J., and Stafford, E. (2007). 'Economic Catastrophe Bonds', Harvard Business School Working Paper 07-102.

Cross, G. (2004). 'The German Pfandbrief and European Covered Bond Markets', in F. J. Fabozzi and M. Choudhry (2004), *The Handbook of European Structured Financial Products,* Chichester: Wiley.

Dell'Ariccia, G., Igan, D., and Laeven, L. (2008). 'Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market', International Monetary Fund Working Paper 106.

DeMarzo, P. (2005). 'The Pooling and Tranching of Securities: A Model of Informed Intermediation', *The Review of Financial Studies* 18: 1–35.

—— and Duffie, D. (1999). 'A Liquidity Based Model of Security Design', *Econometrica* 67: 65–99.

Diamond, D. W. (1984). 'Financial Intermediation and Delegated Monitoring', *The Review of Economic Studies* 51: 393–414.

Duffie, D. (2007). 'Innovations in Credit Risk Transfer: Implications for Financial Stability', Stanford University Working Paper.

—— and Garleanu, N. (2001). 'Risk and Valuation of Collateralized Debt Obligations', *Financial Analysts Journal* 57: 41–59.

——and Singleton, K. J. (2003). *Credit Risk: Pricing Measurement and Management*, Princeton, NJ: Princeton University Press.
Duffee, G. and Zhou, C. (2001). 'Credit Derivatives in Banking: Useful Tools for Managing Risk?', *Journal of Monetary Economics* 48: 25–54.
Eckner, A. (2007). 'Risk Premia in Structured Credit Derivatives', Stanford University Working Paper.
European Central Bank (2007). 'Corporate Finance in the Euro Area', Occasional Paper 63.
——(2008a). 'Securitisation in the Euro Area', *European Central Bank Monthly Bulletin*, February.
——(2008b). 'Financial Stability Review', *European Central Bank Monthly Bulletin*, June.
——(2008c). 'The Role of Banks in the Monetary Policy Transmission Mechanism', *European Central Bank Monthly Bulletin*, August.
European Securitisation Forum (2006). 'European Securitisation: A Resource Guide'.
Fabozzi, F. J. and Choudhry, M. (2004). *The Handbook of European Structured Financial Products,* Chichester: Wiley.
——David, H. A., and Choudhry, M. (2006). *Introduction to Structured Financial,* Chichester: Wiley.
Feldhuetter, P. (2007). 'An Empirical Investigation of an Intensity-Based Model for Pricing CDO Tranches', Copenhagen Business School Working Paper.
Fender, I. and Mitchell, J. (2005). 'Structured Finance: Complexity, Risk and the Use of Ratings', *Bank for International Settlements Quarterly Review*, June: 67–79.
——and Scheicher, M. (2008). 'The ABX: How do the Markets Price Subprime Risk', *Bank for International Settlements Quarterly Review*, September: 67–81.
Financial Stability Forum (2008). 'Report of the Financial Stability Forum on Enhancing Market and Institutional Resilience', March.
Fitch Ratings (2007). 'Asset-Backed Commercial Paper and Global Banks: Ten Key Questions', Special Report, September 2007.
Franke, G. and Krahnen, J. P. (2005). 'Default Risk Sharing between Banks and Markets: The Contribution of Collateralized Debt Obligations', Center for Financial Studies Working Paper 2005/06.
Holmström, B. and Tirole, J. (1997). 'Financial Intermediation, Loanable Funds, and the Real sector', *The Quarterly Journal of Economics*, 112: 663–91.
Gale, D. (1992), 'Standard Securities', *Review of Economic Studies*, 59: 731–55.
Gibson, M. (2004). 'Understanding the Risk of Synthetic CDOs', Federal Reserve Board Finance and Economics Discussion Series 2004-36.
Golin, J. (2006). *Covered Bonds beyond Pfandbriefe: Innovations, Investment and Structured Alternatives*, London: Euromoney Books.
Gorton, G. B. and Pennacchi, G. G. (1995). 'Banks and Loan Sales. Marketing Nonmarketable Assets', *Journal of Monetary Economics* 35: 389–411.
——and Souleles, N. (2006). 'Special Purpose Vehicles and Securitization', in R. M. Stulz and M. Carey (eds.), *The Risks of Financial Institutions*, Chicago: University of Chicago Press.
Hirtle, B. (2007). 'Credit Derivatives and Bank Credit Supply', Federal Reserve Bank of New York, Staff Report 276.
Huddart, R. and Picone, L. (2007). 'Synthetic CDOs: Made to Measure', Dresdner Kleinwort Debt Research.

Innes, R. (1990). 'Limited Liability and Incentive Contracting with Ex Ante Choices', *The Journal of Economic Theory* 52: 45–67.

Isla, L. and Willemann, S. (2007). 'Correlation Modelling: From Vanilla to Exotic', Barclays Capital Structured Credit Strategist.

Jeffrey, P. (2006). *A Practicioners' Guide to Securitisation,* London: City and Financial Publishing.

Jiangli, W. and Pritsker, M. (2008). 'The Impacts of Securitization on US Bank Holding Companies', Mimeo.

J.P. Morgan (2006). *Credit Derivatives Handbook,* Corporate Quantitative Research.

Joint Forum (2008). 'Credit Risk Transfer: Developments from 2005 to 2007', Basel Committee on Banking Supervision, International Organization of Securities Commissions, International Organization of Insurance Supervisors, April.

Kendall, L. T. and Fishman, M. (1996). *A Primer on Securitization*, Cambridge, MA: MIT Press.

Keys, B., Mukherjee, T., Seru, A., and Vig, V. (2008). 'Did Securitization Lead to Lax Screening?: Evidence from Subprime Loans 2001–2006', Mimeo.

Kodres, L. (2008). 'Structured Finance: What Is It and How Did It Get It So Large?', International Monetary Fund Global Financial Stability Report, April 2008.

Kothari, V. (2006). *Securitization: The Financial Instrument of the Future*, Chichester: Wiley.

Lehnert, A. (2008). 'Residential Mortgages', in A. Berger, P. Molyneux, and J. Wilson (eds.), *Handbook of Banking,* Oxford: Oxford University Press.

Lejot, P., Arner, D., and Schou-Zibell, L. (2008). 'Securitisation in East Asia', Asian Development Bank Working Paper Series on Regional Economic Integration 12.

Longstaff, F. and Rajan, A. (2008). 'An Empirical Analysis of the Pricing of Collateralized Debt Obligations', *Journal of Finance* 64: 529–63.

Loutskina, E. (2005). 'Does Securitization Affect Bank Lending?: Evidence from Bank Responses to Funding Shocks', European Finance Association 2005 Moscow Meetings.

—— and Strahan, P. E. (2006). 'Securitization and the Declining Impact of Bank Finance on Loan Supply: Evidence from Mortgage Acceptance Rate', National Bureau of Economic Research Working Paper 11983.

Merton, R. (1974). 'On the Pricing of Corporate Debt: The Risk Structure of Interest Rates', *Journal of Finance* 29: 449–70.

Mian, A. and Sufi, A. (2007). 'The Consequences of Mortgage Credit Expansion: Evidence from the 2007 Mortgage Default Crisis', National Bureau of Economic Research Working Paper 13936.

Minton, B., Stulz, R., and Williamson, R. (2006). 'How Much Do Banks Use Credit Derivatives to Reduce Risk?, National Bureau of Economic Research Working Paper 11579.

Morrison, A. (2005). 'Credit Derivatives, Disintermediation and Investment Decisions', *Journal of Business* 78: 621–47.

Packer F., Stever, R., and Upper C. (2007). 'The Covered Bond Market', *Bank for International Settlements Quarterly Review*, September: 43–55.

—— and Zhu, H. (2005). 'Contractual Terms and CDS Pricing', *Bank for International Settlements Quarterly Review*, March: 89–100.

RAJAN, N. (2005). 'Has Financial Development Made the World Riskier?', National Bureau of Economic Research Working Paper 11728.

SCHEICHER, M. (2008). 'How Has CDO Market Pricing Changed during the Turmoil? Evidence from CDS Index Tranches', European Central Bank Working Paper.

SHILLER, R. J. (2008). *The Subprime Solution: How Today's Global Financial Crisis Happened, and What to Do About it*, Princeton, NJ: Princeton University Press.

SHIN, H. S. (2008). 'Securitisation and Financial Stability', Mimeo.

SMALLMAN, J. D. and SALMAN, S. (2006). 'Residential Mortgage-Backed Securities', in R. Watson and J. Carter (eds.), *Asset Securitisation and Synthetic Structures: Innovations in the European Credit Markets*, London: Euromoney Books.

STIGLITZ, J. E. and GREENWALD, B. (2003). *Towards a New Paradigm in Monetary Economics*, Cambridge: Cambridge University Press.

TARASHEV, N. and ZHU, H. (2007). 'The Pricing of Correlated Default Risk: Evidence from the Credit Derivatives Market', Bank for International Settlements Working Paper.

VAN ORDER, R. (2007). 'On the Economics of Securitization: A Framework and Some Lessons from U.S. Experience', Ross School of Business Working Paper 1082.

WAGNER, W. (2007). 'Aggregate Liquidity Shortages, Idiosyncratic Liquidity Smoothing and Banking Regulation', *Journal of Financial Stability* 3: 18–32.

WATSON, R. and CARTER, J. (2006). *Asset Securitisation and Synthetic Structures: Innovations in the European Credit Markets*, London: Euromoney Books.

WHEELER D., KANE M. E., GAON S., and PARULEKAR S. (2008), Does the World Need Securitisation?, Citigroup Securitized Products Strategy, 12 December 2008.

This page intentionally left blank

PART IV

MACROECONOMIC PERSPECTIVES IN BANKING

This page intentionally left blank

CHAPTER 25

SYSTEMIC RISK IN BANKING

AN UPDATE

OLIVIER DE BANDT
PHILIPP HARTMANN
JOSÉ LUIS PEYDRÓ[1]

The Fed will work closely and actively with the Treasury and other authorities to minimize systemic risk.

Ben Bernanke, Chairman of the Federal Reserve (Oct. 2008)[2]

The failure of Lehman Brothers in September triggered an unprecedented deterioration in the degree of confidence in the banking sector which ran the risk of undermining its fundamental function of financial intermediation... Since the onset of the financial turmoil, increased uncertainty, reflected particularly in counterparty risk, has led banks to

[1] This chapter builds on and updates de Bandt and Hartmann (2002). Any views expressed in this paper are the authors' own and do not necessarily reflect those of the European Central Bank, the Banque de France or the Eurosystem.

[2] 'Stabilizing the Financial Markets and the Economy', address at the Economic Club of New York <http://www.america.gov/st/texttrans-english/2008/October/20081016110841eaifas9.330386e-02.html>.

> hoard liquidity. This has resulted in a significant decline of trading in the interbank money markets.
>
> Jean-Claude Trichet, President of the European Central Bank (Dec. 2008)[3]

Introduction

After a long period of relative quiescence, the credit market crisis, which broke out in August 2007 and which by the time of the last revision of this chapter had developed into a severe international financial crisis, has brought back concerns about the stability of national and international financial systems (see Ferguson, et al., 2007). For example, the need to avoid systemic risk to materialize was an important factor in the Fed's decisions to facilitate the takeover of Bear Stearns by J.P. Morgan in March 2008 or to purchase assets of AIG in September 2008. The collapse of Lehman Brothers, in conjunction with a number of other events, seems to have had significant systemic implications. The two quotes above illustrate the policy relevance of assessing systemic risk. Previously, the 1990s had been characterized by a number of crises, such as the Nordic and Japanese banking or the Asian financial crises. While the earlier and the present episodes partly share some common features and partly are characterized by specific factors, 'systemic risk' is now widely accepted as a fundamental underlying concept for the study of such severe financial instabilities and possible policy responses.[4] This applies particularly to the new interest in macro-prudential supervision and regulation, as reflected for example in the establishment of the European Systemic Risk Board in the European Union and the proposed Financial Services Oversight Council in the United States (see High-level Group on Financial Supervision in the EU, 2009, Council of the European Union, 2009, and US Department of the Treasury, 2009).

In this chapter we provide a comprehensive analysis of systemic risk, as the primary ingredient to understand financial crises and as a main rationale for banking regulation, prudential supervision, and crisis management. In a first step we bring together the most important analytical elements of systemic risk and integrate them into a coherent working concept, which could be used as a baseline for prudential and monetary policy decisions to preserve the stability of financial systems. While the 'special' character of banks continues to play an important role,

[3] Address to the European Parliament <http://www.ecb.int/press/key/date/2008/html/sp081208_2.en.html>. Lehman Brothers' failure became the largest bankruptcy in US history, listing liabilities of $613 billion in its filing.

[4] Extensive further discussions, reports or papers on the current crisis are, for example, Acharya and Richardson (2009), Ashcraft and Schuermann (2008), Evanoff et al. (2009), Federal Reserve Bank of Kansas City (2008), Financial Stability Forum (2008), Greenlaw et al. (2008), Institute of International Finance (2008), Senior Supervisors Group (2008), and many financial stability reports of central banks.

systemic risk goes much beyond the traditional view of single banks' vulnerability to depositor runs. At the heart of the concept is the notion of 'contagion', a particularly strong propagation of failures from one institution or system to another. Especially, nowadays, the way in which wholesale interbank, derivatives, and securitization markets function and how risk is shared among various intermediaries can play an important role in the way shocks may propagate, and actually be amplified, through the bank system and the financial system as a whole.

In a second step we review the existing theoretical and empirical literature about systemic risk in the light of the previously developed general concept. This could also help in identifying areas in which future research is needed. By focusing primarily on the quantitative literature we mainly review the work formulating and testing specific hypotheses with the most advanced techniques.

The remainder of the chapter is organized as follows. The second section contains the general conceptual discussion. It provides the framework within which the theoretical and empirical literature will be interpreted in the subsequent parts and briefly discusses its relevance for economic policy. The third section gives an overview of theoretical models explaining systemic risks in banking markets (including payment systems). The fourth section surveys econometric tests and some other quantitative empirical assessments of the various facets of systemic risk described before, focusing particularly on bank contagion but also on joint crises—for example, as caused by systematic shocks or the unravelling of imbalances that have built up over time. The fifth section concludes.

The concept of systemic risk

Systemic risk in a very general sense is by no way a phenomenon limited to economics or the financial system. Maybe the most natural illustration of the concept is in the area of health and epidemic diseases, where widespread contamination with a disease may wipe out a significant portion of a population. In the area of economics, it has been argued that systemic risk is a special feature of financial systems, in particular the banking system. While contamination effects may also occur in other sectors of the economy, the likelihood and severity in financial systems is often regarded as being considerably higher. A full systemic crisis in the financial system may then have strong adverse consequences for the real economy and general economic welfare.

The objective of this section is to provide a framework for the economic analysis of systemic risk and explain the reasons why financial systems can be regarded as being more vulnerable to systemic risk than other parts of economic systems. We also distinguish between pure self-fulfilling systemic events and those that can be regarded as individually rational responses to information. The relevance of systemic risk in banking for public policy is briefly examined.

Systemic events

We briefly clarify a number of concepts, highlighting here the thrust of the arguments developed in our earlier work (de Bandt and Hartmann, 2002), to which the reader is kindly asked to refer for further details. The main distinction is between a 'systemic event' (i) in the 'narrow' sense, where the release of 'bad news' about a bank, or even its failure, leads in a sequential fashion to considerable adverse effects on one or several other banks (for example, their failure); and (ii) in the 'broad' sense in order to include simultaneous adverse effects on a large number of banks as a consequence of severe and widespread ('systematic') shocks or the unravelling of significant imbalances that have built up over time.

A systemic event in the narrow sense is 'strong', if financial institutions actually fail as a consequence of the initial shock, although they have been fundamentally solvent *ex ante*. We denote these strong instances of systemic events in the narrow sense as 'contagion'. Otherwise (that is, if the external effect is less than a failure) we denote a systemic event in the narrow sense as 'weak'. Similarly, systemic events related to systematic shocks or the unravelling of imbalances are strong (weak), if a significant part of the banks simultaneously affected by them (do not) actually fail. Of course, there is a continuum of intermediate types of shocks (e.g., sector-wide or regional) between the theoretical extremes of idiosyncratic and wide systematic shocks.

Based on this terminology a 'systemic crisis' (in the narrow and broad senses) can be defined as a systemic event that affects a considerable number of banks in a strong sense, thereby severely impairing the general functioning of a major part of the financial system, in which case the effectiveness and efficiency of the financial system to direct savings into real investments is compromised. 'Systemic risk' (in the narrow and broad senses) can then be defined as the risk of experiencing systemic events in the strong sense.

One may also distinguish a 'horizontal' view on the concept of systemic risk, in which the focus is limited to events in the financial sector alone (particularly through the bankruptcy of banks) from a 'vertical' view on systemic risk, where the impact of a systemic event on output can proxy the severity of such an event. In many of the papers discussed below real effects play some role. However, in order to keep the scope of the paper manageable we concentrate the discussion on the horizontal dimension.[5]

[5] See Hoggarth and Saporta (2001) for syntheses of the output effects of a large number of bank crisis situations. The relationship between the performance of the real and financial sectors can go in both directions. This raises the issue of causality. Dell'Ariccia, Detragiache, and Rajan (2008) argue that if banking crises reduce real activity, then sectors more dependent on external finance should perform relatively worse during them. Financial crises can, however, also be the expression of an economy taking greater risks in financing real investment, which could be part of a long-term growth strategy. See Ranciere, Tornell, and Westermann (2008).

The 'financial fragility hypothesis'

Why is it then that systemic risk—in particular potential contagion effects—is of special concern in the financial system? There are three interrelated features of financial systems that can provide a basis for this 'financial fragility hypothesis'.

(i) First, the structure of banks' balance sheets matters, as a result of maturity transformation activity. Traditionally, commercial banks take fixed value deposits that can be withdrawn (unconditionally and at fixed value) at very short notice and lend long term (Bryant, 1980; and Diamond and Dybvig, 1983). Moreover, bank loans are difficult to value. In fact, the present credit-market crisis illustrates that the pricing of bank assets remains a challenge and that the expansion of securitization activity in the preceding years may have masked fundamental valuation problems. Therefore, the health of a bank not only depends on its success in picking profitable investment projects for lending but also on the 'confidence' of depositors in the value of the loan book and, importantly, in their confidence that 'other' depositors will not run on the bank. In fact, coordination problems among depositors may cause solvent but illiquid banks to fail (Goldstein and Pauzner, 2005). This 'special' character of banks does not apply to many other financial intermediaries, such as insurance companies, securities houses, and the like. However, if banks and other intermediaries belong to the same financial entity, or the former are exposed to the latter, non-bank intermediaries' problems might still become a source of bank fragility. Obviously, the more depositors are protected through some deposit insurance scheme the less likely it is that confidence crises occur through retail deposits. In fact, during the current global financial crisis, the deposit insurance thresholds were increased worldwide to contain runs on banks. Another important feature of the current crisis is the role that short-term wholesale funding structures play. Most major investments banks, for example, had such funding structures, and as some main assets and hedging instruments became illiquid or turned out to be much less liquid than previously thought, effectively faced substantial maturity mismatches. As market stresses deepened, wholesale financiers became unwilling to roll over the short-term debt, in ways reminiscent of depositor runs. This affected not only investment banks, but also other banks with undiversified funding structures and off-balance sheet vehicles used for funding structured finance investments. Northern Rock, for example, relied to 75 percent on interbank deposits and therefore collapsed when the money market became dysfunctional.[6] Wholesale deposits lead us to the second special feature of financial systems.

[6] Many banks used off-balance sheet vehicles to invest during the growth period of securitization in structured finance products. Funding was short term, as they issued commercial paper backed by their (long-term) assets. Once valuations of structured products were in doubt, investors were not inclined

(ii) There is a complex network of exposures among banks (and potentially some other financial intermediaries). One channel is through exposures in interbank money markets, derivative markets and large-value (wholesale) payment and security settlement systems. As the current financial crisis has again illustrated, malfunctioning interbank markets have immediately systemic effects due to the extensive participation of most major banks in them. The functioning and important risks of wholesale and retail payment and settlement systems tend to be less well known outside the expert community. Policy and market initiatives over the last two decades have significantly improved their safety so that they played no specific role in the present crisis. Badly designed payments and settlement systems, however, imply substantial systemic risk. Another channel of exposures emerges through banks investing in similar or correlated assets.

(iii) The third feature is, more generally, the information and control intensity of financial contracts, which rely on promises and expectations about future payments (e.g. Stiglitz, 1993). For example, the willingness of agents to extend credit depends on their confidence that borrowers reimburse them in the future. When asymmetric information emerges, uncertainty increases, or the credibility of a financial commitment starts to be questioned, market expectations may shift substantially and, in an 'individually rational' way, in short periods of time leading to equally volatile investment and disinvestment decisions. For example, after the failure of Lehman Brothers and other negative events in September 2008, a general loss of confidence emerged in which many banks preferred to hoard funds rather than lend them out. Or, in the summer of 2007, significant doubts emerged about the viability of structured finance products, so that investors—who could not distinguish good from bad products—in general stopped rolling over asset-backed commercial paper financing the investments of off-balance sheet vehicles.

'Efficient' versus 'self-fulfilling' systemic events

General uncertainty and agents' awareness of potential asymmetries of information highlight the role that expectations can play in systemic events. In fact, systemic events driven by expectations can be individually rational but not socially optimal. It is useful to distinguish between three potential causes of narrow systemic events related to asymmetric information and expectations. We illustrate them with the example of uninsured bank deposits, but similar mechanism can also occur for other information-intensive debt contracts.

any longer to roll over this short-term debt and the vehicles had to be taken back on banks' balance sheets to avoid their failure. It is expected that these off-balance sheet vehicles will lose importance after the crisis or even disappear.

(i) If the information about bank losses is released in full, it is individually rational for depositors to withdraw their funds and force those banks into liquidation if bank losses are sufficiently high—that is, if the bank is insolvent. *Ceteris paribus*, such an outcome, which can be denoted as a 'fully revealing' equilibrium, would also be 'efficient', as opposed to a scenario where the bank continues to accumulate losses.

(ii) Suppose that the information about bad loans and interbank exposures is not revealed in full but that depositors only receive imperfect information, a 'noisy' common signal, from some outside source, which from their point of view increases the likelihood of the bank's adverse position. In such a situation it might still be rational for them to try and withdraw their funds early and thereby force the default of those banks. Whether the signal has been 'right' or 'wrong' would determine, *ceteris paribus*, whether this outcome is 'efficient' *ex post* or not. As it is triggered by imperfect information on fundamentals, this type of contagion could be denoted as 'information-based'.

(iii) Signals can also coordinate depositors in their strategies to run or not in one or multiple banks. If the signals are not related to the health of banks and are common, they could help to coordinate depositors in running. This would be the sunspot mechanism à la Cass and Shell (1983). Bank failures stemming from sunspots are inefficient. On the other hand, private signals to depositors that convey bank information help to predict how good the health of the bank is, *and also* what other depositors think of bank fundamentals (see Goldstein and Pauzner, 2005).[7] Depending on the noise structure, the unique equilibrium of depositors will be to run if the private signal regarding bank fundamentals is below a threshold. In this case, there will be failure even when banks are solvent but close to insolvency. This is because when the signal is low this conveys that the bank is not so good, but, more importantly, it conveys that other depositors also think that bank fundamentals are close to insolvency. This strategic uncertainty among depositors combined with uncertainty in fundamentals may result in bank failures when the bank is illiquid but solvent. This implies an *ex post* inefficiency that should be contrasted with the *ex ante* disciplining incentives provided by short-term debt (Rochet and Vives, 2004; and Calomiris and Kahn, 1991).

The presence of asymmetric information also illustrates how banking problems can build up over an extended period of time before an 'efficient' or 'inefficient' crisis occurs. In other words, the systemic event is only the manifestation of a more fundamental underlying problem—for example, reckless lending and bad loans—which has been hidden from investors or policymakers for some time or remained unaddressed by either of them. The general repricing of risk that triggered the present crisis started only around June–July 2007, although we know that mortgage lending

[7] If the private signals convey information on bank fundamentals, then private signals of the depositors will be positively correlated and, hence, a signal conveys (i) information on the quality of the bank and (ii) on what other depositors know about the bank.

and complex forms of securitizations had been on an unsustainable path for some years and imbalances built up undetected for a while. There may have been however also other factors contributing to the long build-up period to the present crisis.

Systemic risk and public policy

Strong systemic events, such as contagious failures, may involve externalities; the private costs of the initial failure can be lower than the social costs.[8] As a consequence, individually rational bank management or individually rational (wholesale or retail) depositor behavior may lead to a higher level of systemic risk than would be socially optimal. This is one fundamental rationale for the regulation and supervision of banks; an '*ex ante*' (or pre-emptive) policy to avoid the emergence of systemic problems (in contrast to '*ex post*' polices, such as crisis management). Notice that, in this sense, the socially optimal probability of bank failures is *not* zero. However, for a socially optimal outcome, the probability of 'pure' contagion (a self-fulfilling systemic event, as described above) and certain cases of 'information-based' contagion *is* zero. Apart from investor protection considerations, this point is sometimes also brought forward as a rationale for the introduction of deposit insurance. Another element of the safety net and of crisis management that has been widely debated is emergency liquidity assistance by the central bank to *individual financial institutions* in distress, the default of which may trigger contagion effects. Moreover, since *any* systemic event might involve payment and settlement system problems, which may exacerbate externalities, it also provides a rationale for *ex ante* policies to ensure the safety of those systems (oversight, collateral requirements, position caps, real-time settlement, etc.).

Second, a systemic crisis affecting a large number of banks can—via a 'credit crunch' or a debt deflation—lead to a recession or even to a depression. In such situations (*ex post*) macroeconomic stabilization policies, such as expansionary monetary or fiscal expansions, may be used to dampen the recessionary impact of the financial crisis on the real economy. From an *ex ante* perspective, however, too expansionary macro-policies can also fuel asset price bubbles and excessive risk taking and, thereby, contribute to the emergence of imbalances that imply systemic risks.[9] In the case of systemic risk, allocation and stabilization problems can be closely intertwined. If contagion is very strong, then the microeconomic risk allocation problem can degenerate to a macroeconomic destabilization. So, the

[8] See Slovin, Sushka, and Polonchek (1993) for the social costs incurred by the stakeholders of the initially failing bank, and Iyer and Peydró (2009) for the social costs induced by contagion on the stakeholders of the banks that are financially linked to the initial failed bank.

[9] Jiménez, et al. (2007) and Ioannidou, Ongena, and Peydró (2007) find empirical evidence that expansionary monetary policy encourages bank credit risk taking in the medium term, which may lead to solvency problems.

ex ante (regulation and supervision) and *ex post* (crisis management) policies described in the previous paragraph can both be seen as stabilization policies.

The concept of systemic risk developed above is important for the debate about the role a central bank can play as 'lender of last resort' (LOLR). Some authors include in this the central bank's option to use monetary policy—that is, changes in short-term interest rates (or even, under some special circumstances, 'non conventional monetary policies'). Others limit the LOLR role of central banks to liquidity policies. First, central bank liquidity polices can be used to 'lend to the market', the 'banking system as a whole'. By definition, 'lending to the market' is not sterilized, but any surplus liquidity could normally be taken out of the banking system at a later stage, when the crisis is over. This is the traditional distinction between 'banking' or 'liquidity' policy designed to provide transitory liquidity to banks, and 'monetary' policy focused on price stability. Second, the central bank can provide emergency liquidity assistance to individual banks, so preventing individual failures that, in the absence of such emergency lending, have a high likelihood of causing contagion (systemic risk in the 'narrow' sense). Individual emergency loans can be sterilized through opposite transactions vis-à-vis the market as a whole. For various reasons, the literature about the LOLR showed much more controversy regarding this second type of activity than about the first one (see Goodfriend and King, 1988; and Goodhart and Huang, 1999 for two opposite views).[10] This is one main motivation for developing a more complete concept of systemic risk, for distinguishing between a 'narrow' and a 'broad' type of systemic risk and for putting considerable emphasis on the empirical evidence of contagion in the fourth section of this chapter.

Our concept is also relevant for the observation that systemic bank failures or crises lead to public bailouts. In order to preserve the financial health and monetary policy independence of central banks, their LOLR activity tends to be limited to various forms of liquidity support. As soon as solvency support is required to ensure systemic stability, fiscal authorities need to be involved.

It is now widely recognized that public and private safety nets, whether they take the form of deposit insurance, public bailouts, or LOLR facilities, bear the risk of creating moral hazard. For example, if deposit insurance premiums do not reflect the banks' relative portfolio risks, then the protection may incite the insured to take on higher risks (Merton, 1978). Moreover, market expectations could be created that large banks with substantial market, clearing, and settlement links, with many other players in the financial system are 'too big to fail', or 'too complex to fail'.[11] Such effects may be countered by very effective financial regulation and

[10] Holmström and Tirole (1998) derive an even broader role for the state to provide liquidity to the economy more generally when financial frictions become severe.

[11] The takeover of Bear Stearns by JPMorgan Chase at a very low price supported by the US Fed in March 2008 is widely interpreted as an example of the 'too complex to fail' case.

prudential supervision, as, for example, suggested by Kareken and Wallace (1978), Buser, Chen, and Kane (1981) and Furlong and Keeley (1989) for the case of deposit insurance. They also create a case for 'constructive ambiguity' vis-à-vis the potential use of public emergency lending (see Rochet and Vives, 2004).[12] However, if the measures to control moral hazard are not successful, then the insured institutions could take on more risk and become more vulnerable to adverse shocks. This could contribute to the accumulation of imbalances and also enhance the likelihood that instability propagates across institutions. This latter scenario would imply a higher level of systemic risk through inadequate safety net provisions or, in other words, high costs of maintaining the safety net.

When considering different policies to contain systemic risk, the issue of market-oriented approaches to deal with banking instabilities is also raised. Whereas historical analogies bear the risk of neglecting the considerable differences between today's and former bank systems, many industrialized countries have market-oriented elements in their safety net provisions.

Theoretical models of systemic risk in banking

We now consider in greater detail the forms that systemic risk may take in banking, including risks in the interbank payment infrastructure. The theoretical literature in this area is surveyed in the light of the concept discussed in the previous section.

Theories of bank contagion have now developed significantly. Traditionally, many systemic banking panics have been associated with recessions and macroeconomic shocks (systemic risk in the 'broad' sense; see, e.g., Gorton, 1988), but formal theories beyond individual bank run models have been scarce. We start in the next subsection with the recent bank contagion literature. Then, we discuss the literature on systemic banking panics as a consequence of macroeconomic shocks and lending booms.

[12] Some observers wonder whether the decision by the US Fed and Treasury to let Lehman Brothers fail in September 2008 was an application of 'constructive ambiguity'. The loss of confidence following this episode could suggest that such an approach also involves risks. Fed chairman Bernanke, however, argued that the two US authorities did not have the authority to absorb the large expected losses, which was necessary to facilitate the acquisition of Lehman by another firm <http://www.america.gov/st/texttrans-english/2008/October/20081016110841eaifas9.330386e-02.html>. Caballero and Krishnamurthy (2008) challenge 'constructive ambiguity' policies in the case of aggregate 'Knightian' uncertainty (where agents even have no information about the probability distribution of asset returns). They argue that central banks should instead announce that they stand ready to provide liquidity in case of a crisis in order to avoid that investors show 'flight to quality' behavior. This would amount to a 'constructive clarity' approach.

Contagion

As has been observed numerous times in the past, banks may, in the absence of a safety net, be prone to runs. The banking literature in the last twenty years has developed sophisticated models of *single* banks' fragility starting from their balance sheet structure and the speciality of 'the bank contract' (Bryant, 1980; Diamond and Dybvig, 1983; or Chari and Jaganathan, 1988). However, regarding systemic risk this is only part of the story. One should distinguish between a 'run' which involves only a single bank and a 'banking panic' where more than one bank is affected (Calomiris and Gorton, 1991; and Bhattacharya and Thakor, 1993). While the theory of individual runs is relatively well developed, the same did *not*—until quite recently—apply to bank contagion, which requires the consideration of multiple banks and incorporates the *systemic* component.

One can distinguish two main channels through which contagion in banking markets can work: the 'real' or exposure channel and the informational channel. The former relates to the potential for 'domino effects' through direct exposures in interbank markets and payment systems or common exposures to similar non-bank assets. The information channel relates to contagious depositor withdrawals or other funding problems when creditors are imperfectly informed about the type of shocks hitting banks (idiosyncratic or systematic) and about their physical exposures (asymmetric information). In principle, these two fundamental channels can work in conjunction as well as quite independently.

In what follows we organize the bank contagion literature in four subsections according to whether the transmission of instability across banks works through retail depositor reactions, interbank markets, payment and settlement systems, or other factors reflecting endogenously emerging risks.

Interactions of retail depositors across banks

Chen (1999) combines an extension of the bank run models to multiple banks with the rational herding approach. There are two externalities in this model that cause contagious bank runs: a payoff externality through the first-come-first-served rule for servicing withdrawing depositors, and an information externality through the Bayesian updating of beliefs about the macroeconomic situation as a function of observed failures. Chen shows that, even when depositors choose the Pareto-dominant equilibrium, there is a critical number of early failures above which a run on the remaining banks in the system will always be triggered. This critical number is decreasing both in the a priori probability for low investment returns in the economy and in the payoff for early deposit withdrawals, and (weakly) decreasing in the payoff for late withdrawals. Thus the model is also linked to the 'broad' sense of systemic risk. Finally, Chen shows that there also exists a deposit insurance scheme that could eliminate any contagious bank runs in this model.

Interbank markets

A further important step is provided by theories of crises transmitted through the interbank market. Rochet and Tirole (1996a) present a model of the interbank market, where peer monitoring among banks in this market solves the 'moral hazard' problem between bank debt holders and bank shareholder-managers, but also induces contagion risk. The authors show that for certain parameter values of the model, a small increase in the size of the liquidity shock hitting any of the banks can lead to the closing down of the entire banking system, a particularly severe case of contagion.

Allen and Gale's (2000) model focus on the 'physical exposures' of banks in different regions and on the 'real' linkages between regions, as represented by the correlation of liquidity needs of the respective depositors. In their model, both depositors and banks choose deposits to insure against liquidity shocks. Liquidity shocks across regions fluctuate randomly, with aggregate liquidity staying constant with no bank failures. However, in an unexpected state of the world, to which all agents assign a probability of zero *ex ante*, one bank faces additional withdrawals, so that aggregate liquidity is not sufficient to serve all depositors. The authors show that contagion can occur in this situation. Whether and how much propagation of failures emerge depends on the structure of the banking system: with more 'complete' markets (each bank has lending relationships with all the other regions) the system is likely to be more stable.

In a related paper, Freixas, Parigi, and Rochet (2000) discuss physical interbank lending exposures as a consequence of depositors' uncertain geographical consumption decisions. Two pure strategy equilibria are possible when banks are solvent. In the 'credit line equilibrium' efficient interbank lending takes place, all obligations are honored and no contagious runs occur. In the 'gridlock equilibrium' depositors cause inefficient and contagious bank runs for fear of insufficient reserves in the system. Contagious failures occur more easily in the 'credit chain' scenario (analogous to the 'incomplete' case in Allen and Gale, 2000 than in the 'diversified' lending case (credit lines between any two banks exist), although in the 'diversified' case withdrawals can occur more easily.

Following the above, there is a literature that uses 'network theory' to model the different possible sources of connections between financial institutions, stemming from both the asset and the liability side of their balance sheet. By providing means to model the specifics of interbank linkages, network analysis is well designed to explain interbank linkages and contagion through the interbank system. In this type of models, narrow shocks can lead to strong and widespread systemic events. Babus (2007) considers network formation (i.e., optimal interbank arrangements) in order to reduce the risk of contagion. The network is formed endogenously and serves as an insurance mechanism. Better-connected networks are more resilient to contagion, with similar intuition as the one developed by Allen and Gale (2000). The model predicts a connectivity threshold above which contagion does not occur, and banks form links to reach this threshold.

Leitner (2005) studies the trade-off between risk sharing and the potential for contagion in a network model. More interbank linkages imply better risk sharing among banks but also a higher potential for contagion with the possibility of multiple bank failures (systemic event). In the model, the return of a bank depends on the investments of other banks it is linked to. Banks, therefore, may be willing to bail out other banks, in order to prevent the collapse of the whole network.

Conversely, Acharya, Gromb, and Yorulmazer (2008) suggest that surplus banks in the interbank market may strategically under-provide lending to cash stricken banks and, thereby, induce inefficient sales of assets that transmit a crisis. This provides a rationale for the provision of emergency liquidity assistance to individual banks by the central bank.

Brusco and Castiglionesi (2007) extend Allen and Gale (2000) by modeling contagion in the interbank deposit market as an endogenous phenomenon, and introducing moral hazard problems in banks. Banks establish interbank links to insure against liquidity problems and accept the risk of contagion only when the risk is not too large. The main implication is that contagion is a rare phenomenon, since otherwise the banks would avoid establishing financial linkages. In addition, in their model, the extent of contagion is the greater the larger the number of interbank deposit cross-holdings since more banks will get affected by the initial failure. In consequence, the more interbank links banks have, the higher the potential for contagion.

Mishkin (1991) and Davis (1994; 1995) argued that 'adverse selection' play an important role in the transmission of financial crises. Flannery (1996) sketched a model of interbank market crises due to asymmetric information among competing banks. Banks receive imperfect signals about the quality of prospective borrowers. Following a large shock in the financial system, banks may become more uncertain about their rivals' screening ability. As they feel less able to distinguish between banks exposed to more or less risky borrowers, lenders raise interest rates across the board. If the loan rate becomes too high, 'good' banks might not be able to repay their interbank loans any more, so that illiquid but solvent banks may go bankrupt.

Ferguson, et al. (2007) and Cassola et al. (2008) argue that adverse selection problems also played an important role in the transmission of the credit market crisis starting in August 2007. Heider, Hoerova, and Holthausen (2008) develop a theoretical model starting from this idea. They study the functioning and possible breakdown of the (unsecured) interbank market due to asymmetric information about counterparty risk in a Diamond-and-Dybvig-type model. The novel feature is that banks are privately informed about the risk of their illiquid investment. Since asset risk creates counterparty risk in lending, asymmetric information creates frictions in the interbank market as suppliers of liquidity protect themselves against lending to 'lemons'. The model generates several interbank market regimes that are in line with observed developments before and during the present financial crisis.

There are some applications of global games theory to interbank contagion. The advantage of using this technique is to analyze at the same time coordination

problems among depositors (panics) with depositor disciplining based on the health of the banking system (see Goldstein and Pauzner, 2005; and Rochet and Vives, 2004). In fact, the level of (rational) panics depends in equilibrium on the level of banks' fundamentals and the equilibrium is unique. Dasgupta (2004) analyzes crossholding of deposits among banks as a source of contagious breakdowns. He shows that failure in one bank reduces the value of creditor banks thereby increasing its probability of failure. In Iyer and Peydró's (2007) model, depositors may start running when there are bank shocks forcing banks to unwind their positions in the interbank market to pay back the depositors thus generating a strong systemic event.

Fecht, Grüner, and Hartmann (2007a) discuss the relationships between financial integration and systemic risk. They compare segmentation with three forms of inter-regional risk sharing: (i) integration through the secured interbank market; (ii) integration through the unsecured interbank market; and (iii) integration of retail markets. The secured interbank market is an optimal risk sharing device when banks report liquidity needs truthfully. It allows diversification without the risk of cross-regional bank contagion. However, free-riding on the liquidity provision in this market restrains the achievable risk sharing as the number of integrated regions increases. In too large an area this moral hazard problem becomes so severe that either unsecured interbank lending implying contagion risk or, ultimately, the penetration of retail markets is preferable. Financial integration can also promote specialization in lending (and therefore production), which enhances risk sharing but also increases cross-border contagion risk even though this may be optimal from a welfare perspective (Fecht, Grüner, and Hartmann, 2007b).

Payment and settlement systems

By providing the technical infrastructure through which wholesale bank market transactions are settled, large-value payment systems determine the physical exposures among financial institutions. Beyond the explicit interbank lending, one needs to take into account the implicit lending that may arise in payment system. In a way, looking at payment systems is like looking at the network of interbank exposures with a magnifying glass. Hence, depending on their internal organization they also influence how shocks may propagate through the financial system, in particular how severe bank contagion can be.

There are three main types of interbank payment systems: net settlement systems, gross settlement systems, and correspondent banking. See our previous survey, de Bandt and Hartmann, 2002 for details. We just note here that the risks of netting systems explain the spreading of real-time gross settlement (RTGS) systems across the world (Bech and Hobijn, 2007). Most real-life systems have specific additional institutional features in order to reduce systemic risk or liquidity costs

(and 'gridlock' risk) in both net and gross systems, so that the two types that are very different in theory can become quite similar in practice ('hybrid' systems).

The development of theoretical models describing the risks of different payment and settlement systems is not a very active research field. Angelini (1998) argues that in a game of an RTGS system where intraday liquidity is available from the central bank for a fee proportional to the size of the overdraft (rather than collateralized) the competitive (Nash) equilibrium is not welfare optimal, since the cost of intraday credit induces banks to delay payments rather than to draw on the overdraft facility. These payment delays result in network externalities, since payees attempt to free ride on other banks' reserves, thereby reducing overall liquidity. He concludes that fees should therefore be low enough and variable over the day so as to discourage late payments. In contrast, Humphrey (1989) has argued that payment delays in gross systems with uncollateralized overdraft facilities may be desirable to reduce the actual overdrafts and therefore systemic risk (or the costs for the system guarantor). The systemic risk related to overdraft facilities are controlled differently on both sides of the Atlantic. The Eurosystem, running the Trans-European automated real-time gross settlement express transfer system (TARGET), requires complete collateralization. The US Federal Reserve, running Fedwire™, charges daylight overdraft fees. Alternatively, routine queuing facilities can be established, which, however, imply similar risks as net settlement systems.[13]

In a more elaborate model, Freixas and Parigi (1998) (building on McAndrews and Roberds, 1995) introduce geographical consumption preferences in a Diamond–Dybvig-type model, which lead to 'interbank payments' between two regions. 'Gross settlement' imposes relatively high opportunity cost through foregone interest on liquidated investments but is free from contagion in this framework. With 'net settlement' the banks can extend credit lines to each other in order to finance future consumption of 'foreign' consumers. The 'net system' exhibits systemic risk and potential welfare losses in so far as inefficient banks may stay open for longer. Holthausen and Rønde (2002) study the implications of coexisting

[13] Schoenmaker (1995) compares the costs of multilateral net settlement systems (à la CHIPS in the US) with collateralized RTGS systems (à la TARGET in Europe). It turns out that the average loss through settlement failures is higher in the net than in the gross system. But the costs of settlement delays (or even gridlock) and collateral requirements are lower in the net systems. This might explain why central banks often prefer 'safer' gross systems while market participants favor 'less costly' net systems, and it also reflects the trade-off between risks and costs described in Berger, Hancock, and Marquardt (1996). Elaborating on Schoenmaker's comparative approach and using a theoretical framework similar to that of Angelini (1998), Kobayakawa (1997) provides a broad analysis of the efficiency of multilateral net settlement and both types of RTGS, with full collateralization of intraday overdrafts ('EU-type') and with fees on uncollateralized overdrafts ('US-type'). However, he focuses on their relative efficiency and (apart from externalities through payment delays) he does not derive any firm conclusions on 'strong' systemic events.

international gross- and net settlement systems for cross-border systemic risk, when bank supervisory information is only generated at the national level.

Kahn, McAndrews, and Roberds (2003) use an incomplete contract approach to explain gridlock in gross settlement systems without delivery versus payment (DVP). When the underlying trade does not occur at the same time as settlement, there is a possibility to default on obligations ('strategic default'). Even in the absence of fundamental uncertainty about the value of banks' investments, if banks have distrust in their mutual creditworthiness, they may find default as their best option and will not make the pay-ins agreed upon. Two equilibria are possible, either all banks settle normally or no trade occurs ('autarky'). Gridlock does not arise, however, if the central bank guarantees normal settlement. Since banks are fundamentally solvent, the central bank incurs no cost. This is not the case for net settlement systems, which can economize on collateral requirements and avoid trading delays.

The incomplete contract approach is also a useful approach to understand systemic risk in the settlement of foreign exchange transactions, traditionally characterized by a promise to deliver the appropriate currency two days after the initial trade.[14] In volatile foreign exchange markets it is not possible to draft 'complete' contracts including all contingencies. Banks' willingness to carry out the pay-ins in the amount they have promised depends therefore on the appropriate incentives. In this regard, as indicated by Kahn and Roberds (2001), the CLS Bank (for continuous linked settlement) offers a new system increasing banks' compliance with payment obligations. Indeed, established in 2002, CLS implemented payment versus payment (PVP—one leg of a transaction is never settled without the other) and the CLS bank becomes the counterparty. However, from an overall welfare point of view, it might desirable to let some trade go through, so that payments to banks in distress continue. All in all, CLS achieves better bank compliance, at the cost of *ex post* misallocation as well as lower incentives to monitor counterparties' quality.[15]

'Fire sales', liquidity problems, and endogenous risk

Diamond and Rajan (2005) argue that banks are characterized by specific knowledge about borrowers, which make their assets particularly illiquid. So, if a bank fails, then the common pool of liquidity shrinks creating or exacerbating aggregate

[14] In 1974, a number of Bankhaus Herstatt's counterparties had settled their Deutschmark obligations in anticipations of offsetting dollar settlements. The latter could not be made since Herstatt was declared insolvent in between, resulting in a loss of principal for Herstatt counterparties (see Kahn and Roberds, 2001).

[15] Yamazaki (1996) focuses on the relative importance of systemic exposures in bilateral as compared to multilateral (decentralized with loss sharing among participants and without a clearing house) foreign exchange netting. He finds that for single failures multilateral netting reduces other banks' exposures as compared to bilateral netting, if the initial loss is not 'extreme'. However, when a chain reaction of failures occurs, he shows that there are plausible cases in which they are more severe under-multilateral netting (which has moral hazard implications) than under bilateral netting.

liquidity shortages, which in turn may cause further failures. Carletti, Hartmann, and Spagnolo (2007) link individual bank and aggregate interbank market liquidity with competition in the loan market. When interbank markets are relatively efficient, bank concentration can exacerbate aggregate liquidity fluctuations. If a central bank does not offset them through liquidity provision, then system-wide liquidity shortages can become more severe and frequent. Unfortunately, liquidity and solvency problems interact and can cause each other, making it hard to determine the cause of a crisis. In Acharya and Yorulmazer (2008a), as the number of bank failures increases, the set of assets available for acquisition by the surviving banks expands but the total amount of available liquidity within the surviving banks falls. Since financiers not belonging to the banking system do not have the specific knowledge characterizing banks the previous results imply 'cash-in-the-market' pricing (below the 'fair value') for liquidation of banking assets.

Fecht (2004) asks whether an individual bank run has more severe systemic consequences in bank- or market-oriented financial systems. He finds that a bank run on a single bank causes contagion via the financial market only in moderately bank-dominated financial systems where 'fire sales' of long-term financial claims by a distressed bank cause a sudden drop in asset prices, which hurts other banks. Cifuentes, Shin, and Ferrucci (2005) present a model where financial institutions are connected via common portfolio holdings. Contagion is mainly driven here by changes in asset prices through forced sales of assets by some banks that depress the market price inducing further distress to other banks. In Allen and Carletti (2008b), asset prices in some markets may reflect the amount of liquidity available in the market rather than the future earning power of the asset. Mark-to-market accounting is not a desirable way to assess the solvency of a financial institution in such circumstances since it can lead to contagion where none would occur with historic cost accounting.[16]

Brunnermeier and Pedersen (2009) analyze the interplay between market liquidity (i.e., the ease of trading) and funding liquidity (i.e., the availability of funds). Traders in general (or banks in particular) provide market liquidity and their ability to do so depends on their funding—that is, their capital and the margins charged by their financiers (e.g., wholesale depositors). In times of crisis, reductions in market liquidity and funding liquidity are mutually reinforcing, paving the way for a liquidity spiral which can have devastating effects on prices and quantities. This endogenous amplification of initial problems in a crisis has also been studied by Adrian and Shin (2008). They consider a model of leverage and balance sheet size for financial intermediaries which fund their activities through collateralized borrowing. Leverage and balance sheet size decrease together when measured risks increase. When the loss distribution is exponential, the behavior of intermediaries conforms to the value-at-risk rule, in which exposure is adjusted to maintain a constant probability of default. In a system context, increased risk reduces the debt capacity

[16] See also Allen and Gale (2005).

of the financial system as a whole, giving rise to amplified de-leveraging by institutions through the chain of repo transactions.

Macroeconomic fluctuations, aggregate shocks, and lending booms

Surprisingly, the literature on systemic banking risk related to aggregate fluctuations is by no means as well developed as is the literature on bank contagion. There are two forms of these risks: large and widespread shocks and the widespread accumulation of imbalances, such as credit booms. The former is an *ex post* argument in the sense that many banks may get into trouble at the same time when economic downturns or widespread financial market crashes occur. The latter is an *ex ante* argument in the sense that there are mechanisms in financial systems that encourage similar forms of risk taking being pursued by many banks at the same time. So, widespread imbalances can accumulate over time, which may unravel (more or less) violently only much later, be it through a macroeconomic shock or other events. Given their different nature we discuss the two sources of systemic risk in separate subsections.

Aggregate shocks to the banking sector

It has been observed that many banking crises have occurred in conjunction with cyclical downturns or other aggregate shocks, such as interest rate increases, stock market crashes, or exchange rate devaluations (see, e.g., Gorton, 1988 and below). Why is it that banks simultaneously get in trouble in those events (included in the concept of systemic risk in the 'broad' sense according to the terminology given above), even in the absence of direct interbank contagion, and why are prudent banks not better protected than imprudent ones? One answer could be given on the basis of the individual bank run models. News about a cyclical downturn, for example, could provide the negative signal about banks' loans to all or a subset of depositors. Allen and Gale (1998) take issue with the interpretation of bank runs as random phenomena, because of their historical association with severe 'business cycle fluctuations'. They argue that in this framework if depositors make their withdrawal decisions based on a leading indicator for business cycles, the first-best outcome can occur in spite of the non-contingent character of the deposit contract. However, the result breaks down when early withdrawals are costly, so that a public intervention is necessary to restore the first-best outcome. Finally, Chen (1999) shows within his model that an adverse macroeconomic shock will also increase the likelihood of bank contagion.[17]

[17] Of course, there can also be the reverse causality. Restrictions in bank lending due to financial fragility may affect the business cycle, thereby creating adverse acceleration or feedback effects. See, in particular, Mishkin (1991); and Bernanke, Gertler, and Gilchrist (1999). In the current global financial crisis this direction of causality was the more relevant one.

Hellwig (1994) studies the efficient allocation of 'interest rate risk' induced by technology shocks and argues that part of it should be borne by agents with urgent liquidity needs—that is, early withdrawing depositors. This is, however, not the case in standard bank deposit contracts. In a similar vein, Cukierman (1991) provides a macroeconomic model in which an unexpected decline in the supply of deposits occurs after long-term loan contracts have been made, inciting banks to increase their deposit rates to attract new depositors. This suggests that interest rate changes and bank profits are inversely correlated. He derives from this fact an *ex post* rationale for the US Federal Reserve to smooth interest rates in attempts to stabilize the financial system.

Another source of systematic shocks to the banking sector can be 'financial market crashes' or 'market liquidity crises'—in particular when they concern any of the major markets and when they are contagious across markets (see, e.g., Morgenstern, 1959; King and Wadhwani, 1990; and Hartmann, et al., 2004). Commercial and universal banks have become more involved in financial market trading (as opposed to traditional lending). Moreover, as part of the securitization trend over the last decade and more active credit risk management, banks have invested a lot in asset-backed securities and structured credit products, using also credit derivatives very actively. As a consequence, their trading books have grown significantly, exposing them more to shocks originating in financial markets. This implies that the structurally higher systemic risk in banking markets will be more dependent on market risk, (supposedly) tradable credit risk and liquidity in those markets than has previously been the case.

Various negative events in financial markets may increase uncertainty and the ability and willingness to trade in these markets. Imperfect information about asset valuations may lead to significant credit spreads even at shorter maturities or even credit rationing (Duffie and Lando, 2001; and Tirole, 2008a, b). Market makers might increase bid-ask spreads to reduce the likelihood of being hit by a transaction ('price rationing') or even 'refuse' to trade at all ('quantity rationing'). Such a liquidity 'freeze' could involve a systematic shock on all those banks and non-bank financial institutions whose risk management strategies depend on the ability to trade in these markets. In the ongoing credit market crisis valuation uncertainties for credit products had these effects and led to illiquidity in structured product and major money markets (Cassola, et al. 2008).

One key question is how banks deal with aggregate shocks to the banking sector. Allen and Gale (2004; 2007) develop a general equilibrium framework for analyzing the normative aspects of financial crises with a more significant emphasis on the relationships between asset prices and banking crisis. They consider the interaction of banks and markets and focus on fundamental shocks (as opposed to coordination problems among depsitors) as the driver of financial crises. Financial intermediaries provide liquidity insurance to consumers against idiosyncratic liquidity shocks. Markets allow financial intermediaries and their depositors to share

aggregate liquidity and return shocks. The authors show that, when markets are incomplete, asset prices must be volatile to provide incentives for liquidity provision (otherwise, agents would not find liquidity holdings valuable). This asset price volatility can lead to costly and inefficient crises. There is a market failure that potentially provides a justification for regulation and other kinds of intervention to improve the allocation of resources.[18]

Credit booms

An issue related to the real macroeconomic shocks discussed above is why banks expand so much credit, implying risks that can bring many of them into trouble at turning points, even though they know they cannot pass on the risk to depositors. The lending boom literature has addressed this question. Minsky (1977; 1982) believed that the post-World War II free-market economy has a natural tendency toward financial instability at the aggregate level. In good times agents consume and invest, generating more income. As 'euphoria' and 'gregarious behavior' pick up, more speculative or even 'Ponzi' finance is undertaken, as opposed to safer 'hedging' finance.[19] The boom is fed by an over-expansion of bank credit until some exogenous outside shock to the macroeconomic systems ('displacement') brings it to an end. Kindleberger ([1978]1996) shares the basic idea, although perhaps being more moderate in pointing out that the market system 'occasionally' faces such bubbles leading to financial crises. These early writers emphasized the role of uncertainty (of the 'Knightian' type) as opposed to risk and the inability of banks to take the appropriate decisions in some circumstances. For example, Guttentag and Herring (1984) develop a simple model of credit expansion and discuss the consequences of 'Knightian' uncertainty about catastrophic shocks on investment returns and default risk premiums. On the basis of results from psychology they also argue that the subjective probabilities attached to catastrophic events will decline as time elapses after the realisation of such an event. This 'disaster myopia' will lead to a widespread underestimation of the likelihood of extreme events that could question the health of banks.[20]

Related explanations for lending booms are found in the more recent rational expectations literature on 'herding' in investment and loan decisions. For example, Banerjee (1992) or Bikhchandani, et al. (1992) introduce formal models of information externalities that can lead to herding. Each agent only observes the actions of other agents and uses Bayesian updating to derive his or her own subjective probabilities of future returns on his investment decisions. Scharfstein and Stein

[18] Allen and Carletti (2008a) develop a model in which credit risk transfer can lead to contagion. In fact, they argue that credit risk transfer can be detrimental to welfare.

[19] Some observers may regard the Madoff scandal that became known during the present crisis as an example of the greater scope for Ponzi schemes during boom times.

[20] See also Caballero and Krishnamurthy (2008).

(1990) model managers' incentives to mimic others in investment or loan decisions, when their own evaluation and reputation depends on their performance relative to the rest of the market.[21] Acharya and Yorulmazer (2007; 2008b) argue that banks may take correlated loan decisions so that in case of problems there will be 'too many banks to fail' and, hence, banks will get financial support from the regulator. However, a problem of this literature is that it does not give clear explanations as to which events can start a herding wave and when it breaks down.

The stance of 'monetary policy' may also affect bank risk taking and, in general, asset prices. The development of microeconomic banking models with such monetary channels is at an early stage. Allen and Gale (1998; 2000; and 2007) and Diamond and Rajan (2006), among others, have made steps in this direction. In Diamond and Rajan (2006) banks take higher liquidity risk when monetary policy is expansive. In their model, which provides 'a liquidity version of the lending channel' of the monetary policy transmission mechanism, banks finance illiquid long-term projects with very liquid demand deposits. This mismatch makes banks reluctant to grant loans in times of liquidity shortages. Depending on the aggregate real liquidity conditions, monetary intervention can play a useful role by limiting the depositors' incentives to withdraw. Banks will respond by continuing, rather than curtailing, risky credit.[22] In line with such arguments, Dell'Ariccia and Marquez (2006) develop a model in which banks' incentives to screen borrowers diminish as interest rates become lower.[23] In sum, too lose monetary policy may contribute to the emergence of credit booms.

A further branch of the literature explains excessive or too-risky lending by banks with 'moral hazard' (see also above). These writings refer to features of banking markets that normally do *not* exist for other industries. For example, Merton (1977; 1978) develops a model showing how fixed rate deposit insurance premiums that are insensitive to banks' portfolio risks (as observed in many countries) may lead them to increase risk taking in order to maximize the put-option value on the insurance corporation's funds. Boot and Thakor (1993) further argue that such deposit insurance can lead to an inefficiently low level of monitoring. Applying modern corporate finance models of firms' capital structures to the case of banks, Dewatripont and Tirole (1994) argue that banks' excessive reliance on debt financing (partly related to their provision of retail

[21] They quote Gwynne's (1986) description of a typical credit analyst's behavior in lending decisions to less developed countries: 'His job would never be measured how correct his country risk analysis was. At the very least, Herrick was simply doing what hundreds of other large international banks had already done, and any ultimate blame for poor forecasting would be shared by tens of thousands of bankers around the globe; this was one of the curious benefits of following the herd.'

[22] See also Rajan (2006). [23] See also Ruckes (2004).

payment services to a large number of small and relatively uninformed depositors) can also lead to more risk taking in lending. Owing to the existence of explicit or implicit government guarantees for financial institutions, the issue of moral hazard as also been raised in the context of the US savings and loans crisis (Kane, 1989) or regarding the lending boom that partly led to the East Asian crisis (Krugman, 1998). Goodhart and Huang (1999), however, show that a positive level of moral hazard resulting from safety net provisions, such as lending of last resort, might be unavoidable or even optimal to contain the systemic costs or monetary disturbances associated with financial crises.

This credit boom literature addresses the issue of systemic risk in an indirect way. Banks' (or other financial intermediaries') herding and credit over-expansion leads to the (potentially slow) build-up of imbalances that imply vulnerabilities for a large number of banks (or even other firms and households), which increase the likelihood as well as the severity of systemic events. Once a negative aggregate shock or other event makes the non-sustainability of the boom apparent many banks may face similar problems simultaneously. According to Minsky, Kindleberger, and others such financial cycles emerge endogenously as an inherent part of a market economy with relatively unregulated financial markets. The modern expression for this phenomenon is the strong 'pro-cyclicality' of financial systems. Many argue that the financial crisis that started in the summer of 2007 is also a reflection of it.

Empirical evidence on systemic risk in banking

In this section we survey the existing empirical evidence on systemic events and systemic bank crises. The objective is to review how much we know scientifically about how pervasive the different channels of systemic risk are in different national and international contexts. Unfortunately, for reasons of the availability of data points and the still ongoing nature of the present financial crisis, econometric analyses of this important episode were hardly available when this chapter was written. Keeping this in mind, there is 'anecdotal' and descriptive evidence of systemic events since the summer of 2007. On the one hand, large and complex financial institutions failed (notably Lehman Brothers) and major markets became dysfunctional (notably money and securitization markets). On the other hand, many large and complex financial institutions were saved or received other public support, as their failures were expected to pose significant systemic risks. In the judgment of the authors of this chapter, probably few experts would challenge the

view that we are experiencing a systemic financial crisis at present. In this section we will first address the large number of econometric papers that attempt to identify contagion effects and then the few econometric papers dealing with joint crises and aggregate fluctuations.

Evidence on bank contagion

Testing for bank contagion amounts to testing whether 'bad news' about the failure of a specific bank (or group of banks) adversely affects the health of other banks. On the other side, systemic risk in banking markets in the 'broad' sense also includes simultaneous bank failures—for example, as a consequence of macroeconomic shocks or following the build-up of imbalances such as lending booms. In the subsection we will first address the large number of empirical papers that attempt to identify contagion effects and then the few empirical papers dealing with joint crises and aggregate fluctuations.

Intertemporal correlation of bank failures

The common ground of this first branch of the bank contagion literature is a test for autocorrelation in bank failures. Provided that all macroeconomic shocks are effectively covered by the control variables, a positive and significant autocorrelation coefficient indicates that bank failures and periods of tranquillity cluster over time, which is consistent with the contagion hypothesis. These tests have to be undertaken for historical periods in countries without strong (public) safety nets.

Hasan and Dwyer (1994) and Schoenmaker (1996) have substantially refined Grossman's (1993) approach and provide evidence of intertemporal failure clustering in times of 'free banking' in the US. This approach seems to have been relatively successful in supporting the contagion hypothesis, under the caveat that autocorrelated macroeconomic factors that have been omitted would cloud any 'evidence' of contagion.

Duration of bank survival unexplained by fundamentals

In a study of US banking crises during the Great Depression, Calomiris and Mason (2003) apply a microeconometric duration model in which bank survival time is explained by micro and macroeconomic fundamentals, 'panic' dummies, and the level of deposits at failing banks in the same county. The results indicate the presence of some bank contagion effects in specific episodes, but also that they seem to have been rather contained—namely, limited to a specific region of the US. The authors point out, however, that some effects might still be related to unobservable regional or national fundamentals.

Event studies on stock price reactions

The most popular approach to test for contagion effects turned out to be event studies of bank stock price reactions in response to 'bad news', such as the announcement of an unexpected increase in loan-loss reserves or the failure of a commercial bank or even of a country to service its debt. The presence of contagion is usually tested by measuring 'abnormal' bank stock returns (measured by the deviation from a standard capital asset pricing model (CAPM) on historical data) following the announcement of 'bad news' for other banks.

Aharony and Swary (1983) were the forerunners of this approach, who looked at the effects of the three largest bank failures in the US before 1980. The literature also includes Swary (1986) for the Continental Illinois National Bank; Wall and Peterson (1990) for the Latin American debt crisis; Jayanti and Whyte (1996) for banks with significant LDC debt exposures after Continental's failure; and Peavy and Hempel (1988) for regional repercussions after the Penn Square Bank failure of Oklahoma in 1982. In a similar vein, Madura and McDaniel (1989), Docking, Hirschey, and Jones (1997), and Slovin, Sushka, and Polonchek (1999) find that investors better anticipate unfavorable announcements from the large and 'visible' money center banks than from regional banks.

The early results of adverse 'external' stock market reactions to 'bad news' triggered a debate about whether they can be interpreted as evidence of 'pure' contagion effects or whether they rather reflect rational investor choices in response to the revelation of new information. Cornell and Shapiro (1986), Smirlock and Kaufold (1987), Musumeci and Sinkey (1990), Karafiath, Mynatt, and Smith (1991), and Madura, Whyte, and McDaniel (1991) conclude that the strength of abnormal returns during the 1980s was linked to banks' own exposures. Kho, Lee, and Stulz (2000) analyze the impact of the emerging market crises in the 1990s and the LTCM crisis on American banks and find that the market was able to discern between exposed and non-exposed banks. The LTCM crisis had no significant contagion effects in the banking sector, but banks that participated in the LTCM rescue experienced negative stock returns when the rescue was announced. The general result of this debate was that abnormal returns varied in proportion to banks' exposures to countries with problems, which is consistent with the hypothesis of rational investor choice.[24] Extensions outside the US are scarce. See, however, Gay, Timme, and Yung (1991) for Hong Kong during the problems in the 1980s.

In terms of the concepts developed above, this literature considers *weak* systemic events, since stock price fluctuations do not imply failures. However, many of the systemic events studied seem to have been 'efficient'—that is, in

[24] However, there is one early study, Schoder and Vankudre (1986), that challenges the market efficiency hypothesis during the August 1982 Mexican debt crisis.

proportion to actual exposures (see above). Also, several of the cases studied rather represent systemic repercussions in the 'broad' sense, since, for example, events related to the LDC debt crisis could be regarded as caused by an aggregate shock.

Examinations of bank debt risk premiums

Some related research has also been carried out with market prices of bank debt instruments, such as certificates of deposits, bonds, and interbank deposits (Carron, 1982; Saunders, 1986; Saunders, 1987; Karafiath, Mynatt, and Smith, 1991; Cooperman, Lee, and Wolfe, 1992; and Jayanti and Whyte, 1996). The results, however, were quite mixed and the evaluation of the event study approach applied to risk premiums in debt rates, as a test for contagion effects, is of course similar to the application to equity returns (see above).

Extreme bank stock price spillovers

Related to but methodologically distinct from the above event studies are analyses of bank stock price spillovers to measure bank contagion. In this approach, if a large negative return of bank *i* (or a group of banks) is associated with a large negative return of bank *j*, then this is taken as evidence of bank contagion.

De Nicolo and Kwast (2002) use simple correlations, market return models, and generalized autoregressive conditional heteroskedasticity (GARCH) models with trends in correlations for US large and complex banking organizations (LCBOs) during the period 1990–9. They find that stock return linkages between them have significantly increased after the mid-1990s and that consolidation among these institutions has contributed to this increase in 'systemic risk'.

Inspired by the cross-country equity market work by Bae, Karolyi, and Stulz (2003), Gropp and Moermann (2004) estimate 'co-exceedances' (defined as two banks exceeding the 90 or 95 percentile of the empirical distribution at the same time) of stock returns and equity-derived distances-to-default in a sample of European banks between 1991 and 2003. The authors detect a complex network of within and across-country linkages between European banks.

Hartmann, Straetmans, and de Vries (2005) apply for the first time extreme-value theory (EVT) to the banking sector. They develop a measure of multivariate extreme spillovers between bank stocks, which is based on the conditional probability of any set of banks facing a dramatic decline in its stock price given a dramatic decline of the stock prices of any set of other banks. The results for LCBOs between 1992 and 2004 suggest that multivariate extreme bank spillover risk in the US is both economically and statistically higher than in the euro area. This measure of banking system risk has gradually increased during the second half of the 1990s, although only to a very limited extent in Europe, and stayed at the more elevated levels until the end of the sample period.

Gropp, Lo Duca, and Vesala (2007) use the multinomial logit model introduced in the contagion literature by Bae, Karolyi, and Stulz (2003). For a sample of banks from six major European countries during the period 1993–2002, they estimate how the number of co-exceedances in a specific country is affected by the number of co-exceedances in the other countries, with various control variables. They find significant cross-border contagion effects, measured by lagged co-exceedances, which are higher in the 2000s compared with the 1990s.

As for the event study approach, stock market spillovers can only be identified for listed banks and they tend to signal systemic events in the weak sense. Further progress could also be made in better controlling for aggregate shocks. Having said that, three of the above papers found increased systemic risk in major banking systems before the current crisis broke out.

Analyses of deposit flows

Another test of contagion measures the reaction of depositors (wholesale and retail) to 'bad news'. If in response to problems revealed about bank *i* (or a group of banks), depositors also withdraw funds from bank *j*, then there is evidence of a contagious bank run. In this sense an analysis by Saunders (1987) suggests that the 'denial of rumours' by the US Office of the Comptroller of the Currency about the health of Continental Illinois Bank in May 1984 seems to have triggered a 'flight to quality' by large US banks but not a general run.

Saunders and Wilson (1996) provide evidence during the Great Depression in the US that is consistent with panic-type 'pure' (regional) contagion effects between 1930 and 1932 and with 'flight-to-quality' phenomena in 1929 and 1933. However, they also observe that the level of withdrawals at failing banks was always significantly higher than at non-failing banks, which could be interpreted as higher levels of 'informed' withdrawals at unhealthy banks as compared to 'uninformed'/'purely contagious' withdrawals at banks that in the end turned out to be healthy. Again, in our terminology, this approach can only address the occurrence of ('narrow') systemic events in the 'weak' sense. However, the possibility that some of the failing banks considered in this chapter collapsed as a consequence of 'uninformed' withdrawals, while being fundamentally solvent, cannot be excluded. Calomiris and Mason (1997) examine the June 1932 Chicago bank panic and conclude that only weaker banks *ex ante* actually failed during the panic, which is consistent with the hypothesis that 'pure' contagious failures, or 'strong' systemic events (in the 'narrow' sense), did not occur. They explain this finding with the existence of private co-operative arrangements among banks. In Calomiris and Mason (2003) they argue that the regional 1930 bank panic was associated with greater deposit withdrawals than could have been predicted from bank level micro data, and regional and national fundamentals, but they question the notion

that uniform withdrawals, unexplained by fundamentals, have happened on the national level before 1933.

In the international context, van Rijckeghem and Weder (2003) study the direction of bank flows during various crises. They provide evidence that spillovers from a crisis country to other countries was caused by creditor banks' exposures after the Mexican and Asian crises in the 1990s, but not after the Russian crisis in 1998.

Iyer and Peydró (2009) test whether interbank lending exposures can explain contagious deposit withdrawals, as predicted, for example in Allen and Gale (2000), using a detailed micro dataset for a large idiosyncratic Indian bank failure. Banks with higher interbank exposure to the failed bank experience higher deposit withdrawals. This relationship is non-linear and is stronger for banks whose fundamentals are weaker. In addition, more exposed banks suffer further as other banks do not renew their interbank loans. Finally, both households and firms suffer from the relationship in terms of reduced deposits and loans, respectively.

Simulations of chain reactions through interbank exposures

Theoretical models of contagion suggest that banking contagion may also happen directly through interbank exposures. Kaufman (1994) reports that, shortly before the failure of Continental Illinois, sixty-five financial institutions had uninsured exposures to the bank in excess of their capital. However, Continental's actual losses finally reached 5 percent, which was below the 60 percent threshold that would have triggered insolvency of exposed banks.

There is a recent central bank research literature using confidential and often incomplete data on interbank exposures to assess the risk of contagion using counterfactual simulations. Single or multiple banks are assumed to fail and the simulations derive which other banks would fail as a consequence of this, everything else equal. Some of these studies suggest that contagion effects are relatively limited in most scenarios. For example, Furfine (2003) finds this for Fed Funds transactions by US commercial banks settled through the Fedwire™ real-time gross settlement system in 1998. He also shows that the degree of systemic risk depends dramatically on the recovery rate that is assumed for bank failures.[25] Elsinger, Lehar, and Summer (2006a; 2006b) combine information about interbank exposures with that on macroeconomic fluctuations for Austria and the UK. It turns out that contagion risk is generally quite low and dominated by risks from correlated assets. Contagion risk through interbank exposures is also measured to be relatively low in Belgium, although it changes over time with increasing risk

[25] In addition, Furfine (2003) studies the federal funds market during the LTCM and Russian crises, finding that risk premiums on overnight lending were largely unaffected and lending volumes increased.

emanating from foreign banks (Degryse and Nguyen, 2007). Mistrulli (2007) reaches qualitatively similar conclusions for Italy, where the risks from foreign banks are higher than the ones from domestic banks. In contrast to Degryse and Nguyen (2007), however, he finds that the move from a complete interbank lending structure to a money center bank structure has increased rather than decreased contagion risk, which is in line with the theoretical predictions discussed above.

A few other studies simulate greater contagion risk through interbank exposures. Upper and Worms (2004) estimate for Germany that the failure of a single bank could lead to the breakdown of up to 15 percent of the banking system in terms of assets. Van Lelyveld and Liedorp (2004) find significant contagion risk for the Netherlands. The greatest source of this are foreign banks, notably from Europe and North America. The more material results are, however, often the consequence of more extreme assumptions on recovery rates or other variables.[26]

The general conclusion from this literature is that the banking systems demonstrate a relatively high level of resilience. Widespread contagion tends to be simulated to occur only in cases of very large shocks or very low recovery rates, which assigns a relatively low probability to it. This result is, however, subject to two caveats. First, it heavily depends on how the linkages between banks, represented by credit exposures in the interbank market, are estimated. For many countries, data are extracted from banks' balance sheets, which only provide information on the aggregate exposure of the reporting institution vis-à-vis all other banks. To estimate bank-to-bank exposures, it is generally assumed that banks spread lending as evenly as possible—that is, maximum entropy, an approach that was introduced by Sheldon and Maurer (1998) when they analyzed the Swiss interbank system. In effect, this assumption requires that banks are connected in a complete network. In consequence, it might bias the results in the light of the theoretical findings that better connected networks are more resilient to the transmission of shocks. In fact, Mistrulli (2007) finds that the Italian interbank system, which shifted from a diversifed interbank system toward a structure with multiple money center banks, exhibited increasing contagion risk. In contrast, Van Lelyveld and Liedorp (2006) argue that the bias from this assumption is low when using Dutch data. The second caveat is that this approach to assess bank contagion does not consider endogenous responses by the market to failures and risks or safety nets. In theory, these could either increase or decrease systemic risk. The experience of the ongoing financial crisis seems to be that endogenous responses by market participants can be very important and amplify risks considerably.

[26] Further studies following the simulation approach include Wells (2004) for the UK; Amundsen and Arnt (2005) (using payment system data) for Denmark; Blavarg and Nimander (2002) for Sweden; and Lubloy (2005) for Hungary. Upper (2007) provides a survey of this literature.

Simulation of chain reactions in interbank payment systems

An earlier literature already assessed contagion risks in payment systems using simulations. Humphrey (1986) suggests that a large share of all clearing house interbank payments' system (CHIPS) participants could default and that the institutions affected by the initial failure were quite different between the two days examined. Angelini, Maresca, and Russo (1996) apply a substantially generalized simulation exercise to the Italian net settlement system. From these simulations, the systemic risk in the Italian settlement system seems to be lower than that for CHIPS.

McAndrews and Wasilyew (1995) undertake a study of systemic risk in net systems with unwinding provisions based entirely on Monte Carlo simulations. It turns out that system-wide repercussions of a failure increase with the average size of bilateral payments, the number of system participants, and with the degree of 'connectedness' between the participants (as measured by the likelihood that any two banks exchange payments). (See also Furfine, 2003, discussed above.)

Recent simulation literature suggests that contagion risks in payment systems are nowadays relatively contained (see, e.g., Bech, Nartop, and Madsen, 2002, on the Danish Interbank Netting System, or Galos and Soramaki, 2005, taking TARGET payment flows and simulating contagion risks under different settlement rules than real-time gross settlement). But, similar caveats apply here as for the simulation approach for interbank markets. For example, this approach does not allow for reactions of other payment-system participants to initial failures and might therefore either overstate contagion risk (e.g., if banks manage to undertake hedging transactions quickly) or understate contagion risk (e.g., if adverse selection phenomena in relation to banks' health in a crisis situation would lead market participants to hold back payments). Moreover, nowadays many net payment systems reduced or removed potential unwinding of transactions for exactly the reason that they might enhance systemic risk. The fact that payment systems did not seem to play any significant role in the ongoing financial crisis is in line with improvements in the safety of them over time and with the latest results of the literature regarding limited risks.

Banking crises, aggregate fluctuations, and lending booms

Whereas there are numerous descriptive accounts of banking crises referring to macroeconomic fluctuations, the number of econometric papers is much more scarce.[27] According to Gorton (1988), during the US National Banking Era

[27] There is, of course, a fairly broad literature about early warning indicator systems to predict individual bank failures and banking system crises for micro- and macro-prudential purposes that we do not cover for reasons of space.

(1865–1914) widespread banking panics did not occur as random events à la Diamond and Dybvig (1983) or Waldo (1985), but rather as 'normal' widespread reactions of depositors to severe cyclical downturns. In contrast, during the Great Depression panics appeared to be much more special events.

Gonzalez-Hermosillo, Pazarbaşioglu and Billings (1997) and Gonzalez-Hermosillo (1999) study the determinants of bank 'distress' in various episodes in the US. It turns out that market and liquidity risk factors played a role in explaining 'distress', whereas the role of credit risk and moral hazard is more case specific. However, the introduction of aggregate variables, such as macroeconomic fundamentals and regional variables, significantly improved the predictive power of the models tested, providing evidence in favor of the macro explanation of systemic bank difficulties in the 'broad' sense.

Demirgüç-Kunt and Detragiache (1998) study the macroeconomic and structural determinants of banking crises in forty-five to sixty developing and industrial countries between 1980 and 1994. Consistently with the business cycle hypothesis for bank crises, in all specifications, GDP growth, real interest rates, and inflation are highly significant. However, private sector credit growth has only explanatory power in some specifications, providing mixed evidence in favor of the lending-boom hypothesis. In contrast, the evidence in favor of the moral-hazard hypothesis regarding explicit deposit insurance schemes is stronger. Since data for crises and non-crises times are pooled, this study can claim to isolate the factors causing full-scale banking crises from those only causing a gradual increase in financial fragility or single bank failures.

Hartmann, Straetmans, and de Vries (2005) apply the tail-beta method (i.e., an extreme-value equivalent of the beta in the CAPM, which measures how very extreme downturns by the market factor affect the propensity of bank stocks to crash as well) to euro area and US large and complex banking organizations during the 1990s and early 2000s. The results suggest that this extreme systematic bank risk is of significant and similar magnitude on both sides of the Atlantic and that in both economies it increased over the sample period.

Gourinchas, Valdés, and Landerretche (2001) examine more narrowly the properties of lending booms in a sample of ninety-one industrial and developing countries between 1960 and 1996 and link them to the likelihood of banking and currency crises. In fact, the unconditional probability of banking crises directly *after* lending boom periods is higher than during tranquil periods. Somewhat contrary to conventional wisdom, they also find that the build-up and ending phases of booms are fairly symmetric, so that on average abrupt and crash-type ends are not consistent with their data.

Dell'Ariccia, Igan and Laeven (2008) link the current subprime mortgage crisis to a decline in lending standards associated with the rapid expansion of this market. They show that lending standards declined more in areas that experienced

larger credit booms and house price increases. In addition, lending standards declined more in areas with higher mortgage securitization rates. This final finding has also been found by Mian and Sufi (2008).

An important issue is also how monetary policy relates to banking system risk. Jiménez, et al. (2007) and Ioannidou, Ongena, and Peydró (2007; 2008) find empirical evidence that lower levels of short-term interest rates increases the credit risk-appetite of banks (implying banks granting loans with higher default probability). Controlling for the macroecomic environment, bank, borrower, and loan characteristics, they find that lower short-term interest rates imply that banks give more loans to either borrowers with bad credit histories or to borrowers with no credit histories, or to borrowers with subprime ratings. More importantly, the new loans have a higher hazard rate. They also find that lower interest rates or higher inflation reduce the default risk of outstanding loans, which implies that in the short run expansive monetary policy reduces credit risk. Ioannidou, Ongena, and Peydró (2007), using the credit register of Bolivia, where the banking system is almost completely 'dollarized', find that not only do banks take on higher credit risk when short-term rates (federal funds rates) are lower, but also loan spreads are reduced, especially for the banks with worse monitoring, thereby suggesting higher bank risk taking.[28]

Conclusion

In this chapter we discussed the various elements of systemic risk in banking, which are essential for the understanding of financial crises. The overall concept developed can be used as a baseline for financial and monetary policies when attempting to maintain stable banking systems. At the heart of systemic risk (in the narrow sense) is the notion of contagion—often a strong form of externality—working from one institution or system to another. In a broad sense, the concept also includes the unraveling of imbalances that have built up over time and the consequences of wide systematic shocks which adversely affect many banks at the same time. In this sense, systemic risk goes much beyond the vulnerability of single banks to runs in a fractional reserve system.

[28] See also Rajan (2006); and Calomiris (2008).

We reviewed the quantitative literature in the light of our concept of systemic risk. Some important new theoretical contributions have appeared in this literature over the last decade. First of all, a considerable number of theoretical studies have now directly addressed the issue of contagion through interbank markets and the relative stability associated with different lending patterns in them. Second, many models stress the interactions between banks and asset prices in crisis periods, emphasizing cumulative disruptions induced by forced asset sales. Third, the global games approach has illustrated the close relationship between information structures and possible multiple equilibria in financial stability problems, and also between liquidity and solvency crises. Owing to unique equilibrium results this approach allows for undertaking comparative statics, which is essential for conducting policy analysis in a model. Finally, some progress has been made in better understanding the role of liquidity for banking system stability. Overall, an important direction of work is how the transmission of shocks in financial systems is amplified through endogenously emerging risks.

Available empirical research still continues to focus more on the assessment of bank contagion phenomena than on the macroeconomic causes for banking system instability. This remains so, even though macroeconomic causes of systemic risk have been identified in a less ambiguous way than the case for contagion. Interesting new developments include the use of actual interbank exposures for counter-factual simulations of contagion risk, the application of extreme-value theory to banking system risk, and a starting focus on factors that lead to the build-up of imbalances in financial systems. Further progress could still be made in controlling for aggregate factors in contagion analyses. The difficulty to distinguish between macroeconomic causes of banking problems and contagion does not allow general answers to the relative merits of different crisis management policies, such as liquidity support to the market as a whole or macroeconomic stabilization policies versus emergency liquidity assistance to individual banks. Studies of systemic risks in payment and settlement systems remain relatively rare. Even though the practical and unambiguous identification of concrete contagion cases continue to be a challenge, the overall understanding of bank contagion risks has increased significantly over the last decade.

The ongoing financial crisis has highlighted the relevance of systemic banking stability. The available literature illustrates a number of mechanisms that played important roles in it. But, other important factors related inter alia to new financial instruments or bank business models should be better understood than is the case, which gives ample room for future research on systemic risk. The analysis of systemic risk as surveyed in this chapter and further progress in this literature will be of great value to the efforts launched recently to strengthen macro-prudential supervision and regulation.

References

Acharya, V. V., Gromb, D., and Yorulmazer, T. (2008). 'Imperfect Competition in the Interbank Market as a rationale for Central Banking', manuscript.

—— and Richardson, M. (eds.,) (2009). *Restoring Financial Stability: How to Repair a Failed System*, Wiley.

—— and Yorulmazer, V. (2007). 'Too Many to Fail: An Analysis of Time-inconsistency in Bank Closure Policies', *Journal of Financial Intermediation*, 16: 1–31.

—— —— (2008a). 'Cash-in-the-Market Pricing and Optimal Resolution of Bank Failures', *Review of Financial Studies*, 21: 2705–42.

—— —— (2008b). 'Information Contagion and Bank Herding', *Journal of Money, Credit and Banking*, 40: 215–231.

Adrian, T. and Shin, H. (2008). 'Financial Intermediary Leverage and Value at Risk', New York Federal Reserve, Mimeo.

Aharony, J. and Swary, V. (1983). 'Contagion Effects of Bank Failures: Evidence from Capital Markets'. *Journal of Business*, 56: 305–17.

Allen, F. and Carletti, V. (2008a). 'Credit Transfer and Contagion', *Carnegie Rochester Conference in Economic Policy*, 53: 89–111.

—— —— (2008b). Mark-to-Market Accounting and Liquidity Pricing *Journal of Accounting and Economics*, 45: 358–78.

—— and Gale, D. (1998). 'Optimal Financial Crises', *Journal of Finance*, 53: 1245–84.

—— —— (2000). 'Financial Contagion', *Journal of Political Economy*, 108(1), 1–33.

—— —— (2004). 'Financial Fragility, Liquidity and Asset Prices', *Journal of the European Economic Association*, 2: 1015–48.

—— —— (2005). 'From Cash-in-the-Market Pricing to Financial Fragility', *Journal of the European Economic Association*, 3: 535–46.

—— —— (2007). *Understanding Financial Crises*, Oxford: Oxford University Press.

Amundsen and Arnt (2005). 'Contagion Risk in the Danish Interbank Market', Danmarks Nationalbank Working Paper 29.

Angelini, P. (1998). 'An Analysis of Competitive Externalities in Gross Settlement Systems', *Journal of Banking & Finance*, 22: 1–18.

—— Maresca, G., and Russo, D. (1996). 'Systemic Risk in the Netting System', *Journal of Banking & Finance*, 20: 853–68.

Ashcraft, A. B. and Schuermann, T. (2008). 'Understanding the Securitization of Subprime Mortgage Credit', Federal Reserve Bank of New York Staff Report 318, March.

Avery, C. and Zemsky, P. (1998). 'Multidimensional Uncertainty and Herd Behaviour', *American Economic Review*, 88: 724–48.

Babus, A. (2007). 'The Formation of Financial Networks', Working Paper 69, Fondazione Eni Enrico Mattei.

Bae, K., Karolyi, G., and Stulz, R. (2003). 'A New Approach to Measuring Financial Contagion', *Review of Financial Studies*, 16: 717–63.

Bandt, O. (de) and Hartmann, P. (2002). 'Systemic Risk in Banking: A Survey', in *Financial Crises, Contagion and the Lender of Last Resort: A Reader*, C. Goodhart and G. Illing (eds.), Oxford: Oxford University Press.

Banerjee, A. V. (1992). 'A Simple Model of Herd Behaviour', *Quarterly Journal of Economics*, 107: 797–811.

Bech, M. and Hobijn, B. (2007). 'Technology Diffusion within Central Banking: The Case of RTGS', *International Journal of Central Banking*, 3: 147–81.

Bech, M., Nartop, L., and Madsen, B. (2002). 'Systemic Risk and the Danish Interbank Netting System', Working Paper 8, Danmarks National Bank.

Berger, A. N., Hancock, D., and Marquardt, J. C. (1996). 'A Framework for Analyzing Efficiency, Risks, Costs, and Innovations in the Payment System', *Journal of Money, Credit, and Banking*, 28: 696–732.

Bernanke, B. S. (1990). 'Clearing and Settlement during the Crash', *Review of Financial Studies*, 3: 133–51.

—— Gertler, M., and Gilchrist, S. (1999). 'The Financial Accelerator in a Quantitative Business Cycle Framework', in *Handbook of Macroeconomics*, I, J. B. Taylor and M. Woodford, (eds.), Amsterdam: Elsevier.

Bhattacharya, S. and Thakor, A. (1993). 'Contemporary Banking Theory', *Journal of Financial Intermediation*, 3: 2–50.

Bikhchandani, S., Hirshleifer, V., and Welsh, V. (1992). 'A Theory of Fads, Fashions, Customs and Cultural Changes as Informational Cascade', *Journal of Political Economy*, 100: 992–1026.

Blavarg, M. and Nimander, P. (2002). 'Inter-Bank Exposures and Systemic Risk', *Sveriges Riksbank, Economic Review*, 2: 19–45.

Boot, A. and Thakor, A. (1993). 'Bank Regulation, Reputation and Rents: Theory and Policy', in *Capital Markets and Financial Intermediation*, C. Mayer and X.Vives, (eds.), Cambridge: Cambridge University Press.

Borio, C. E. V. (2003). 'Towards a Macroprudential Framework for Financial Regulation and Supervision', Bank for International Settlements Working Paper 128, February.

—— and Van den Bergh, P. (1993). 'The Nature and Management of Payment System Risks: An International Perspective', *Bank for International Settlements Economic Papers*, 36.

Brimmer, A. F. (1989). 'Distinguished Lecture on Economics in Government: Central Banking and Systemic Risks in Capital Markets', *Journal of Economic Perspectives*, 3: 3–16.

Brunnermeir, M. and Pedersen, L. (2009). 'Market Liquidity and Funding Liquidity', *Review of Financial Studies*, 22(6), 2201–38.

Brusco, S. and Castiglionesi, F. (2007). 'Liquidity Coinsurance, Moral Hazard, and Financial Contagion', *Journal of Finance*, 62: 2275–302.

Bryant, J. (1980). 'A Model of Reserves, Bank Runs, and Deposit Insurance', *Journal of Banking & Finance*, 4: 335–44.

Buser, S. A., Chen, A. H., and Kane, E. J. (1981). 'Federal Deposit Insurance, Regulatory Policy, and Optimal Bank Capital', *Journal of Finance*, 35: 51–60.

Caballero, J. and Krishnamurthy, A. (2008). 'Collective Risk Management in a Flight to Quality Episode', *Journal of Finance*, 63: 2195–230.

Calomiris, C. W. (2008). 'The Subprime Turmoil: What's Old, What's New, What's Next?', Paper Presented at the *Jacques Polak Conference* (IMF).

—— and Kahn, C. (1991). 'The Role of Demandable Debt in Structuring Optimal Banking Arrangements', *American Economic Review*, 81: 497–513.

—— and Gorton, G. (1991). 'The Origins of Banking Panics: Models, Facts, and Bank Regulation', in *Financial Markets and Financial Crises*, G. Hubbard (ed.), Chicago, IL: The University of Chicago Press.

—— and Mason, V. (1997). 'Contagion and Bank Failures during the Great Depression: The June 1932 Chicago Banking Panic', *American Economic Review*, 87: 863–83.

—— and J. Mason (2003). 'Fundamentals, Panics and Bank Distress during the Depression', *American Economic Review*, 93: 1615–47.

Carletti, E., Hartmann, P., and Spagnolo, G. (2007). 'Bank Mergers, Competition and Liquidity', *Journal of Money, Credit and Banking*, 39: 1067–105.

Carron, A. S. (1982). 'Financial Crises: Recent Experience in U.S. and International Markets', *Brookings Papers on Economic Activity*, 2: 395–418.

Cass, D. and Shell, K. (1983). 'Do Sunspots Matter?', *Journal of Political Economy*, 91: 193–227.

Cassola, N., Drehmann, M., Hartmann, P., Lo Duca, M., and Scheicher, M. (2008). 'A Research Perspective on the Propagation of the Credit Market Turmoil', *European Central Bank Research Bulletin* 7 (June).

Chari, V. and Jagannathan, R. (1988). 'Banking Panics, Information, and Rational Expectations Equilibrium', *Journal of Finance*, 43: 749–60.

Chen, Y. (1999). 'Banking Panics: The Role of the First-Come, First-Served Rule and Information Externalities', *Journal of Political Economy*, 107: 946–68.

Cifuentes, R., Shin, V., and Ferrucci, V. (2005). 'Liquidity Risk and Contagion', *Journal of the European Economic Association*, 3: 556–66.

Cooperman, E. S., Lee, W. B., and Wolfe, G. A. (1992). 'The 1985 Ohio Thrift Crisis, the FSLIC's Solvency, and Rate Contagion for Retail CDs', *Journal of Finance*, 47: 919–41.

Cornell, B. and Shapiro, A. C. (1986). 'The Reaction of Bank Stock Prices to the International Debt Crisis', *Journal of Banking & Finance*, 10: 55–73.

Council of the European Union (2009). 'Presidency Conclusions', Brussels European Council, 18/19 June 2009, 11225/09, 19 June.

Cukierman, A. (1991). 'Why Does the Fed Smooth Interest Rates?, in *Monetary Policy on the 75th Anniversary of the Federal Reserve System: Proceedings of the Fourteenth Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis*, M. T. Belongia (ed.), Dordrecht: Kluwer Academic, 111–47.

Dasgupta, A. (2004). 'Financial Contagion through Capital Connections: A Model of the Origin and Spread of Bank Panics', *Journal of the European Economic Association, 6: 1049–84.*

Davis, E. P. (1994). 'Market Liquidity Risk', in *The Competitiveness of Financial Institutions and Centres in Europe*, D. Fair, (ed.), Dordrecht: Kluwer Academic Publishers.

—— (1995). *Debt, Financial Fragility and Systemic Risk*, 2nd edition, Oxford: Clarendon Press.

Degryse, H., Nguyen, G. (2007). 'Interbank Exposures: An Empirical Examination of Contagion Risk in the Belgian Banking System', *International Journal of Central Banking*, 3: 123–171.

Dell'Ariccia, G., Detragiache, E., and Rajan, R. (2008). 'The Real Effect of Banking Crises', *Journal of Financial Intermediation*, 17: 89–112.

—— —— and Marquez, R. (2006). 'Lending Booms and Lending Standards,' *Journal of Finance*, 61(5): 2511–46.

—— Igan, D., and Laeven, L. (2008), 'Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market', International Monetary Fund, Washington, DC.

Demirgüç-Kunt, A. and Detragiache, E. (1998). 'The Determinants of Banking Crises in Developing and Developed Countries', *IMF Staff Papers*, 45: 81–109.

De Nicolo, G. and Kwast, M. L. (2002). 'Systemic Risk and Financial Consolidation: Are They Related?, *Journal of Banking & Finance*, 26: 861–80.

Dewatripont, M. and Tirole, J. (1994) *The Prudential Regulation of Banks*, Cambridge: MIT Press.

Diamond, D. V. and Dybvig, P. (1983). 'Bank Runs, Deposit Insurance, and Liquidity', *Journal of Political Economy*, 91: 401–19.

Diamond, D. V. and Rajan, R. (2005). 'Liquidity Shortages and Banking crises', *Journal of Finance*, 60: 615–47.
——— (2006). 'Money in a Theory of Banking', *American Economic Review*, 96: 30–53.
Docking, D. S., Hirschey, M., and Jones, V. (1997), 'Information and Contagion Effects of Bank Loan-Loss Reserve Announcements', *Journal of Financial Economics*, 43: 219–40.
Duffie, D. and Lando, D. (2001). 'Term Structure of Credit Spreads with Incomplete Accounting Information', *Econometrica*, 69: 633–64.
Elsinger, H., Lehar, A., and Summer M. (2006a). 'Risk Assessment for Banking Systems', *Management Science*, 52: 1301–14.
——— ——— (2006b). 'Using Market Information for Banking System Risk Assessment', *International Journal of Central Banking*, 2: 137–66.
Evanoff, D., Hartmann, P., and Kaufman, G. (eds.,) (2009). *The First Credit Market Turmoil of the 21st Century*, Hackensack, NJ: World Scientific Publishers.
Fecht, F. (2004). 'On the Stability of Different Financial Systems', *Journal of the European Economic Association*, 2: 969–1014.
——— Grüner, H. P., and Hartmann, P. (2007a). 'Welfare Effects of Financial Integration', Center for Economic Policy Research Discussion Paper 6311, May.
——— ——— (2007b). 'Financial Integration, Specialization and Systemic Risk', Mimeo., European Central Bank.
Federal Reserve Bank of Kansas City (2008). *Maintaining Stability in a Changing Financial System*, Jackson Hole, WY.
Ferguson, R., Hartmann, P., Panetta, F., and Portes, R. (2007). 'International Financial Stability', *Geneva Report on the World Economy*, 9, November.
Financial Stability Forum (2008). 'Enhancing Market and Institutional Resilience', Basel, 7 April.
Flannery, M. (1996). 'Financial Crises, Payment System Problems, and Discount Window Lending', *Journal of Money, Credit, and Banking*, 28: 804–24.
Folkerts-Landau, D. (1991). 'Systemic Financial Risk in Payment Systems', in 'Determinants and Systemic Consequences of International Capital Flows', IMF Occasional Paper 77.
Freixas, X. and Parigi, B. (1998). 'Contagion and Efficiency in Gross and Net Interbank Payment Systems', *Journal of Financial Intermediation*, 7: 3–31.
——— ——— and Rochet, J. C. (2000). 'Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank', *Journal of Money, Credit and Banking*, 32: 611–40.
Furfine, C. H. (2003). 'Interbank Exposures: Quantifying the Risk of Contagion', *Journal of Money, Credit and Banking*, 35: 111–28.
Furlong, F. T. and Keeley, M. C. (1989). 'Capital Regulation and Bank Risk-Taking: A Note', *Journal of Banking & Finance*, 13: 883–91.
Galos, P. and Soramaki, K. (2005). 'Systemic Risk in Alternative Payment System Designs', European Central Bank Working Paper 508, July.
Gay, G., Timme, S., and Yung, K. (1991). 'Bank Failure and Contagion Effects: Evidence from Hong Kong; *Journal of Financial Research*, 14: 153–65.
Goldstein, I. and Pauzner, A. (2005). 'Demand Deposit Contracts and the Probability of Bank Runs', *Journal of Finance*, 60: 1293–328.
González-Hermosillo, B. (1999). 'Determinants of Ex Ante Banking System Distress: A Macro-Micro Empirical Exploration of Some Recent Episodes', IMF Working Paper WP/99/33.
——— Pazarbaşioglu, C., and Billings, R. (1997). 'Banking System Fragility: Likelihood Versus Timing of Failure: An Application to the Mexican Financial Crisis', *IMF Staff Papers* 44: 295–314.

Goodfriend, M. and King, R. (1988). 'Financial Deregulation, Monetary Policy, and Central Banking', in *Restructuring Banking and Financial Services in America*, American Enterprize Institute Studies 481: 216–53.

Goodhart, C. A. E. and Huang, H. (1999). 'A Model of the Lender of Last Resort', L.S.E. Financial Markets Group Discussion Paper 313.

Gorton, G. (1988). 'Banking Panics and Business Cycles', *Oxford Economic Papers*, 40: 751–81.

Gourinchas, P.-O., Valdés, R., and Landerretche, O. (2001). 'Lending Booms: Latin America and the World', *Economia*, 1: 2.

Greenlaw, D., Hatzius, J., Kashyap A., and Shin, H. S. (2008). 'Leveraged Losses: Lessons from the Mortgage Market Meltdown', US Monetary Policy Forum Report 2.

Gropp, R. and Moerman, G. (2004), 'Measurement of Contagion in Banks' Equity Prices', in *Banking, Development and Structural Change*, I. Hasan, J. Lothian, and J. Tarkka (eds.), Special Issue of the *Journal of International Money and Finance*, 23: 405–59.

—— Lo Duca, M., and Vesala, J. (2007). 'Cross–border Bank Contagion in Europe,' *Working Papers Series*: Finance and Accounting, 175, Department of Finance, Goethe University, Frankfurt and Main.

Grossman, R. (1993). 'The Macroeconomic Consequences of Bank Failures under the National Banking System', *Explorations in Economic History*, 30: 294–320.

Guttentag, J. M. and Herring, R. J. (1984). 'Credit Rationing and Financial Disorder', *Journal of Finance*, 39: 1359–82.

Hartmann, P., Straetmans, S., and de Vries, C. (2004). 'Asset Market Linkages in Crisis Periods', *Review of Economics and Statistics*, 86: 313–26.

—— —— —— (2005). 'Banking System Stability: A Cross-Atlantic Perspective', National Bureau of Economic Research Working Paper 11698, October.

Hasan, I. and Dwyer, G. (1994). 'Bank Runs in the Free Banking Period', *Journal of Money, Credit, and Banking*, 26: 271–88.

Heider, F., Hoerova, M., and Holthausen, C. (2008) 'Liquidity Hoarding and Interbank Market Spreads: The Role of Counterparty Risk', Paper presented at the 2008 UniCredit Conference on Banking and Finance, Vienna, 6–7 November.

Hellwig, M. (1994). 'Liquidity Provision, Banking, and the Allocation of Interest Rate Risk', *European Economic Review*, 38: 1363–89.

High-Level Group on Financial Supervision in the EU (2009). 'Larosière Report', Brussels, 25 February.

Hoggarth, G. and Saporta, V. (2001). 'Costs of Banking System Instability: Some Empirical Evidence', *Bank of England Financial Stability Review*, June, 148–65.

Holmström, B. and Tirole, J. (1998). 'Private and Public Provision of Liquidity', *Journal of Political Economy*, 106: 1–40.

Holthausen, C. and Rønde, T. (2002). 'Regulating Access to International Large-Value Payment Systems', *Review of Financial Studies*, 15(5): 1561–86.

Humphrey, D. B. (1986). 'Payments Finality and Risk of Settlement Failure', in *Technology and the Regulation of Financial Markets: Securities, Futures, and Banking*, A. Saunders and L. J. White (eds.), Lexington, MA: Lexington Books, 97–120.

Humphrey, D. B. (1989). 'Market Responses to Pricing Fedwire Daylight Overdrafts', Federal Reserve Bank of Richmond, *Economic Review*, 75: 23–34.

Institute of International Finance (2008). 'Final Report of the IIF Committee on Market Best Practice: Principles of Conduct and Best Practice Recommendations', Washington, July.

Ioannidou V., Ongena, S., and Peydró, J. L. (2007). 'Monetary Policy, Risk-Taking and Pricing: Evidence from a Natural Experiment'. University of Tilburg Working Paper.
——— ——— ——— (2008), 'The Impact of Short-Term Interest Rates on Risk Taking: Hard Evidence', in *The First Global Financial Crisis of the 21st Century*, A. Felton and Reinhart C. (eds.), Centre for Economic Policy Research, 41–5.
Iyer, R. and Peydró, J. L. (2007). 'How does a Shock Propagate? A Model of Contagion in the Interbank Market due to Financial Linkages', University of Amsterdam Working paper.
——— ——— (2009). 'Interbank Contagion at Work: Evidence from a Natural Experiment', *The Review of Financial Studies*, forthcoming.
Jayanti, S. V. and Whyte, A. M. (1996). 'Global Contagion Effects of the Continental Illinois Failure', *Journal of International Financial Markets, Institutions and Money*, 6: 87–99.
Jiménez, G., Ongena, S., Peydró, J. L., and Saurina, J. (2007). 'Hazardous Times for Monetary Policy: What do Twenty-Three Million Bank Loans Say about the Effects of Monetary Policy on Credit Risk-Taking?', Center for Economic Policy Discussion Paper 6514.
Kahn, C., McAndrews, J., and Roberds, W. (2003). 'Settlement Risk under Gross and Net Settlement', *Journal of Money, Credit and Banking*, 35: 591–608.
——— and Roberds, W. (2001). 'The CLS Bank: A Solution to the Risks of International Payments Settlements', *Carnegie Rochester Conference in Public Policy*, 54: 191–226.
Kane, E. (1989). *The S and L Insurance Mess: How Did it Happen*? Washington, DC: Urban Institute Press.
Karafiath, I., Mynatt, R., and Smith, K. L. (1991). 'The Brazilian Default Announcement and the Contagion Effect Hypothesis', *Journal of Banking & Finance*, 15: 699–716.
Kareken, J. H. and Wallace, N. (1978). 'Deposit Insurance and Bank Regulation: A Partial-Equilibrium Exposition', *Journal of Business*, 51: 413–38.
Kaufman, G. G. (1994). 'Bank Contagion: A Review of the Theory and Evidence', *Journal of Financial Services Research*, 7: 123–50.
Kho, B. C., Lee, D., and Stulz, R. (2000). 'US Banks Crises, and Bailouts: From Mexico to LTCM', *American Economic Review Papers and Proceedings*, 90: 28–31.
Kindleberger, C. P. (1978/1996), *Manias, Panics and Crashes. A History of Financial Crises*, 3rd edn., London: Macmillan.
King, M. and Wadhwani, S. (1990). 'Transmission of Volatility between Stock Markets', *Review of Financial Studies*, 3: 5–35.
Kobayakawa, S. (1997). 'The Comparative Analysis of Settlement Systems', Center for Economic Policy Discussion Paper 1667.
Krugman, P. (1998). 'What Happened to Asia?', mimeo., Massachusetts Institute of Technology, January.
Lehar, A. (2005). 'Measuring Systemic Risk: A Risk Management Approach', *Journal of Money, Credit and Banking*, 29: 2577–603.
Leitner, Y. (2005). 'Financial Network: Contagion, Commitment and Private Sector Bailouts', *Journal of Finance*, 60: 2925–53.
Lubloy, A. (2005). 'The Domino Effect on the Hungarian Interbank Market', *Kozgazdasagi Szemle* (Economic Review), 42: 377–401.
Madura, J. and McDaniel, W. R. (1989). 'Market Reaction to Increased Loan Loss Reserves at Money-Center Banks'. *Journal of Financial Services Research*, 3: 359–69.
——— Whyte, A. M., and McDaniel, W. R. (1991). 'Reaction of British Bank Share Prices to Citicorp's Announced $3 Billion Increase in Loan-loss Reserves', *Journal of Banking & Finance*, 15: 151–63.

McAndrews, J. J. and Roberds, W. (1995). 'Banks, Payments, and Coordination', *Journal of Financial Intermediation*, 4: 305–27.

—— and Wasilyew, G. (1995). 'Simulations of Failure in a Payment System', Working Paper 95-19.

Merton, R. C. (1977). 'An Analytical Derivation of the Cost of Deposit Insurance and Loan Guarantees: An Application of Modern Option Pricing Theory', *Journal of Banking and Finance*, 1: 3–11.

—— (1978). 'On the Cost of Deposit Insurance When There Are Surveillance Costs', *Journal of Business*, 51: 439–52.

Mian, A. and A. Sufi (2008). 'The Consequences of Mortgage Credit Expansion: Evidence from the 2007 Mortgage Default Crisis', *The Quarterly Journal of Economics*, forthcoming.

Minsky, H. P. (1977). 'A Theory of Systemic Fragility', in *Financial Crises*, E. I. Altman and A. W. Sametz, (eds.), New York, NY: Wiley, 138–52.

—— (1982). 'The Financial-Instability Hypothesis: Capitalist Processes and the Behaviour of the Economy', in *Financial Crises: Theory, History, and Policy*, C. P. Kindleberger and J.-P. Laffargue (eds.), Cambridge, UK: Cambridge University Press, 13–39.

Mishkin, F. S. (1991). 'Asymmetric Information and Financial Crises: A Historical Perspective', in *Financial Markets and Financial Crises*, G. Hubbard (ed.), Chicago, IL: University of Chicago Press.

Mistrulli, P. E. (2007). 'Assessing Financial Contagion in the Interbank Market: Maximum Entropy versus Observed Interbank Linkages', Banca d'Italia, *Temi di Discussione* 641, September.

Morgenstern, O. (1959). *International Financial Transactions and the Business Cycle*, National Bureau of Economic Research Studies in Business Cycles, Princeton, NJ: Princeton University Press.

Musumeci, J. J. and Sinkey, J. F. (1990). 'The International Debt Crisis, Investor Contagion, and Bank Security Returns in 1987: The Brazilian Experience', *Journal of Money, Credit, and Banking*, 22: 210–33.

Peavy, J. W. and Hempel, G. H. (1988). 'The Penn Square Bank Failure: Effect on Commercial Bank Security Returns: A Note', *Journal of Banking & Finance*, 12: 141–50.

Rajan, Raghuram (2006). 'Has Finance Made the World Riskier?', *European Financial Management* 12: 499–533.

Rancière, R., Tornell, A., and Westermann, F. (2008). 'Systemic Crises and Growth', *Quarterly Journal of Economics*, 123: 359–406.

Rochet, J.-C. and Tirole, J. (1996a). 'Interbank Lending and Systemic Risk', *Journal of Money, Credit, and Banking*, 28: 733–62.

—— —— (1996b). 'Controlling Risk in Payment Systems', *Journal of Money, Credit, and Banking*, 28: 832–62.

—— and Vives, X. (2004). 'Coordination Failures and the Lender of Last Resort: Was Bagehot Right After All?', *Journal of the European Economic Association.*

Ruckes, M. (2004). 'Bank Competition and Credit Standards', *Review of Financial Studies*, 17: 1073–112.

Saunders, A. (1986). 'An Examination of the Contagion Effect in the International Loan Market', *Studies in Banking and Finance*, 3: 219–47.

—— (1987). 'The Inter-Bank Market, Contagion Effects and International Financial Crises', in *Threats to International Financial Stability*, R. Portes and A. K. Swoboda (eds.), New York, NY: Cambridge University Press, 196–232.

Saunders, A. and Wilson, B. (1996). 'Contagious Bank Runs: Evidence from the 1929–33 Period', *Journal of Financial Intermediation*, 5: 409–23.

Scharfstein, D. S. and Stein, J. C. (1990). 'Herd Behaviour and Investment', *American Economic Review*, 80: 465–79.

Schoder, S. and Vankudre, P. (1986). 'The Market for Bank Stocks and Banks' Disclosure of Cross-Border Exposure', *Studies in Banking and Finance*, 3: 179–202.

Schoenmaker, D. (1995). 'A Comparison of Alternative Interbank Settlement Systems', L.S.E. Financial Markets Group Discussion Paper 204.

—— (1996). 'Contagion Risk in Banking', L.S.E. Financial Markets Group Discussion Paper 239.

Schwartz, A. J. (1986). 'Real and Pseudo-Financial Crises', in *Financial Crises and the World Banking System*, F. Capie and G. E. Wood (eds.), London: Macmillan, 11–31.

Senior Supervisors Group (2008). 'Observations on Risk Management Practices during the Recent Market Turbulence', 6 March.

Sheldon, G. and Maurer, M. (1998). 'Interbank Lending and Systemic Risk: An Empirical Analysis for Switzerland', *Swiss Journal of Economics and Statistics*, 134: 685–704.

Slovin, M. B., Sushka, M. E. and Polonchek, J. A. (1993). 'The Value of Bank Durability: Borrowers as Bank Stakeholders', *Journal of Finance,* 48: 247–66.

—— —— —— (1999). 'An Analysis of Contagion and Competitive Effects at Commercial Banks', *Journal of Financial Economics*, 54: 197–225.

Smirlock, M. and Kaufold, H. (1987). 'Bank Foreign Lending, Mandatory Disclosure Rules, and the Reaction of Bank Stock Prices to the Mexican Debt Crisis', *Journal of Business*, 60: 347–64.

Stiglitz, J. E. (1993). 'The Role of the State in Financial Markets', Paper Presented to the Annual World Bank Conference on Development Economics.

Swary, I. (1986). 'Stock Market Reaction to Regulatory Action in the Continental Illinois Crisis', *Journal of Business*, 59: 451–73.

Tirole, J. (2008a). 'Liquidity Shortages: Theoretical Underpinnings', Special Issue on Liquidity, *Banque de France Financial Stability Review*, 11, February.

—— (2008b). 'Leçons d'une crise', *Toulouse Sciences Economiques Notes.*

Upper, C. (2007). 'Using Counterfactual Simulations to Assess the Danger of Contagion in Interbank Markets', Bank for International Settlements Working Paper 234.

—— Worms, A. (2004), 'Estimating Bilateral Exposures in the German Interbank Market: Is There a Danger of Contagion?', *European Economic Review*, 48: 827–49.

US Department of the Treasury (2009). 'Financial Regulatory Reform—A New Foundation: Rebuilding Financial Supervision and Regulation', Washington, DC, 17 June.

Van Lelyveld, I. and Liedorp, F. (2006). 'Interbank Contagion in the Dutch Banking Sector', *International Journal of Central Banking*, 2(2): 99–133.

Van Rijckeghem, C. and Weder, B. (2003). 'Spillovers through Banking Centers: A Panel Data Analysis of Bank Flows', *Journal of International Money and Finance*, 22: 483–509.

Waldo, D. G. (1985). 'Bank Runs, the Deposit Currency Ratio and the Interest Rate', *Journal of Monetary Economics*, 15: 269–77.

Wall, L. D. and Peterson, D. R. (1990). 'The Effect of Continental Illinois' Failure on the Financial Performance of Other Banks', *Journal of Monetary Economics*, 26: 77–99.

Wells, S. (2004). 'Financial Interlinkages in the United Kingdom's Interbank Market and the Risk of Contagion', *Bank of England Working Paper*, 230.

Yamazaki, A. (1996). 'Foreign Exchange Netting and Systemic Risk', Institute for Monetary and Economic Studies Discussion Paper Series 96 <http://www.imes.boj.or.jp/english/publication/edps/edps96_index.html>.

CHAPTER 26

BANKING CRISES

GERARD CAPRIO JR
PATRICK HONOHAN[1]

Introduction

The great international banking crisis that broke out in August 2007 is only the latest episode in a lengthy history of recurrent banking crises around the world. Failures of banks have often been sudden—with depositors scrambling to withdraw their funds or refusing to renew their maturing deposits. They have been costly, both in direct cash costs to bank creditors or to the governments who have bailed them out, and indirectly in the associated spillover effects on economic activity including that caused by reduced access to credit. Some financial crises have had their focus elsewhere, as in government debt, exchange rate, and stock market crises, but banks have typically played a central or important supporting role.

Although bank solvency is often the victim of adverse shocks arising elsewhere in the economy, and while panic can result in unnecessarily large and damaging depositor withdrawals, this chapter argues that the most damaging of systemic banking crises—including but not limited to the current one—have ultimately involved or were significantly exacerbated by what we call bad banking and bad policies—those that permitted or encouraged excessive risk taking and even

[1] We would like to thank Thorsten Beck, Roger Bolton, Stijn Claessens, Asli Demirgüç-Kunt, James Hanson, Luc Laeven, Philip Lane, Millard Long, Peter Montiel, Steven Nafziger, Sergio Schmukler, and Andrew Sheng for comments. Nonetheless, the responsibility for any errors and omissions lies with the authors.

'looting' of other people's money. With each crisis there is an inevitable chorus of calls for more official prudential regulation and supervision to prevent a recurrence. However, cross-country empirical evidence suggests that policy is best directed toward ensuring a dynamic approach to regulation focusing on the information that is being disclosed to market participants, the degree of market discipline on the behavior of bankers, and the incentives in the financial system, including those for regulators.

The second section briefly sketches the historical background, noting the 'boom in busts' of the post-Bretton Woods period following a thirty-year lull. Not all crises are the same and the third section highlights the distinct role of mismanagement, government interference, and macroeconomic shocks. The fourth section reviews the aspects of crises which have received attention from economic theoreticians seeking to understand their recurrence and severity. The fifth section discusses the costs of crises. The size of these explains the importance of prevention and corrective policy and these are discussed in the sixth section. In conclusion, the seventh section suggests that despite an inevitable overhaul of regulation in the coming years, crises will recur periodically, and the goal should be to minimize their frequency and cost without sacrificing the benefit to economic growth and income equality that a well-functioning financial system can deliver.

Early history

It is no exaggeration to say that banking crises—for now, the widespread insolvency of banks leading to closures, mergers, takeovers, or injections of government resources—are virtually as old as banking. When modern banking emerged as a development of money-changing in thirteenth-century Europe, bankers faced information problems more severe than in the least developed countries today. Clients' trade was subjected to a variety of shocks—wars, plague, shortage of coins, losses in trade (e.g., ships sinking or being plundered), defalcation by borrowers, etc.—that made lending hazardous. And depositors faced the risk that their bankers would not survive these shocks, or would themselves abscond with funds. Repeated failures led to some drastic remedies: a Barcelonan banker was executed in front of his failed bank in 1360—a far cry from the limited liability that protected bank owners in later times (Kohn, 2009: chap. 8). Sovereigns were less likely to impose such extreme sanctions when they were the source of the problem, and bankers often succumbed to the temptation or were required (literally for their survival) to lend to the monarch. Such famous early Italian banking houses as the Riccardi of Lucca, the Bardi, the Peruzzi, and even the illustrious Medici of

Florence, owed their banking downfall in whole or large part to kings and princes that would not or could not repay. Financing the loser in a war was a sure route to failure, but even winners reneged, leading to a higher interest rate spread on loans to kings and princes than to the more business-minded town governments (Homer and Sylla, 1996: 94).

That bank failures have come in waves is suggested by the list assembled by Kindleberger (1978, and with Aliber, 2005) and covering mostly the more advanced economies since the seventeenth century, and which displays, for example, the rather regular ten-yearly recurrence of crises through most of the nineteenth century and through to World War II. Emerging economies experienced a higher frequency of crises in the interwar period (Bordo, et al., 2001). The post-World War II era saw a period of exceptional quiescence that lasted through the early 1970s. Against the background of a relatively benign macroeconomic environment, regulations that restricted banking competition and product innovation, including cross-border activities, probably contributed to this stability. Gradually, however, these regulations became unsustainable as communications technology and financial innovation (including the emergence of nearbank competitors) led to evasion.

Liberalization of banking and of capital flows, together with increasingly volatile macroeconomic conditions (themselves associated with weakened fiscal discipline, the abandonment of the Bretton Woods exchange rate pegs and surges in inflation rates) were followed by a return to banking crises at a frequency comparable to what had been experienced before. Already by 1997, over three out of every five member states of the IMF had experienced banking problems severe enough to be regarded as systemic or at least borderline systemic (Lindgren, Garcia, and Saal, 1996; and Caprio, et al., 2005). But the etiology of these crises varied.

Diverse origins: Management, government, and macroeconomics in recent crises

Many of the most spectacular systemic banking crises of recent decades have been inextricably linked with macroeconomic crises in a way that makes the direction of causality hard to unravel. However, it is important not to neglect the role of fraud and mismanagement, on the one hand, and government interference, on the other. Indeed, one or the other of these two—bad banking and bad policies[2]—has been at the

[2] We use 'bad banking' to embrace a range of management practice, from fraud to miscalculations of risk to deliberate exploitation of the put option inherent in deposit insurance, that heightens the

root of quite a number of systemic banking crises, not only in the developing world (Honohan, 1997; and Caprio and Honohan, 2005) but also in the latest great international crisis that emerged in the US and Europe.

Two very large individual bank failures in the Caribbean area can be taken as classic examples where *fraud or mismanagement* were at the root of the problem—namely, that in Venezuela (1994) and the Dominican Republic (2003). Both appear to be cases of the diverted deposits fraud, in which some of the deposits accepted by the bank are not recorded as liabilities and the corresponding resources are looted by insiders even though the bank still appears solvent on paper and even though its recorded assets may be properly performing. In each of these cases, the bank involved was of systemic importance[3] and the sums were so large that the loans that eventually were made by the central bank to enable the bank to make the depositors whole, destabilized the macroeconomy. And, in Venezuela, high deposit rates in the 'rogue' bank forced up rates, and risk taking, at other banks. Another very large failure in which the diverted deposits fraud appears to have been present was that of the international group BCCI. This group, headquartered in Luxembourg and London, was operating in about seventy countries and its failure was of systemic importance in some African countries where it had attained a sizable market share (cf. Herring, 2005). The diverted deposits fraud typically involves the acquiescence of audit professionals; the official supervisor can then be hard-pressed to detect such frauds because of the complexity of the false accounting structures that are created.

Inadequate management of 'rogue' traders has caused several sizable bank failures, most famously that of Barings Bank in 1995, but, although the losses involved in some of these cases have run into ten figures, no known cases have been of systemic importance. In January 2008, Société Générale reported the largest single bank loss (over $7 billion) ever attributed to fraud by a lone 'rogue' trader. As with the even larger Ponzi scheme uncovered in Bernard Madoff's investment firm in late 2008, typically fraud is discovered in a period of asset market decline following a long run of over-optimism. Other forms of mismanagement weakness can be cited, none more apparent than the case of Crédit Lyonnais in the 1990s, where grandiosity and exaggerated ambition in lending policy led to the largest single bank loss in the industrial world: without the French government's bailout,

likelihood of bank failure. Of course, all banking involves risk, not least because of the ever-present information problems of adverse selection and moral hazard, but these are managed and adequately priced in normal banking operation. Pressure of circumstances can turn good bankers into bad bankers, as is graphically characterized by de Juan (2002).

[3] The payments system can create a strong short-term interdependency of banks, so that the failure of one major bank could disrupt the entire system of payments and short-term credit on which much of day-to-day economic activity depends. For this reason, some banks of systemic importance are perceived as being TBTF, requiring official support for their continued operation even if they are insolvent.

Crédit Lyonnais would have proved insolvent. Lack of management capacity on the part of new controlling insiders also brought insolvency in 1995 to the long-established Meridien BIAO bank in Western and Central Africa—although that bank had already been severely weakened by the effects of government intervention.

While the Mexican Tequila crisis (1994–5) crystallized around a currency collapse, which hit the banks because of speculative derivative contracts that gave them a de facto long position on local currency, the underlying weakness of the Mexican banks was subsequently traced to insider lending and a long period of evasion of minimum capitalization requirements dating back to their privatization. With little shareholder equity at stake, banks were free to move out on the risk frontier and lend to the few sectors with the highest return, as confirmed by Caprio and Wilson (2000), Haber (2005), and Wilson, Saunders, and Caprio (2000).

Significant regime changes in the economy often devalue both the financial and skills portfolio of banks, sharply increasing the risk of a banking crisis. The introduction of new instruments or opportunities for risk taking often leads some to take on new risks without adequate attention to their downside potential. Likewise, liberalization of economic policies has definitely been associated with a surge of bank failures in countries with weaker information and governance institutions (Demirgüç-Kunt and Detragiache, 1999). Liberalization of entry into banking increased competitive pressures for banks, liberalization of interest rates heightened repayment and market risks, and liberalization of other aspects of economic policy impacted on the creditworthiness of borrowers in ways that were not always easy to perceive, often entailing large changes in relative prices. And, to the extent that pre-liberalization portfolios were controlled, the lifting of controls often led banks to expand simultaneously. However, simultaneous portfolio shifts by the banking sector can move asset prices, making the shift look like a safe proposition, as in the case of the Malaysian property boom of the late 1970s and early 1980s, which led to a mid-1980s crisis. In addition to a skewed portfolio, liberalized banks inherit a staff that is short on banking skills, unfortunately precisely when they are greatly needed, just as the government begins with bank supervisors skilled only in checking that banks are complying with various government commands and not at all trained in modern risk-based bank supervision. Although even the best bankers and supervisors would be challenged during liberalization, those with weak skills are even more likely to fail.

In particular, the process of economic transition from socialist or planned economies proved fertile for banking crises, many of which can be attributed to inexperienced or reckless management. Although the first wave of post-transition inflation wiped out much of the real value of their pre-existing deposits, and reduced the debt burden of their borrowers, many transition economy banks—especially in Eastern Europe—misjudged the difficulty of credit appraisal especially in the fluid conditions of the transition. As a result, many made a new round of poor or self-serving loans, which soon fell into non-performing status.

Even where Transition was managed without a surge of high inflation, as in China and Vietnam, large banking losses were socialized. Indeed, in China, cumulative injections of government funds into the four main government-owned banks alone from 1998 to 2006 amounted to over $350 billion, or about 30 percent of 2001 GDP, with further injections still considered necessary to restore full capitalization on a realistic evaluation of the recoverability of the loan portfolio (see Barth and Caprio, 2007; and Honohan, 2008). This massive bailout was accomplished without loss of depositor confidence, reflecting the ability and undisputed willingness of the State to ensure that depositors at its banks would not suffer. Indeed, expressed as a percentage of GDP, bank deposits in China have been higher than almost anywhere else in the developing world, aside from offshore financial centers. These growing funds were effectively applied up to the mid-1990s as a transitional and partial substitute for the former budgetary allocations made under the planned system to key unprofitable state-owned enterprises (Lardy, 1998). Made as loans, these could never have been fully serviced, as was gradually recognized through the various bank restructuring measures adopted from 1998 on. The Chinese case, then, provides a conspicuous example of how government policy—specifically government-directed lending policy—has led to loan losses large enough to erode the banks' capital many times over.

Many of the poorest developing-country economies that were not subject to a centrally planned regime also experienced explicit or implicit government policies of directed credit. When these were enforced by statist regimes without regard to the viability of the lending banks, the result was losses, erosion of capital, and a weakening of financial autonomy and motivation of bank managers, often resulting in insolvency. The true financial condition of state-owned or heavily controlled banks of this sort was often acknowledged only at a time of regime change or a sizable policy reform. Even in non-socialist economies, government influence has often had similar effects. A good example comes from francophone West Africa where the banks in several countries made what proved to be unrecoverable loans to parastatals and government suppliers, unwisely taking comfort in the fact that these loans were being rediscounted by the regional central banks. A similar problem arises with provincial governments relying on the national authorities to bail out failing provincial banks, as was seen in Brazil.

Banks have always been dependent to a degree on the willingness of the state to allow them to function profitably. Even where directed credit is not an issue, quasi-fiscal impositions such as unremunerated reserve requirements have weakened bank profitability. Arbitrary exchange rate and exchange-control regulations also have a tax-like effect. The most dramatic example of this was the forced conversion to local currency of foreign currency deposits and loans at Argentine banks in late 2001. Because the conversion was not at market rates and furthermore was asymmetric, with a much larger effective write-down of bank loans than of bank deposits, this arbitrary measure created systemic bank insolvency at a stroke.

Although the roles of management and government are never irrelevant in a banking crisis, what has dominated many of the larger episodes of systemic crisis is a dynamic instability in widely held expectations about macroeconomic and business prospects generally. A wave of *over-optimism* about economic growth, often manifested in a real estate price boom, results in expansion of credit by most banks, especially to the sectors specifically favored by the optimism. The resulting increase in leverage often is fuelled in part by capital inflows—as in Mexico and East Asia in the 1990s, but also in the recent mortgage finance booms in the US, Ireland, and elsewhere. Because of the optimism, loan–loss provisioning is lower than will prove necessary, and this for a time is justified by low delinquencies as the overall economic boom financed by credit expansion makes it easy for borrowers to service their debt. This could explain by itself why rapid credit expansion is a predictor of crises. In addition, of course, rapid credit expansion places stresses on credit appraisal capacity and results in errors even conditional on the overall optimism. Various forms of contagion or herd effect come into play. Even banks whose managers do not share the optimism feel pressure to relax credit approval standards for fear of losing market share. The formation of banker expectations can be influenced by peer observation, magnifying and generalizing emerging overconfidence. As a latecomer to the South Sea Bubble (John Martin, of Martin's Bank) said, 'when the rest of the world are mad we must imitate them in some measure' (Dale, 2004: 113), words that were echoed in July 2007 by the soon-to-be-sacked CEO of Citigroup, Chuck Prince, who told the *Financial Times* that 'As long as the music is playing, you've got to get up and dance'.

Whereas experienced bankers are normally alert to isolated indications of unsound practices among their peers, in contrast, during the euphoria of the boom phase, they are unlikely to detect even fatal weaknesses. These waves of over-optimism are sufficiently rare in any one country for learning to be imperfect. Disaster myopia prevails, with decision-makers disregarding the relevance of historical experience at home and abroad (Guttentag and Herring, 1986). Eventually, however, the unsustainability of the fundamentals on which the credit expansion was predicated becomes evident and the process goes into reverse. Sharp falls in property prices reveal the unrecoverability of property-related loans and erode the value of collateral, currency depreciation creates insolvency among unhedged borrowers, asset sales by distressed borrowers seeking liquidity drive down the prices of other securities too, and the resulting economic disruption also undermines the solvency of borrowers in unrelated sectors.

Previous examples of the boom-and-bust syndrome are provided by the correlated crises in Scandinavia around 1990, as well as the East Asian crisis of 1997–8, in which extensive failure of banking systems especially in Thailand, Indonesia, and Korea were associated with currency collapse and a sharp—albeit transitory—contraction of economic activity following a long period of rapid growth and capital inflows. The sudden withdrawal of what had previously been readily

available foreign funds was an aggravating factor in several other crises, notably Chile in 1982. Exchange rate collapse, too, has been a feature in many episodes; indeed, anticipations of currency movements during crises can result in sizable depositor withdrawals exacerbating bank liquidity problems. To be sure, in all of these cases, connected lending and excessive risk taking were a good part of the story, as they often are in large crises (World Bank, 2001; and Harvey and Roper, 1999).

Although some features of the international crisis that began in 2007, such as the role of derivative securities, seemed new, this crisis in fact displays many familiar features (cf. Reinhart and Rogoff, 2008). In particular, it exhibits a wave of over-optimism leading to extreme leverage, and unsound management and regulatory responses to financial innovation. At its center were the growing market in US-originated mortgage-backed securities and the boom in housing prices in many industrial countries. Provided by Basel I with a clear incentive to reduce required capital by shifting loans off their balance sheet, and notwithstanding the well-understood adverse selection problem that historically had limited loan sales, banks in the US and other countries had increasingly turned to an 'originate and distribute' model, in which standardized loans, mostly mortgages, could be bundled and sold as securities without recourse to the originating bank, thereby leaving that institution free to reuse its capital elsewhere.[4] Non-depository financial intermediaries jumped into the same business, given the ability to earn fees and yet not retain credit risk. By careful structuring of these securities and in particular their priority in receiving cash flow from the servicing of the original portfolios, favorable credit ratings were obtained for most of the securities sold, seemingly overcoming the adverse selection problem that had hitherto prevented such loan sales (buyers assumption that sellers would only part with their worst loans). However, knowing that the loans they originated would be sold to others reduced the incentive to make careful credit assessment. Indeed, US banks and finance companies originated a large number of high-risk mortgages (e.g., 'no money down', interest only or less as the initial payment, with no documentation on borrowers' capacity to pay and initial 'teaser' interest rates that would adjust upwards even if market rates remained constant).

Rating agencies seemed to become the partners of those doing the securitization, rather than serving as unbiased arbiters of credit quality. The initial ratings they had attached to the securities proved over-optimistic and most had to be downgraded sharply. As the US housing market cooled and rates adjusted (from teaser levels, and then with the tightening of monetary policy), defaults spread leading to sizable loan losses for most of the world's leading international banks, and inducing them to raise additional capital to strengthen their balance sheets. Thanks to

[4] According to the Basel system, various loans and other assets were assigned different risk weights, thereby leading to the incentive to shed assets with a higher risk charge.

securitization, US banks had retained only part of the mortgage risk, passing much of it to European and other banks and investment funds. Indeed, the first bank failures from the US subprime mortgages were two German banks which had taken unwarranted risks in this market.

Banks were not really able to assess the risk of the increasingly complex securities that were being created. This very complexity destroyed information and has made resolution and workout of distressed debt enormously more difficult and uncertain. The dramatic decline in the market value of these securities both illustrated and precipitated a growing revulsion for complex and risky lending, and a retreat to liquidity leading to a global 'credit crunch'.

Even after the existence of a bank solvency crisis has been publicly acknowledged, the scale of the crisis is rarely evident at first. Bank insiders have many reasons to conceal weaknesses as long as possible. Almost all recent systemic crises have involved several waves of intervention, generally spread over a period of months or even years.

Panic and contagion: Explaining sudden and fast-moving banking crises

A sudden and irresistible depositor run, the classic form in which systemic crises have been seen as crystalizing, and which dominates the theoretical literature, has actually only featured in a minority of recent cases. Even in Argentina, 1995, the response of depositors to fears of a spillover from Mexico's 1994 Tequila event, aggregate depositor withdrawals from the system were little more than 20 percent, spread over several months. In this case, when depositor concerns shifted from the health of specific banks to the prospects for the currency peg, they exited the system altogether. This pattern was repeated in 2001, only then depositors were justified in that the government did subsequently abandon the currency peg.

But, even if depositor runs are not as common as a reading of textbooks would suggest, the sudden onset of correlated bank failures that have characterized some systemic banking crisis with widespread consequences for economic activity raises the question of what is special about banks that might make banking systems prone to such dramatic collapses.

Five distinctive and interrelated features of banking stand out as contributing factors to this vulnerability. First, the highly leveraged nature of modern banks; second, the degree of maturity transformation (or liquidity creation) with which they are associated; third, the demandable or very short-term nature of the bulk of their liabilities; fourth, the opaque nature of bank assets; and fifth, the fact that the

bulk of their assets and liabilities are denominated in fiat currency. Of course, each of these features represents a key contribution of banking to the economy, which is probably part of the explanation as to why authorities have not adopted proposals for 'narrow' banking—few are disposed to give up these benefits.

That high leverage has a role seems obvious: it is why much policy effort focuses on limiting leverage through capital adequacy regulation (even though the risk-reducing goal of such regulation can often be nullified by bankers' offsetting assumption of higher risks in unregulated dimensions). Opacity also matters: just as banks are at an informational disadvantage vis-à-vis borrowers, so too are depositors and other creditors (as well as supervisors) in relation to banks. Much recent theory has developed around the second and third of these features (Allen and Gale, 2008a). It is not just the liquidity problems that can arise if depositors wish to withdraw more than expected from a bank that has committed its resources to loans that can be liquidated early only at a loss. There is the consideration that even depositors who have no immediate need to withdraw might do so if they foresee a bank failure. The possibility of self-fulfilling depositor panics not based on any fundamental change in the bank's asset portfolio or any special liquidity shock to its depositors has been known to theoreticians for decades, though the real-world relevance of self-fulfilling panics unwarranted by weak fundamentals has been much debated. From this theoretical perspective, there is no difference between the visible retail depositor run and the 'silent run' of the bank's wholesale creditors, including other banks through the interbank market. Indeed, in practice it is often the better-informed wholesale market that undermines a failing bank's liquidity and, as in the case of Northern Rock in 2007, leads to a run in the retail market. Better-informed wholesale market participants might have reason to suspect that the bank's problem is less liquidity and more solvency. In theory, liquidity runs can lead to insolvency by forcing a 'fire sale' of assets at unfavorable prices, but in practice it is difficult to distinguish this case from insolvency due to excessive risk taking.

One structural feature of banking implicated in panics is the demandable nature of deposit liabilities, which has the effect of encouraging early withdrawals (Calomiris and Kahn, 1991). It is 'first come, first served' for bank depositors (known as 'sequential service' in the theoretical literature). Until an insolvent bank closes its doors, early withdrawing depositors will receive their full deposit, paid out of the bank's liquid assets; while those that arrive too late will bear between them the full capital deficiency. Even a small overall initial deficiency could result in the remaining depositors suffering severe losses if enough others have withdrawn before the bank is closed. Awareness of this risk makes astute depositors alert to signs of trouble and indeed serves to ensure that there will be an incentive for large depositors to monitor the performance of the bank managers. As is confirmed by well-documented cases such as that of Continental Illinois bank (Stern and Feldman, 2004), as well as from less precise information from the

changing size distribution of deposits in crises in developing countries (Schmukler and Halac, 2005), it is wholesale depositors and interbank lenders who have been the first to withdraw.

Some system-wide bank failures may be simply due to numerous banks being hit by a common shock external to the banking system. But the speed with which several very large systemic crises have emerged without apparent warning and the depth of the ensuing financial and economic crisis has suggested a contagious transmission and amplification of the problems of one bank to others. Furthermore, even if the failure of a number of banks is attributable to an exogenous macroeconomic shock, the consequences of that failure on aggregate credit availability and on the value of asset prices may in turn amplify the macroeconomic downturn feeding back again into the banking system.

Models of contagion focus on different aspects. Contagion can occur through depositor panic, as the failure of one bank causes a reassessment by depositors' of the default risks associated with other banks, and the loss of liquidity from one bank failure may cause depositors to withdraw from other banks in the system. At the broader national level, both such factors seem to have been at work in the international crises of 1997–8 and in the liquidity and 'credit crunch' of 2007–9. Information and fears can be transmitted through several distinct channels including the prices of bank equity, credit default swaps, and the secondary market in bank debt, as well as ratings announcements. Regulations suspending short-selling of bank equities, introduced during 2008, reflect official suspicion of market manipulation in some of these markets at times of panic.

On the asset side too, bank distress can be transmitted through the system. If it forecloses on some of its borrowers or is unable to extend credit, a bank's distress will be spread to the customers of those borrowers in turn worsening the loan-loss experience of other banks. The weakening of asset portfolios will become general if there is a scramble for liquidity in asset markets, which drives down prices including of assets used as collateral. Pure informational cascades, where pessimistic opinions of the part of some bankers or investors become generalized, have also been studied as channels of contagion. The use by banks of the same or similar mechanical risk assessment technologies could have the unfortunate effect of coordinating banks' responses to shocks, thereby amplifying their effect (International Monetary Fund, 2007). Indeed, the depth of the 'credit crunch' in 2007–9 reflects the correlated realization by leading bankers that the risk management paradigm that they all shared had failed.

Models of such feedback can exhibit multiple equilibria: a good equilibrium in which investors' confidence is validated by high asset prices boosting the creditworthiness of borrowers with productive and profitable investments, and a bad equilibrium where investors' skepticism is justified by low asset prices, a lack of creditworthiness, weak aggregate demand, and business and bank insolvency. The equilibrium value of the nominal or real exchange rate is at the heart of several of

these models, reflecting the central role of currency collapses in some of the largest crises. If there are multiple equilibria, the occurrence of a crisis can be considered a coordination failure (Diamond and Dybvig, 1983; and Allen and Gale, 2008a).

Costs of crises

Two approaches have been adopted to calculating the cost of banking crises. The first approach focuses narrowly on the revealed capital deficiency of the banks and specifically on the fiscal and quasi-fiscal costs incurred by efforts to indemnify depositors of failing institutions. The other approach has sought to calculate system-wide economic costs of the failure. The two approaches have generated rather different figures for specific events, though on average across countries they come up with roughly similar total costs, expressed as a percentage of GDP. Thus, taking thirty-nine systemic crises for which both economic costs and fiscal costs have been calculated, the fiscal costs—ranging up to 55 percent of GDP (Argentina, 1982)—averaged 12.5 percent, whereas the estimated economic costs averaged 14.6 percent. The correlation between the two sets of costs was only 0.43, however (Hoggarth, Reis, and Saporta, 2002; and Honohan and Klingebiel, 2003). (In view of the large government guarantees and central bank asset purchases that have been undertaken, and the rapidly-changing estimates of GDP in the coming years, it is too soon to give even an approximate figure for the fiscal or economic costs of the current international crisis in the countries most severely affected.)

Neither approach to measuring costs is wholly satisfactory. The fiscal costs approach refers to what in principle is a concrete concept, though changing prices, exchange rates, and asset values in the months and years following the crisis greatly complicate the calculation. For example, favorable property price movements in Norway and Sweden allowed the authorities to recover most if not all of the outlays they had initially made in respect of failing banks. To the extent that the sums expended by the authorities are to fill resource gaps resulting from loss-making economic activity by borrowers, the fiscal costs can be considered as an estimate of true economic costs. But, since some of the fiscal outlays simply go to compensate depositors for resources that were diverted to others, and as such represent a transfer, this would overstate true economic costs. On the other hand, the distortions created by poor banking practice will have affected decision-making more widely, resulting in losses and missed opportunities that are not captured in the fiscal costs.

Attempts to measure true economic costs from analysis of a dip in growth rates around the time of the crisis lack credibility to the extent that the economic downturn (which exposed the bank insolvencies) may have been triggered by unrelated factors. To attribute all of the downturn to the banking problems probably overstates the costs. On the other hand, some episodes have not been followed by an economic downturn. These include cases where the impact on economic growth was spread over a long number of years. Thus, the calculations are sensitive to the conjectural nature of the counterfactual macroeconomic growth path against which the actual is compared. Many crises are preceded by an economic boom, part of which was attributed to the excess optimism in banking and in other sectors. Since some part of the boom might have had sound foundations, backing out the sustainable path is no simple exercise.

Even if it is hard to get a precise estimate, it is clear that the aggregate costs of banking crises around the world have been very substantial indeed. Total fiscal costs of crises in developing countries since the 1970s exceeds US $1 trillion—a sum far in excess of all development aid provided by the advanced economies. The economic costs of crises have been felt across the income spectrum with sharp increases in the fraction of the population below the poverty line (Honohan, 2005; and World Bank, 2001). Notwithstanding these costs, some countries—Chile and Korea, for example—have seen their financial system recover nicely from even large crises. Unfortunately other countries, notably Argentina, have had numerous crises in the last 150 years, pointing to a sizable, even critical, benefit from the application of good policies of prevention, containment, and resolution.

Crisis response and prevention

An ounce of prevention

The design of regulatory policy and practice that could most effectively reduce the risk of banking crises is controversial. The Basel Committee on Bank Supervision, established in 1974, has emerged as a standard setter for bank regulation and supervision. In the Basel II Revised Capital Accord, to be implemented in 2008 and beyond in many countries, the Committee's approach to prudential regulation involves three pillars: capital, supervision, and disclosure. The first pillar defines a minimum amount of capital to be held by banks in relation to the risks that they have assumed; the second pillar is a supervisory regime to ensure compliance with this capital minimum and generally discourage excessive risk taking; the third pillar mandates disclosure of relevant accounting information.

Unfortunately, Basel II's approach to setting required capital is highly controversial (Keating, et al., 2001) not only because of the difficulty of measuring the underlying risks, but because reliance on the mandated approaches could exacerbate herding to the extent that banks adopt similar approaches to modeling risk. Furthermore, cross-country empirical evidence casts considerable doubt on the merits of relying on discretionary action by official supervisors to limit banking failure. Specifically, Barth, Caprio, and Levine (2006) show that this approach does *not* seem to help prevent banking crises. Using their database on bank regulation and supervision around the world, this study compiled indexes that represented the extent of capital regulation, supervisory powers, market monitoring (effectively, the three pillars of Basel II) and other regulatory variables, and related them to the development, efficiency, vulnerability, integrity (lack of corruption), and governance of the banking system, after controlling for other determinants of the latter variables and also dealing with concerns about endogeneity. On vulnerability, they found that none of the three pillars explained the probability of a banking crisis (though private monitoring helped explain the other endogenous variables of interest). Instead, this research indicates that authorities concerned with reducing the likelihood of a crisis should either not adopt or greatly circumscribe deposit insurance, and should encourage banks to diversify both their activities and their geographic and sectoral exposure. Lack of such diversification helps explain the large number of failures in the US (roughly 15,000 bank failures in the period 1920–33), compared with Canada (just 1 in the period). Although this research is by no means the last word on banking crisis, it highlights an approach to regulation that in effect tries to work with market forces, rather than supplant them.

Prevention would be easier if the onset of crises could be predicted, but models are better at showing fragility than predicting timing (Demirgüç-Kunt and Detragiache, 2005). With no effective forecasting system, good containment and resolution policies are also needed to deal with the next crisis when it comes.

A pound of cure

When a crisis hits, government has two key roles: as the lender of last resort (LOLR) and as organizer or participant in the restructuring of troubled entities. The threat of contagion among banks has led many policymakers to intervene to stop a run before healthy banks and borrowers are impaired. Central banks have accepted the role as LOLR since the early nineteenth century, though not uniformly or without contention (Wood, 2003). The advice from Bagehot, that the LOLR should lend freely but at a penalty rate and only to solvent institutions with good collateral, has become conventional wisdom, if not always followed, and his additional lessons—lend quickly before a run takes off, and only use the LOLR rarely to avoid moral hazard—also are regularly quoted by central bankers. This seemingly

straightforward advice is notoriously difficult to apply in practice, as it involves judgments on collateral, solvency, and speed.[5]

Longer-term restructuring and rehabilitation of banks raises issues that go beyond the scope of this chapter (Honohan and Laeven, 2005; and World Bank, 2001). In the spirit of Bagehot, it is worth noting that once authorities decide to intervene it is important that their intervention be comprehensive, dealing with all potential problem banks especially where depositors fear that they will suffer from bank closures. The failure of the initial policy 1997 bank restructuring package in Indonesia (according to the announcement of which only sixteen banks would be closed—both a much smaller number than had been expected by business opinion and than subsequently proved necessary) has been attributed to its less than comprehensive nature. Soon, all of the private banks were run, with depositors putting their funds in what they assumed were safe public banks. The central bank then extended liquidity support to the private banks, who appear to have used the funds to buy foreign exchange, exacerbating the decline of the currency (a chronology of the events by the IMF experts involved can be found in Enoch, et al., 2001). In several crises in Argentina, the public would run to public sector and foreign banks, from the domestic private banks.

Crisis management during 2007–9 evolved slowly. Despite the persistent alarming indications, government intervention in insolvent banks and loss allocation were slow to start, reactive on a case-by-case basis, and piecemeal.

At first, the main efforts were devoted to addressing the illiquidity that suddenly emerged in the interbank and other short-term money markets from August 2007. Progressively over the following eighteen months central banks bought and accepted as collateral an increasingly large and varied range of assets. A few banks were rescued by public authorities during the early months, but they were treated as poorly managed outliers, which had become excessively exposed to the mispriced securities backed by badly underwritten mortgages, rather than as symptomatic of a wider solvency problem. Even with the failure in March 2008 of the important investment bank Bear Stearns, policy remained on a case-by-case basis, including (i) the guarantees provided to the large US Government-sponsored

[5] LOLR actions need an effective communications strategy if they are to be successful in restoring depositors' confidence. When the UK mid-sized mortgage lender Northern Rock in 2007 had difficulty in refinancing its mortgage portfolio in the wholesale markets and was given exceptional liquidity support by the Bank of England (eventually amounting to the equivalent of about US $50 billion, larger than any previous such loan in history), the tone of the accompanying statements seems to have triggered a retail depositor run so unnerving that the authorities issued a temporary open-ended depositor guarantee. It is too soon to know if Northern Rock was solvent at the time of its first request. If it were, the authorities' initial hesitation to assist may have been inconsistent with Bagehot's rule; if not, it demonstrates the difficulty for the LOLR when insolvent banks are not promptly closed before a run begins. This recent case also illustrates the importance of encouraging banks to manage carefully their risks, including liquidity positions, which frequent LOLR support will undermine.

wholesale mortgage banks Fannie Mae and Freddie Mac; (ii) the decision to allow the investment bank Lehman Brothers go into bankruptcy; and (iii) the rescue of the large insurance company AIG which had underwritten credit default swap contracts on a systemic scale. Together, these ad hoc actions, which occurred within a couple of weeks in September, destabilized market expectations, as did the Congressional reaction to the US Treasury's sketchy proposal to use vast budgetary resources to buy from the banks the most opaque securities. The following weeks were marked by sharp declines in stock market values, a widening of credit spreads and the failure or near-failure of several large banks in Europe and the US. The prospect of a generalized freezing of the banking system triggered coordinated international decisions in mid-October 2008 to make public funds available systematically and on a large scale to recapitalize the main banks. This action, which led, for example, to the UK government assuming a majority stake in one of its largest banks, the Royal Bank of Scotland, reassured markets of governments' intention to rescue any other large banks getting into difficulties—as indeed was confirmed by a package announced in late November 2008 for Citigroup.

The delay in recognizing the scale of the solvency issues and the piecemeal official response probably exacerbated the growth in risk aversion in and around banks during 2008. By the end of that year credit remained extremely tight (with interbank rates well above the equivalent swap rates). The 'credit crunch' had begun to depress economic activity worldwide, even in countries whose banks had not been implicated in the excesses and errors of US and European markets. This deepening global turndown began in turn to have a feedback effect on bank loan losses.

In almost all crises, a sizable fraction of the banking system has survived, remaining solvent and liquid (Caprio and Honohan, 2005). An exception: all but one of the seven banks in Guinea, accounting for 98 percent of the banking assets in the country were deemed insolvent and closed following massive frauds. Interestingly, the one bank left open failed several years later. Another exception is Iceland, where all three of the main banks failed in 2008; in this case there was no adequately resourced LOLR because the banks had expanded into international business, denominated in foreign exchange, and had gross liabilities of the order of ten times Iceland's GDP. Although luck can play a part in survival, that some banks typically survive points to the potential for well-managed banks to cope with severe shocks, and to the importance of maintaining an incentive structure that encourages safe-and-sound banking. But, the survivors may not be easily able or willing to expand to fill the gap that would be created if the failed banks are removed from the system and often become more conservative in their lending decisions. Indeed, post-crisis 'credit crunches' are significant contributors to the macroeconomic dips noted earlier, which was one of the reasons for the exhortations of Bagehot.

Toward dynamic regulation

In the years leading up to the current crisis, the regulatory approach in industrial countries, as embodied by the work of the Basel Committee on Bank Supervision, was taken as a model for others. Quantitative risk management was on the rise, and the Basel Committee reflected this by moving away from crude, arbitrary risk weights decided by regulators (Basel I) to a more modern approach that placed reliance either on weights derived from the judgments of credit rating organizations (CROs) or the risk models of the banks themselves. Yet, the locus of failures in the current crisis prominently featured the wildly over-optimistic ratings of the CROs and the utter failure of risk management modeling to protect the banking system. Similarly, prompt corrective action, the key regulatory change in the wake of the US S&L crisis, was too often dormant—neither prompt, nor corrective. Modern financial instruments, which were supposed to parcel out risk to those who could best bear it, were pushed beyond their capacity by users who ignored the danger that variances and correlations might not be stationary over time, and indeed that they might be endogenous to the common behavior of bankers. As a result, these tools instead contributed to a reduction in information and in the incentive to monitor risk.

Although his analysis tends to caricature history, it is striking how many of the features stressed by Minsky (1986) as being typical of banking crises have been present in recent events. Low interest rates did make many debtors look good, and their subsequent rise regularly revealed 'surprises'. Financial safety nets, not limited to the boom in deposit insurance since the 1980s, did increase risk taking. And large current account imbalances contain their own risks, as they are sooner or later eroded by a slowdown in the dynamism of the surplus party (whether that be a UK or a China) or a revision of opinion as to the creditworthiness of the borrower (an Argentina or the US).

But any notion that it is possible to come up with a static set of rules to govern the sector and then leave the system to operate on automatic pilot should be dismissed. Such an approach can too easily be blindsided by some distinctively new and unexpected features, such as—this time—the hidden weaknesses of risk management for highly engineered financial instruments. Finance cannot return to the highly controlled and segmented world of the 1930s. Regulatory arbitrage needs to be acknowledged as a fact of life, and one made infinitely easier thanks to inexpensive communications and computing.

Instead, finance demands a more dynamic approach to regulation. Regulation should be geared to revealing information, not sequestering it in the files of the supervisor, and to be revealing, especially in boom times, the risks that are being taken and the compensation of the risk takers. By publication or by regulatory consequences, financial firms need to be dissuaded from paying out huge rewards

out of current profits. Those managing other people's money need to face the consequences of their decisions rather than be protected by their fulfillment of requirements that they hold highly rated paper. And, those with deeper pockets who lend to intermediaries need to experience losses when they have made poor decisions. In other words, crises focus attention on incentives, and the less regulation is incentive-based, the larger will be the losses to society.

References

Allen, F. and Gale, D. (2006). *Understanding Financial Crises,* New York: Oxford University Press.

Barth, J. and Caprio Jr., G. (2007). 'China's Changing Financial System: Can it Catch Up With, or even Drive Growth', *Milken Review*, 9/3.

——— and Levine, R. (2004). 'Bank Regulation and Supervision: What Works Best', *Journal of Financial Intermediation*, 12: 205–48.

——— ——— (2006). *Rethinking Bank Regulation: Till Angels Govern,* New York: Cambridge University Press.

——— ——— (2007). Changing Bank Regulation: For Better or for Worse? mimeo, http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:20345037~pagePK:64214825~piPK:64214943~theSitePK:469382,00.html, July.

Bordo, M., Eichengreen, B., Klingebiel, D., and Martinez-Peria, M. (2001). 'Is the Crisis Problem Growing More Severe?', *Economic Policy*, 32: 51–82.

Calomiris, C. W. (2008). 'Banking Crises', in S. Durlauf and L. Blume (eds.). *The New Palgrave Dictionary of Economics*, New York: Macmillan.

—— and Kahn, C. (1991). 'The Role of Demandable Debt in Structuring Optimal Banking Arrangements', *American Economic Review*, 81: 497–513.

—— and Mason. J. (2001). 'Causes of U.S. Bank Distress during the Depression', *Bank Structure and Competition Conference Proceedings.* Chicago: Federal Reserve Bank of Chicago.

Caprio Jr., G. and Honohan, P. (2005). 'Starting Over Safely: Rebuilding Banking Systems', in G. Caprio Jr., J. A. Hanson, and R. E. Litan (eds.). *Financial Crises: Lessons from the Past, Preparation for the Future.* Washington, DC: Brookings Institution Press.

—— Klingebiel, D., Laeven, L., and Noguera, G. (2005). 'Banking Crisis Database', in P. Honohan and L. Laeven (eds.). *Systemic Financial Crises.* Cambridge: Cambridge University Press.

—— and Wilson, B. (2000). 'Financial Fragility and Mexico's 1994 Peso Crisis: An Event-Window Analysis of Market Valuation Effects', *Journal of Money, Credit, and Banking*, 32: 450–68.

Dale, R. (2004) *The First Crash: Lessons from the South Sea Bubble*, Princeton, NJ: Princeton University Press.

De Juan, A. (2002). 'From Good Bankers to Bad Bankers', in G. Caprio, P. Honohan, and D. Vittas (eds.). *Financial Sector Policy for Developing Countries: A Reader.* Washington, DC: The World Bank.

Demirgüç-Kunt, A. and Detragiache, E. (1999). 'Financial Liberalization and Financial Fragility', in B. Pleskovic and J. E. Stiglitz (eds.), *Proceedings of the 1998 World Bank Conference on Development Economics*. Washington, DC: World Bank.

—— —— (2005). 'Cross-Country Empirical Studies of Systemic Bank Distress: A Survey', *National Institute Economic Review*, 192.

Diamond, D. and Dybvig, P. (1983). 'Bank Runs, Deposit Insurance, and Liquidity', *Journal of Political Economy*, 91: 401–19.

Enoch, C., Baldwin, B., Frecaut, O., and Kovanen, A. (2001). 'Indonesia: Anatomy of a Banking Crisis: Two Years of Living Dangerously 1997–99', IMF Working Paper 01/52.

Fung, B. J. G., Hohl, S., and Ma, G. (2004). 'Public Asset Management Companies in East Asia: A Comparative Study'. Financial Stability Institute Occasional Paper 3.

Furman, J. and Stiglitz, J. E. (1998). 'Economic Crises: Evidence and Insights from East Asia', *Brookings Papers on Economic Activity*, 1–135.

Goldstein, M. (2005). 'The Next Emerging-Market Financial Crisis: What Might it Look Like?, in G. Caprio Jr., J. A. Hanson, and R. E. Litan, (eds.). *Financial Crises: Lessons from the Past, Preparation for the Future*, Washington, DC: Brookings Institution Press.

Guttentag, J. M. and R. J. Herring (1986). 'Disaster Myopia in International Banking', *Princeton Essays in International Finance*, 164, September.

Haber, S. (2005). 'Mexico's Experiments with Bank Privatization and Liberalization, 1991–2003', *Journal of Banking & Finance*, 29: 2325–53.

Halac, M. and Schmukler, S. L. (2004). 'Distributional Effects of Crises: The Financial Channel', *Economía* 5: 1–67.

Harvey, C. and Roper, A. (1999). 'The Asian Bet', in A. Harwood, R. E. Litan, and M. Pomerleano (eds.). *The Crisis in Emerging Financial Markets*, Washington, DC: Brookings Institution Press.

Herring, R. J. (2005). 'BCCI and Barings: Bank Resolutions Complicated by Fraud and Global Corporate Structure', in D. Evanoff and G. Kaufman (eds.). *Bank Resolutions and Financial Stability*, Cambridge, MA: MIT Press.

Homer, S. and Sylla, R. (1996). *A History of Interest Rates*, New Brunswick, NJ: Rutgers University Press.

Hoggarth, G., Reis, R., and Saporta, V. (2002). 'Costs of Banking System Instability: Some Empirical Evidence', *Journal of Banking & Finance*, 26: 857–60.

Honohan, P. (1997). 'Banking System Failures in Developing and Transition Countries: Diagnosis and Prediction', Bank for International Settlements Working Paper 39, January, http://www.bis.org/publ/work39.htm

—— (2005). 'Banking Sector Crises and Inequality', World Bank Policy Research Working Paper WPS 3659.

—— (2008). 'Protecting Depositors in China: Experience and Evolving Policy', in A. Demirgüç-Kunt, E. Kane, and L. Laeven (eds.). *Deposit Insurance*, Cambridge, MA: MIT Press.

—— and Klingebiel, D. (2003). 'Controlling the Fiscal Costs of Banking Crises', *Journal of Banking & Finance*, 27: 1539–60.

—— and Laeven, L. (2005). *Systemic Financial Distress: Containment and Resolution*, New York: Cambridge University Press.

International Monetary Fund (2007). *The Global Financial Stability Review*, Washington, DC: The International Monetary Fund.

KEATING, C., SHIN, H. S., GOODHART, C., and DANIELSSON, J. (2001). *An Academic Response to Basel II*. Financial Markets Group Special Paper 130, London School of Economics.

KINDLEBERGER, C. P. (1978). *Manias, Panics, and Crashes: A History of Financial Crises*, New York: John Wiley and Sons. (2005 edition by C. P. Kindleberger and R. Z. Aliber, New York: Palgrave MacMillan.)

KOHN, M. (forthcoming). *The Origins of Western Economic Success: Commerce, Finance, and Government in Pre-Industrial Europe*, http://www.dartmouth.edu/~mkohn/.

LARDY, N. R. (1998). *China's Unfinished Economic Revolution*. Washington, DC: Brookings Institution Press.

LINDGREN, C. J., GARCIA, G., and SAAL, M. (1996). *Bank Soundness and Macroeconomic Policy*. Washington, DC: International Monetary Fund.

MARTINEZ-PERIA, M. S. and SCHMUKLER, S. (2001). 'Do Depositors Punish Banks for "Bad" Behavior? Market Discipline, Deposit Insurance and Banking Crises', *Journal of Finance*, 56: 1029–51.

MINSKY, HYMAN P., (1986). *Stabilizing an Unstable Economy*, New Haven: Yale University Press.

RADELET, S. and SACHS, J. (1998). 'The Onset of the East Asian Financial Crisis', National Bureau of Economic Research Working Paper 6680, August.

REINHART, C. M. and ROGOFF, K. S. (2008). 'Is the 2007 U.S. Subprime Financial Crisis So Different? An International Historical Comparison', *American Economic Review*, 98: 339–44.

STERN, G. H. and FELDMAN, R. J. (2003). *Too Big to Fail: The Hazards of Bank Bailouts*. Washington, DC: Brookings Institution Press.

WILSON, B., SAUNDERS, A., and CAPRIO Jr., G. (2000). 'Mexico's Financial Sector Crisis: Propagative Links to Devaluation', *Economic Journal*, 110: 292–308.

WOOD, J. H. (2003). 'Bagehot's Lender of Last Resort: A Hollow Hallowed Tradition', *The Independent Review*, 7: 343–51.

World Bank (2001). *Finance for Growth: Policy Choices in a Volatile World*, Washington, D.C. and New York: The World Bank and Oxford University Press.

CHAPTER 27

THE GREAT DEPRESSION AND OTHER 'CONTAGIOUS' EVENTS

CHARLES W. CALOMIRIS

Introduction

Concerns about the susceptibility of banks to unwarranted withdrawals of deposits during panics, the possibility of bank failures, and contractions of bank credit resulting from unwarranted withdrawals of deposits (which is sometimes described as the result of 'contagious' weakness among banks) and the attendant adverse macroeconomic consequences of bank disappearance or bank balance sheet contraction have motivated much of the public policies toward banks. The global financial crisis of 2007–9 was the most recent illustration of this phenomenon (Calomiris 2008). In reaction to initial bank losses (e.g., on subprime mortgage-related exposures), a scramble for liquidity ensued in which banks reduced their lending and scrambled to shore up their liquidity and reduce their leverage. Interest rate spreads on risky assets skyrocketed, and money market instruments (commercial paper, interbank deposits, and repurchase agreements) contracted sharply, adding to the 'liquidity crunch'.

Several policies have come into existence to deal with such shocks, including assistance mechanisms intended to protect banks from unwarranted withdrawals of deposits (central bank lending during crises, deposit insurance, and government-sponsored bank bailouts), and a host of prudential regulatory policies (intended to promote banking system stability, and especially to prevent banks from taking advantage of government protection by increasing their riskiness—the so-called 'moral-hazard' problem of protection). This chapter reviews the theory and historical evidence related to the prevalence of banking contagion and the effects of the policies designed to mitigate it.

'Contagion' vs. fundamentals as causes of bank failures

Theoretical models have been devised in which banking crises result from systemic 'contagion', when banks that are intrinsically solvent are subjected to large unwarranted withdrawals, and may fail as a consequence of this withdrawal pressure. Advocates of the view that banking systems are inherently vulnerable to such contagion often emphasize that the structure of banks—the financing of illiquid assets with demandable debts, and the 'sequential service constraint' (which mandates that depositors who are first in line receive all of their deposits)—tends to aggravate the tendency for unwarranted withdrawals (see Diamond and, 1983; Allen and Gale, 2000; and Diamond and Rajan, 2002).

Unwarranted withdrawals (that is, those unrelated to the solvency of the bank) can occur, in theory, for a number of reasons. Diamond and Dybvig (1983) develop a banking model with multiple equilibria, where one of the equilibria is a systemic bank run, which occurs simply because depositors believe that others will run. More generally, observers of historical panics sometimes document depositors imitating each other's withdrawal behavior; depositors may line up to withdraw their funds simply because others are doing so, particularly in light of the incentives implied by the sequential service constraint. It is important to recognize, however, that evidence about mimetic withdrawals does not generally confirm the all-or-nothing runs by all depositors imagined by some theoretical models; rather, mimesis may be partial and gradual (see O'Grada and White, 2003; and Bruner and Carr, 2007).

A second possibility, which is particularly relevant for understanding pre-World War I banking panics in the US (e.g., the nationwide US Panics of 1857, 1873, 1884, 1890, 1893, 1907, and some events during the Great Depression, including the Chicago banking panic of June 1932) is that a signal is received by

depositors, which contains noisy information about the health of the various banks. Depositors have reason to believe that a loss has occurred that might cause a bank to become insolvent, but they cannot observe which bank has suffered the loss. In that circumstance, depositors may withdraw large amounts of funds from all banks, including those that are (unobservably) solvent, simply because they would rather not risk leaving their money in a bank that turns out to be insolvent.

Third, exogenous shocks to depositors' liquidity preferences, or to the supply of reserves in the banking system, unrelated to banks' asset condition, may cause an excess demand for cash on the part of depositors relative to existing reserves, which can lead banks to a scramble for reserves, which can produce systemic runs (a banking version of the game 'musical chairs'). Liquidity demand and supply shocks may be related to government policies affecting the reserve market, or to foreign exchange risks that lead depositors to want to convert to cash. This mechanism may have had a role in some banking system crises (notably, the nationwide US Panics of 1837 and 1933).

Withdrawal pressures, whether they are associated with warranted or unwarranted withdrawals, can accumulate over time or can take the extreme form of a 'bank run' (when depositors decide en masse to remove deposits). Some financial historians (notably Friedman and Schwartz, 1963) have pointed to the Great Depression of the 1930s as a time when unwarranted depositor withdrawals, and sometimes 'runs' or 'panics', led to large numbers of bank failures, and rapid declines in deposits of solvent and insolvent banks alike. Bank distress is associated not only with bank failures, but with general macroeconomic consequences resulting from the reduced supply of loans and deposits, which can amplify business cycle downturns and spread panic-induced financial distress from banks to the whole economy (Bernanke, 1983; and Calomiris and Mason, 2003b). Other episodes of banking panics outside the Great Depression have also been identified as possible episodes of unwarranted bank failures, especially in the US during the nineteenth and early twentieth centuries, with similar inferences drawn by some about the contagious causes and costly consequences of bank distress.

Another view of banking distress (which I will label the 'fundamentalist', as opposed to the 'panic', approach), stresses a different direction of causality: a chain of causation from non-panic-related, observable, exogenous adverse changes in the economic conditions of banks, to intrinsic weakening of bank condition, ultimately leading to bank failure. According to this view, fundamental losses to bank borrowers cause losses to banks, which may bankrupt some banks and lead other weakened banks to curtail the supplies of loans and deposits as part of a rebalancing of portfolios to limit default risk in a disciplined market (Calomiris and Wilson, 2004). Endogenous contractions of deposits and loans, just like unwarranted contractions, will limit the supply of money and credit, and thus

they will exacerbate the macroeconomic decline that caused them. Thus, according to the fundamentalist view, banking distress can magnify economic downturns even if banks are not the originators of shocks; banks will tend to magnify macroeconomic shocks through their prudential decisions to curtail the supplies of loans and deposits in response to adverse shocks, even if banks are passive responders to shocks and even if depositors avoid engaging in unwarranted runs or panics.

Differences in opinion about the sources of shocks that cause bank failures have important implications for policy. While both the panic and fundamentalist views can be used to motivate public policy to protect banks (since both views see banks as important magnifiers of macroeconomic disturbance), the panic view provides special motives for public policies to protect banks from withdrawal risk. The fundamentalist view, in contrast, sees banks as inherently stable—that is, neither victims of unwarranted withdrawals, nor a major source of macroeconomic shocks. According to the fundamentalist view, market discipline of banks is not random, and indeed, helps preserve efficiency in the banking system. It may be desirable to limit or even avoid government protection of banks to preserve market discipline in banking (making banks more vulnerable to the risk of depositor withdrawal). Preserving market discipline encourages good risk management by banks, even though bank deposit and credit contractions attendant to adverse economic shocks to bank borrowers may aggravate business cycles. Indeed, some empirical studies have argued that policies that insulate banks from market discipline tend to produce worse magnifications of downturns, due to excessive bank risk taking in response to protection (e.g., Boyd, et al., 2000; and Barth, Caprio, and Levine, 2006).

These two views of the sources of bank distress (the panic view that banks are fragile and highly subject to panic, or, alternatively, the fundamentalist view that banks are stable and generally not subject to unwarranted large-scale withdrawals) do not define the universe of possibilities. One or the other extreme view may do a better job explaining different historical crises, and both fundamentals and unwarranted withdrawals may play a role during some banking crises. The recent empirical literature on banking crises has tried to come to grips with the causes and effects of systemic bank failures in different places and times, to ascertain the dominant causal connections relating banking distress and macroeconomic decline, and to try to draw inferences about the appropriate public policy posture toward banks. The remainder of this chapter selectively reviews the empirical literature on the causes of bank failures during systemic banking crises. This review begins with a lengthy discussion of the Great Depression in the US, which is followed by a discussion of US bank distress prior to the Depression, historical bank distress outsides the US, and contemporary banking system distress (which is discussed more fully in Chapter 26 of this volume, by Caprio and Honohan).

US BANK DISTRESS DURING THE GREAT DEPRESSION

The list of fundamental shocks that may have weakened banks during the Great Depression is a long and varied one. It includes declines in the value of bank loan portfolios produced by waves of rising default risk in the wake of regional, sectoral, or national macroeconomic shocks to bank borrowers, as well as monetary-policy-induced declines in the prices of the bonds held by banks. There is no doubt that adverse fundamental shocks relevant to bank solvency were contributors to bank distress; the controversy is over the size of these fundamental shocks—that is, whether banks experiencing distress were truly insolvent or simply illiquid.

Friedman and Schwartz (1963) are the most prominent advocates of the view that many bank failures resulted from unwarranted 'panic' and that failing banks were in large measure illiquid rather than insolvent. Friedman and Schwartz's emphasis on contagion imagined that bank failures mainly reflected a problem of illiquidity rather than insolvency. Illiquid but solvent financial institutions, in their view, failed purely as the result of withdrawal demands by depositors, particularly during sudden moments of panic. In contrast, an insolvent institution fails to repay depositors as the result of fundamental losses in asset value, rather than the suddenness of depositor withdrawals.

Friedman and Schwartz attach great importance to the banking crisis of late 1930, which they attribute to a 'contagion of fear' that resulted from the failure of a large New York bank, the Bank of US, which they regard as itself a victim of panic. They also identify two other banking crises in 1931—from March to August 1931, and from Britain's departure from the gold standard (21 September 1931) to the end of the year. The fourth and final banking crisis they identify occurred at the end of 1932 and the beginning of 1933, culminating in the nationwide suspension of banks in March 1933. The 1933 crisis and suspension was the beginning of the end of the Depression, but the 1930 and 1931 crises (because they did *not* result in suspension) were, in Friedman and Schwartz's judgment, important sources of shock to the real economy that turned a recession in 1929 into the Great Depression of 1929–33.

The Friedman and Schwartz argument is based upon the suddenness of banking distress during the panics that they identify, and the absence of collapses in relevant macroeconomic time series prior to those banking crises (see charts 27–30 in Friedman and Schwartz, 1963: 309). But there are reasons to question Friedman and Schwartz's view of the exogenous origins of the banking crises of the Depression. As Temin (1976) and many others have noted, the bank failures during the Depression marked a continuation of the severe banking sector distress that had gripped agricultural regions throughout the 1920s. Of the nearly 15,000 bank disappearances that occurred between 1920 and 1933, roughly half pre-date 1930.

And massive numbers of bank failures occurred during the Depression era outside the crisis windows identified by Friedman and Schwartz (notably, in 1932). Wicker (1996: 1) estimates that '[b]etween 1930 and 1932 of the more than 5,000 banks that closed only 38 percent suspended during the first three banking crisis episodes'. Recent studies of the condition of the Bank of US indicate that it too may have been insolvent, not just illiquid, in December 1930 (Joseph Lucia, 1985; and Wicker, 1996). So there is some prima facie evidence that the banking distress of the Depression era was more than a problem of panic-inspired depositor flight.

How can one attribute bank failures during the Depression mainly to fundamentals when Friedman and Schwartz's time series evidence indicates no prior changes in macroeconomic fundamentals? Friedman and Schwartz omitted important aggregate measures of the state of the economy relevant for bank solvency—for example, measures of commercial distress and construction activity may be useful indicators of fundamental shocks. Second, aggregation of fundamentals masks important sectoral, local, and regional shocks that buffeted banks with particular credit or market risks. The empirical relevance of these factors has been demonstrated in the work of Wicker (1980; 1996) and Calomiris and Mason (1997; 2003a).

Using a narrative approach similar to that of Friedman and Schwartz, but relying on data disaggregated to the level of the Federal Reserve districts and on local newspaper accounts of banking distress, Wicker argues that it is incorrect to identify the banking crisis of 1930 and the first banking crisis of 1931 as national panics comparable to those of the pre-Fed era. According to Wicker, the proper way to understand the process of banking failure during the Depression is to disaggregate, both by region and by bank, because heterogeneity was very important in determining the incidence of bank failures.

Once one disaggregates, Wicker argues, it becomes apparent that at least the first two of the three banking crises of 1930–1 identified by Friedman and Schwartz were largely regional affairs. Wicker (1980; 1996) argues that the failures of November 1930 reflected regional shocks and the specific risk exposures of a small subset of banks, linked to Nashville-based Caldwell and Co., the largest investment bank in the South at the time of its failure. Temin (1989: 50) reaches a similar conclusion. He argues that the 'panic' of 1930 was not really a panic, and that the failure of Caldwell and Co. and the Bank of US reflected fundamental weakness in those institutions.

Wicker's analysis of the third banking crisis (beginning in September 1931) also shows that bank suspensions were concentrated in a very few locales, although he regards the nationwide increase in the tendency to convert deposits into cash as evidence of a possible nationwide banking crisis in September and October 1931. Wicker agrees with Friedman and Schwartz that the final banking crisis (of 1933), which resulted in universal suspension of bank operations, was nationwide in scope. The banking crisis that culminated in the bank holidays of February–March 1933

resulted in the suspension of at least some bank operations (bank 'holidays') for nearly all banks in the country by 6 March.

From the regionally disaggregated perspective of Wicker's findings, the inability to explain the timing of bank failures using aggregate time series data (which underlay the Friedman–Schwartz view that banking failures were an unwarranted and autonomous source of shock) would not be surprising even if bank failures were entirely due to fundamental insolvency. Failures of banks were local phenomena in 1930 and 1931, and so may have had little to do with national shocks to income, the price level, interest rates, and asset prices.

The unique industrial organization of the American banking industry plays a central role in both the Wicker view of the process of bank failure during the Depression, and in the ability to detect that process empirically. Banks in the US (unlike banks in other countries) did not operate throughout the country. They were smaller, regionally isolated institutions. In the US, therefore, large region-specific shocks might produce a sudden wave of bank failures in specific regions even though no evidence of a shock was visible in aggregate macroeconomic time series (see the cross-country evidence in Bernanke and James, 1991; and Grossman, 1994). The regional isolation of banks in the US, due to prohibitions on nationwide branching or even statewide branching in most states, also makes it possible to identify regional shocks empirically through their observed effects on banks located exclusively in particular regions.

Microeconomic studies of banking distress have provided some useful evidence on the reactions of individual banks to economic distress. White (1984) shows that the failures of banks in 1930 are best explained as a continuation of the agricultural distress of the 1920s, and are traceable to fundamental disturbances in agricultural markets.

Calomiris and Mason (1997) study the Chicago banking panic of June 1932 (a locally isolated phenomenon). They find that the panic resulted in a temporary contraction of deposits that affected both solvent and insolvent banks, and, in that sense, unwarranted deposit contraction did occur. Fundamentals, however, determined which banks survived. Apparently, no solvent banks failed during that panic. Banks that failed during the panic were observably weaker *ex ante*, judging from their balance sheet and income statements, and from the default risk premiums they paid on their debts. Furthermore, the rate of deposit contraction was not identical across banks; deposits declined more in failing weak banks than in surviving banks.

Calomiris and Wilson (2004) study the behavior of New York City banks during the interwar period, and, in particular, analyze the contraction of their lending during the 1930s. They find that banking distress was an informed market response to observable weaknesses in particular banks, traceable to *ex ante* bank characteristics. It resulted in bank balance sheet contraction, but this varied greatly across banks; banks with higher default risk were disciplined more by the market (that is,

experienced greater deposit withdrawals), which encouraged them to target a low-risk of default.

Calomiris and Mason (2003a) construct a survival duration model of Fed member banks throughout the country from 1929 to 1933. This model combines aggregate data at the national, state, and county level with bank-specific data on balance sheets and income statements to identify the key contributors to bank failure risk and to gauge the relative importance of fundamentals and panics as explanations of bank failure. Calomiris and Mason find that a fundamentals-based model can explain most of the failure experience of banks in the US prior to 1933. They identify a significant, but small, national panic effect around September of 1931, and some isolated regional effects that may have been panics, but, prior to 1933, banking panics were not very important contributors to bank failures compared to fundamentals.

The fact that a consistent model based on fundamentals can explain the vast majority of US bank failures prior to 1933 has interesting implications. First, it indicates that the influence of banking panics as an independent source of shock to the economy was not important early in the Depression. Only in 1933, at the trough of the Depression, did failure risk become importantly de-linked from local, regional, and national economic conditions and from fundamentals relating to individual bank structure and performance. Second, the timing of this observed rise in risk unrelated to indicators of credit risk is itself interesting. In late 1932 and early 1933, currency risk became increasingly important; depositors had reason to fear that President Roosevelt would leave the gold standard, which gave them a special reason to want to convert their deposits into (high-valued) dollars before devaluation of the dollar (Wigmore, 1987). Currency risk, of course, is also a fundamental.

It is also interesting to connect this account of bank distress during the Depression—which emphasizes fundamental shocks, rather than simply illiquidity, as the source of bank distress—with the history of lender of last resort (LOLR) assistance to banks during the Depression. Many commentators have faulted the Federal Reserve for failing to prevent bank failures with more aggressive discount window lending. While it is certainly true that expansionary monetary policy, particularly in 1929–31, could have made an enormous difference in preventing bank distress (through its effects on macroeconomic fundamentals), that is not the same as saying that more generous terms at the discount window (holding constant the overall monetary policy stance) would have made much of a difference. Discount window lending only helps preserve banks that are suffering from illiquidity, which was not the problem for most banks in the 1930s that were experiencing large depositor withdrawals.

Indeed, in 1932, President Hoover created the Reconstruction Finance Corporation (RFC), to enlarge the potential availability of liquidity, but this additional source of liquidity assistance made no difference in helping borrowing banks avoid

failure (Mason, 2001). Commentators at the time noted that, because the collateralized RFC and Fed loans were senior to deposits, and because depositor withdrawals from weak banks reflected real concerns about bank insolvency, loans from the Fed and the RFC to banks experiencing withdrawals did nothing to help, and actually often did harm to banks, since those senior loans from the Fed and the RFC reduced the amount of high quality assets available to back deposits, which actually increased the riskiness of deposits and created new incentives for deposit withdrawals.

In 1933, however, once the RFC was permitted to purchase preferred stock of financial institutions (which was junior to depositors), RFC assistance to troubled banks was effective in reducing the risk of failure (Mason, 2001). Finland enjoyed similar success with its use of preferred stock in the early 1990s. Preferred stock injections were not so successful in resolving Japanese bank distress in 1999 and 2000, which reflected the magnitude of the Japanese banks' problems, problems in the implementation of the program, and the limitations of preferred stock injections for helping resolve problems of deep bank insolvency (Calomiris and Mason, 2004; Calomiris, 2009). Preferred stock injections had limited beneficial effects on large global banks during the 2007–9 crisis, perhaps for similar reasons.

Microeconomic studies of local contagion

As part of their bank-level analysis of survival duration during the Depression, Calomiris and Mason (2003a) also consider whether, outside the windows of 'panics' identified by Friedman and Schwartz, the occurrence of bank failures in close proximity to a bank affects the probability of survival of the bank, after taking into account the various fundamental determinants of failure. Calomiris and Mason recognize that this measure of 'contagious failure' is an upper bound, since in part it measures unobserved cross-sectional heterogeneity common to banks located in the same area, in addition to true contagion. They find small, but statistically significant, effects associated with this measure. The omission of this variable from the analysis raises forecasted survival duration by an average of 0.2 percent. They also consider other regional dummy variables associated with Wicker's (1996) instances of identified regional panics, and again find effects on bank failure risk that are small in national importance.

O'Grada and White (2003) provide a detailed account of depositor behavior based on individual account data during the 1850s for a single bank, the Emigrant Savings Bank of New York, which offers a unique perspective on depositor

contagion during banking panics. In 1854, Emigrant experienced an unwarranted run that can be traced to mimetic behavior among inexperienced, uninformed depositors. This run, however, was easily handled by the bank, which was able to pay off depositors and restore confidence. In contrast, the run in 1857 was an imitative response to the behavior of informed, sophisticated depositors who were running for a reason, and that run resulted in suspension of convertibility. Furthermore, in both of these episodes, mimesis was not sudden: 'In neither 1854 nor 1857 did depositors respond to a single signal that led them to crowd into banks all at once. Instead, panics lasted a few weeks, building and sometimes ebbing in intensity, and only a fraction of all accounts were closed' (O'Grada and White, 2003: 215). O'Grada and White show that contagion can be a real contributor to bank distress, but they also show that runs based on random beliefs tend to dissipate with little effect, while runs based on legitimate signals tend to grow in importance over time. The fact that runs are not sudden, and that many depositors do not participate in them at all, is important, since it implies the ability of events to unfold over time; that is, for a form of collective learning among depositors to take place during panics.

A similar account of mimetic withdrawals based on a random rumor can be found in an article by Nicholas in *Moody's Magazine* in 1907. A bank in Tarpen Springs, Florida experienced an unwarranted outflow of deposits based on a false rumor that was spread through the local Greek-American community, which included many of the bank's depositors. The bank quickly wired to have cash sent from its correspondent bank, which arrived in time to prevent any suspension of convertibility, and brought the run to an end. Nicholas noted that, if the bank had really been in trouble, not only would the correspondent not have provided the funds, but it and other banks would have probably withdrawn any funds it had on deposit at the bank long before the public was aware of the problem (a so-called 'silent run'; see the related discussions in Halac and Schmukler, 2004; and Stern and Feldman, 2003).

US BANK DISTRESS IN THE PRE-DEPRESSION ERA

As many scholars have recognized for many years, for structural reasons, US banks were unusually vulnerable to systemic banking crises that saw large numbers of bank failures before the Depression, compared to banks in other countries (for reviews, see Bordo, 1985, and Calomiris, 2000). Calomiris and Gorton (1991) identify six episodes of particularly severe banking panics in the US between the

Civil War and World War I, and prior to the Civil War, there were other nationwide banking crises in 1819, 1837, and 1857. In the 1920s, the US experienced waves of bank failures in agricultural states, which have always been identified with fundamental shocks to banks, rather than national or regional panics. Other countries, including the US's northern neighbor, Canada, however, did not suffer banking crises during these episodes of systemic US banking system distress. The key difference between the US and other countries historically was the structure of the US banking system. The US system was mainly based on unit banking—geographically isolated single-office banks; no other country in the world imitated that approach to banking, and no other country experienced the US pattern of periodic banking panics prior to World War I, or the waves of agricultural bank failures that gripped the US in the 1920s.

Canada's early decision to permit branch banking throughout the country ensured that banks were geographically diversified and thus resilient to large sectoral shocks (like those to agriculture in the 1920s and 1930s), able to compete through the establishment of branches in rural areas (because of low overhead costs of establishing additional branches), and able to coordinate the banking system's response in moments of confusion to avoid depositor runs (the number of banks was small, and assets were highly concentrated in several nationwide institutions). Coordination among banks facilitated systemic stability by allowing banks to manage incipient panic episodes to prevent widespread bank runs. In Canada, the Bank of Montreal occasionally would coordinate actions by the large Canadian banks to stop crises before the public was even aware of a possible threat.

The US was unable to mimic this behavior on a national or regional scale (Calomiris, 2000; and Calomiris and Schweikart, 1991). US law prohibited nationwide branching, and most states prohibited or limited within-state branching. US banks, in contrast to banks elsewhere, were numerous (e.g., numbering more than 29,000 in 1920), undiversified, insulated from competition, and geographically isolated from one another, thus were unable to diversify adequately or to coordinate their response to panics (US banks did establish clearing houses in cities, which facilitated local responses to panics beginning in the 1850s, as emphasized by Gorton, 1985).

The structure of US banking explains why the US uniquely suffered banking panics despite the fact that the vast majority of banks were healthy, and were able to avoid ultimate failure. Empirical studies show that the major US banking panics of 1857, 1873, 1884, 1890, 1893, 1896, and 1907 were moments of heightened asymmetric information about bank risk. Banking necessarily entails the delegation of decision-making to bankers, who specialize in screening and monitoring borrowers and making non-transparent investments. Bankers consequently have private information about the attendant risks. During normal times, the risk premium banks pay in capital markets and money markets contains a small 'opacity' premium—part of the risk depositors and bank stockholders face and charge for

comes from not being able to observe the value of bank assets moment to moment—that is, not being able to mark bank portfolios to market. During the US panics, the normally small opacity premium became very large, as people became aware that risks had increased and as they also were aware of what they *didn't* know—namely, the incidence among banks of the probable losses that accompanied the observable increased risk.

Calomiris and Gorton (1991) show that banking panics were uniquely predictable events that happened at business cycle peaks. In the pre-World War I period (1875–1913), every quarter in which the liabilities of failed businesses rose by more than 50 percent (seasonally adjusted) and the stock market fell by more than 8 percent, a panic happened in the following quarter. This happened five times, and the Panic of 1907 was the last of those times. Significant national panics (i.e., events that gave rise to a collective response by the New York Clearing House) never happened otherwise during this period.

Bank failure rates, even during these panic episodes, were small, and the losses to depositors associated with them were also small. In 1893, the panic with the highest failure rate and highest depositor loss rate, depositor losses were less than 0.1 percent of GDP (Calomiris, 2007). *Expected* depositor losses during the panics also appear to have been small. Oliver Sprague (1910: 57–8, 423–4) reports that the discount applied to bankers' cashier checks of New York City banks at the height of the Panic of 1873 did not exceed 3.5 percent and, with the exception of an initial ten-day period, remained below 1 percent, and a similar pattern was visible in the Panic of 1893. A 1 percent premium would be consistent with depositors in a New York City bank estimating a 10 percent chance of a bank's failing with a 10 percent depositor loss if it failed. Clearly, banking panics during this era were traceable to real shocks, but those shocks had small consequences for bank failures in the aggregate and even at the height of the crisis those consequences were expected to be small. Historical US panics teach us that even a small expected loss can lead depositors to demand their funds, so that they can sit on the sidelines until the incidence of loss within the banking system has been revealed (usually a process that took a matter of weeks).

Bank failure rates in the 1830s and the 1920s were much higher than those of the other pre-Depression systemic US banking crisis episodes. The 1830s saw a major macroeconomic contraction that caused many banks to fail, which historians trace to large fundamental problems that had their sources in government-induced shocks to the money supply (Rousseau, 2002), unprofitable bank-financed infrastructure investments that went sour (Schweikart, 1988), and international balance of payments shocks (Temin, 1969). The 1920s agricultural bank failures were also closely linked to fundamental problems—in this case, the collapses of agricultural prices at the end of World War I, which were manifested in local bank failures in the absence of regional or national bank portfolio diversification (Calomiris, 1992; and Alston, Grove, and Wheelock, 1994).

Other historical experiences with bank failures

Although the US was unique in its propensity for panics, it was not the only economy to experience occasional waves of bank failures historically. Losses (i.e., the negative net worth of failed banks), however, were generally modest and bank failure rates were much lower outside the US. The most severe cases of banking distress during this era, Argentina in 1890 and Australia in 1893, were the exceptional cases; they suffered banking system losses of roughly 10 percent of GDP in the wake of real estate market collapses in those countries. The negative net worth of failed banks in Norway in 1900 was 3 percent and in Italy in 1893 1 percent of GDP, but with the possible exception of Brazil (for which data do not exist to measure losses), there were no other cases from 1875 to 1913 in which banking losses in a country exceeded 1 percent of GDP (Calomiris, 2007).

Loss rates tended to be low because banks structured themselves to limit their risk of loss by maintaining adequate equity-to-assets ratios, sufficiently low asset risk, and adequate liquidity. Market discipline (the potential for depositors fearful of bank default to withdraw their funds) provided incentives for banks to behave prudently (for a theoretical framework, see Calomiris and Kahn, 1991). The picture of small depositors lining up around the block to withdraw funds has received much attention by journalists and banking theorists, but perhaps the more important source of market discipline was the threat of an informed ('silent') run by large depositors (often other banks). Banks maintained relationships with each other through interbank deposits and the clearing of deposits, notes, and bankers' bills. Banks often belonged to clearing houses that set regulations and monitored members' behavior. A bank that lost the trust of its fellow bankers could not long survive.

Bank failures in the late twentieth century

Recent research on systemic bank failures has emphasized the destabilizing effects of bank safety nets. This has been informed by the experience of the US Savings and Loan industry debacle of the 1980s, the banking collapses in Japan and Scandinavia during the 1990s, and similar banking system debacles occurring in 140 developing countries in the last two decades of the twentieth century, all of which experienced

banking system losses in excess of 1 percent of GDP, and more than twenty of which experienced losses in excess of 10 percent of GDP (data are from Caprio and Klingebiel, 1996, updated in private correspondence with these authors). Empirical studies of these unprecedented losses concluded that deposit insurance and other policies that protect banks from market discipline, intended as a cure for instability, have instead become the single greatest source of banking instability.

The theory behind the problem of destabilizing protection has been well known for over a century, and was the basis for Franklin Roosevelt's opposition to deposit insurance in 1933 (an opposition shared by many). Ironically, federal deposit insurance is one of the major legacies of the Roosevelt presidency, despite the fact that President Roosevelt, the Federal Reserve, the Treasury, and Senator Carter Glass—the primary authorities on banking policy of the time—all were opposed to it on principle. Deposit insurance was seen by them and others as undesirable special-interest legislation designed to benefit small banks. They acquiesced in its passage for practical reasons—to get other legislation passed—not because they wanted deposit insurance to pass per se. Numerous attempts, dating from the 1880s, to introduce federal deposit insurance legislation failed to attract support in the Congress (Calomiris and White, 1994). Opponents understood the theoretical arguments against deposit insurance espoused today—that deposit insurance removes depositors' incentives to monitor and discipline banks, and frees bankers to take imprudent risks (especially when they have little or no remaining equity at stake, and see an advantage in 'resurrection risk taking'); and that the absence of discipline also promotes banker incompetence, which leads to unwitting risk taking.

Research on the banking collapses of the last two decades of the twentieth century have produced new empirical findings indicating that the greater the protection offered by a country's bank safety net, the greater the risk of a banking collapse (see, e.g., Caprio and Klingebiel, 1996; Boyd, et al., 2000; Demirgüç-Kunt and Detragiache, 2000; and Barth, Caprio, and Levine, 2006). Empirical research on prudential bank regulation similarly emphasizes the importance of subjecting some bank liabilities to the risk of loss to promote discipline and limit risk taking (Shadow Financial Regulatory Committee 2000; Mishkin 2001; Barth, et al., 2006).

Studies of historical deposit insurance reinforce these conclusions (Calomiris, 1990). Opposition to deposit insurance in the 1930s reflected the disastrous experience with insurance in several US states in the early twentieth century, which resulted in banking collapses in all the states that adopted insurance. Government protection of banks played a similarly destabilizing role in Argentina in the 1880s (leading to the 1890 collapse) and in Italy (leading to its 1893 crisis). In retrospect, the successful period of US deposit insurance, from 1933 through the 1960s, was an aberration, reflecting limited insurance during those years (insurance limits were subsequently increased), and the unusual macroeconomic stability of the era.

Conclusion

Banking failures, in theory, can be a consequence either of fundamental, exogenous shocks to banks, or, alternatively, unwarranted withdrawals by depositors associated with contagions of fear, or panics. Interestingly, although many economists associate contagions of fear with the banking distress of the Great Depression, empirical research indicates that panics played a small role in Depression-era distress, which was mainly confined to regional episodes (e.g., June 1932 in Chicago) or to the banking collapse of 1933.

More importantly, empirical research on banking distress clearly shows that panics are neither random events nor inherent to the function of banks or the structure of bank balance sheets. Panics in the US were generally not associated with massive bank failures, but rather were times of temporary confusion about the incidence of shocks within the banking system. This asymmetric-information problem was particularly severe in the US. For the late nineteenth and early twentieth centuries, system-wide banking panics like those that the US experienced in that period did not occur elsewhere. The uniquely panic-ridden experience of the US, particularly during the pre-World War I era, reflected the unit banking structure of the US system. Panics were generally avoided by other countries in the pre-World War I era because their banking systems were composed of a much smaller number of banks operated on a national basis, who consequently enjoyed greater portfolio diversification *ex ante*, and a greater ability to coordinate their actions to stem panics *ex post*. The US also experienced waves of bank failures unrelated to panics (most notably in the 1920s), which reflected the vulnerability to sector-specific-shocks (e.g., agricultural price declines) in an undiversified banking system.

More recent banking system experience worldwide indicates unprecedented costs of banking system distress—an unprecedented high frequency of banking crises, many bank failures, and large losses by failing banks, sometimes with disastrous costs to taxpayers who end up footing the bill of bank loss. This new phenomenon has been traced empirically to the expanded role of the government safety net. Government protection removes the effect of market discipline. It thereby encourages excessive risk taking by banks, and also creates greater tolerance for incompetent risk management (as distinct from purposeful increases in risk). Ironically, the government safety net, which was designed to forestall the (overestimated) risks of contagion, seems to have become the primary source of systemic instability in banking.

References

Allen, F. and D. Gale (2004). 'Financial Fragility, Liquidity, and Asset Prices', *Journal of the European Economic Association*, 2: 1015–48.

Alston, L. J., W. A. Grove, and D. C. Wheelock (1994). 'Why Do Banks Fail? Evidence from the 1920s', *Explorations in Economic History*, 30: 409–31.

Barth, J. R., G. Caprio, Jr., and R. Levine (2006). *Rethinking Bank Regulation: Till Angels Govern*, Cambridge: Cambridge University Press.

Bernanke, B. S. (1983). 'Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression', *American Economic Review*, 73: 257–76.

—— and H. James (1991). 'The Gold Standard, Deflation, and Financial Crisis in the Great Depression: An International Comparison', in R. Glenn Hubbard (ed.), *Financial Markets and Financial Crises*, Chicago: University of Chicago Press, 33–68.

Bordo, M. (1985). 'The Impact and International Transmission of Financial Crises: Some Historical Evidence, 1870–1933', *Revista di Storia Economica*, second ser., 2: 41–78.

Boyd, J., P. Gomis, S. Kwak, and B. Smith (2000). 'A User's Guide to Banking Crises'. Conference Paper, The World Bank.

Bruner, R. F. and S. D. Carr (2007). *The Panic of 1907: Lessons Learned from the Market's Perfect Storm*, Chichester: Wiley.

Calomiris, C., W. (1990). "Is Deposit Insurance Necessary? A Historical Perspective," *Journal of Economic History*, 50, 283–95.

—— (1992). 'Do Vulnerable Economies Need Deposit Insurance? Lessons from U.S. Agriculture in the 1920s', in P. L. Brock (ed.), *If Texas Were Chile: A Primer on Bank Regulation*, San Francisco: The Sequoia Institute, 237–349, 450–58.

—— (2000). *U.S. Bank Deregulation in Historical Perspective*, Cambridge: Cambridge University Press.

—— (2007). 'Victorian Perspectives on the Banking Distress of the Late 20th Century', Working Paper.

—— (2008). 'The Subprime Turmoil: What's Old, What's New, and What's Next', in *Maintaining Stability in a Changing Financial System*, Federal Reserve Bank of Kansas City's Jackson Hole Symposium, August 21–22, 2008.

—— (2009). 'Helping Wall Street—And Main Street', Forbes.com, January 21, 2009.

—— and G. Gorton (1991). 'The Origins of Banking Panics: Models, Facts, and Bank Regulation', in R. Glenn Hubbard (ed.), *Financial Markets and Financial Crises*, University of Chicago, 107–73.

—— and C. M. Kahn (1991). 'The Role of Demandable Debt in Structuring Optimal Banking Arrangements', *American Economic Review*, 81: 497–513.

—— and J. R. Mason (1997). 'Contagion and Bank Failures During the Great Depression: The June 1932 Chicago Banking Panic', *American Economic Review*, 87: 863–83.

—— —— (2003a). 'Fundamentals, Panics and Bank Distress During the Depression', *American Economic Review*, 93: 1615–47.

—— —— (2003b). 'Consequences of Bank Distress During the Great Depression', *American Economic Review*, 93: 937–47.

—— —— (2004). 'How to Restructure Failed Banking Systems: Lessons from the U.S. in the 1930s and Japan in the 1990s', in T. Ito and A. Krueger (eds.), *Governance, Regulation, and Privatization in the Asia-Pacific Region*, Chicago: University of Chicago Press, 375–420.

—— and L. Schweikart (1991). 'The Panic of 1857: Origins, Transmission, and Containment', *Journal of Economic History*, 51: 807–34.

—— and E. N. White (1994). 'The Origins of Federal Deposit Insurance', in C. Goldin and G. Libecap (eds.), *The Regulated Economy: A Historical Approach to Political Economy*, Chicago: University of Chicago, 145–88.
—— and B. Wilson (2004). 'Bank Capital and Portfolio Management: The 1930s "Capital Crunch" and Scramble to Shed Risk', *Journal of Business*, 77: 421–55.
Caprio, G. and P. Honohan (2007). 'Banking Crises', this volume, chapter 26.
—— and D. Klingebiel (1996). 'Bank Insolvencies: Cross Country Experience'. Working Paper 1620, The World Bank.
Demirgüç-Kunt, A. and E. Detragiache (2000). 'Does Deposit Insurance Increase Banking System Stability?', Conference Paper, The World Bank.
Diamond, D. and P. Dybvig (1983). 'Bank Runs, Deposit Insurance, and Liquidity', *Journal of Political Economy*, 91: 401–19.
—— and R. Rajan (2002). 'Liquidity Shortage and Banking Crises', National Bureau of Economic Research Working Paper 8937, May.
Friedman, M. and A. J. Schwartz (1963). *A Monetary History of the United States, 1867–1960*, Princeton, NJ: Princeton University Press.
Gorton, G. (1985). 'Clearing Houses and the Origin of Central Banking in the United States', *Journal of Economic History*, 45: 277–83.
Grossman, R. S. (1994). 'The Shoe That Didn't Drop: Explaining Banking Stability During the Great Depression'. *Journal of Economic History*, 54: 654–82.
Halac, M. and S. Schmukler (2004). 'Distributional Effects of Crises: The Financial Channel', *Economia*, 5: 1–67.
Lucia, J. L. (1985). 'The Failure of the Bank of United States: A Reappraisal', *Explorations in Economic History*, 22: 402–16.
Mason, J. R. (2003). 'Do Lender of Last Resort Policies Matter? The Effects of Reconstruction Finance Corporation Assistance to Banks During the Great Depression', *Journal of Financial Services Research*, 20: 77–95.
Mishkin, F. S. (ed.) (2001). *Prudential Supervision: What Works and What Doesn't*, Chicago: University of Chicago Press.
Nicholas, H. C. (1907). 'Runs on Banks', *Moody's Magazine*, December.
O'Grada, C. and E. N. White (2003). 'The Panics of 1854 and 1857: A View from the Emigrant Industrial Savings Bank', *Journal of Economic History*, 63: 213–40.
Rousseau, P. (2002). 'Jacksonian Monetary Policy, Specie Flows, and the Panic of 1837', *Journal of Economic History*, 62: 457–88.
Shadow Financial Regulatory Committee (2000). *Reforming Bank Capital Regulation*, American Enterprise Institute, 2000.
Sprague, O. M. W. (1910). *History of Crises under the National Banking System*, National Monetary Commission.
Stern, G. H. and R. J. Feldman (2003). *Too Big to Fail: The Hazards of Bank Bailouts*, Washington, DC: Brookings Institution Press.
Temin, P. (1969). *The Jacksonian Economy*, New York: W. W. Norton.
—— (1976). *Did Monetary Forces Cause the Great Depression?*, New York: W. W. Norton.
—— (1989). *Lessons from the Great Depression*, Cambridge, MA: MIT Press.
White, E. N. (1984). 'A Reinterpretation of the Banking Crisis of 1930', *Journal of Economic History*, 44: 119–38.

WICKER, E. (1980). 'A Reconsideration of the Causes of the Banking Panic of 1930'. *Journal of Economic History,* 40: 571–83.

—— (1996). *The Banking Panics of the Great Depression*, Cambridge: Cambridge University Press.

WIGMORE, B. A. (1987). 'Was the Bank Holiday of 1933 a Run on the Dollar Rather than the Banks?', *Journal of Economic History,* 47: 739–56.

CHAPTER 28

PAYMENTS SYSTEMS

DAVID HUMPHREY
JAMES McANDREWS

Introduction

Many different payment instruments are used for transactions. Cash, checks, and debit and credit cards are used for small and medium-value transactions for 'retail' payments at the point of sale. Checks, automated clearing house (ACH in the US), or giro (in Europe) credit transfers and direct debits are used, with cards at times used over the Internet or phone, for medium-value bill payments which are also considered 'retail' transactions. For large-value or 'wholesale' payments, wire transfers (in the US and Europe) and giro payments (in Europe) are used. These represent large-value transactions among businesses, between business and government, and for almost all-large value financial transactions in the foreign exchange, bond, equity, derivative, and other financial markets. As retail and wholesale transactions and the policy issues they present are so different, they are treated separately in most of the sections below with retail payments covered first followed by wholesale transactions.

Production structure

A country's payment system is comprised of the payment instruments listed above, the banking institutions directly involved in offering transaction services, the bank

and non-bank firms processing the payments, the transportation firms (for cash and checks) and telecommunications facilities (for electronic payments) needed to move payment information between bank accounts, and the central bank for final settlement of transactions. Cash does not require final settlement—the transfer of good and final funds between accounts—since coin and currency already represents final payment. Central banks are used for final settlement, rather than private banks, owing to their low cost and the fact that they cannot fail (as the government could print money or tax to support them, if needed).

In the US and Europe, banks have an effective monopoly in offering retail payment services. This is enforced by the legal definition of what a bank is and does since typically only banks are allowed access to central bank settlement accounts. In some countries (e.g., Canada, Australia) non-bank financial firms can have limited access to central bank settlement services and in many countries non-bank institutions may offer limited payment services (e.g., money order firms, money transmitters of inter-country remittances) outside of usual banking channels.

A different arrangement exists for large-value payments. Here, volumes are less than 0.5 percent of all non-cash retail transactions but over twelve times the value of retail payments. Due to the importance of these transactions for financial markets, central banks are the primary suppliers (or guarantors) of these payments. However, groups of private banks also process large-value transactions (e.g., CHIPS in the US, Euro1 in Europe) and business and other users purchase them through banks with settlement at the central bank. International payments are made via SWIFT, a message-transfer network using correspondent accounts at banks in sending and receiving countries. As there is no world central bank, funds are moved and settlement is made when banks in two countries transfer funds from one internal account to another as directed by the SWIFT message.

Retail payments

Payment theory

Using simplified models, payment theory focuses on how different payment arrangements evolved among transactors as a substitute for barter. These analyses offer insights into the reasons for different types of payment instruments and on their efficiency in facilitating transactions (Kahn and Roberds, 2007). Recent theoretical work has explored the risk, liquidity benefits, and costs of different settlement arrangements on large-value transfer networks (net settlement versus real-time gross settlement, an issue covered below). In contrast, most empirical analysis of payments relies on well-accepted microeconomic theory dealing with

payment pricing principles, demand estimation, analysis of cost and scale economies, and measurement of competition. The most novel development in payment theory concerns recasting traditional demand theory into a two-sided market framework (Rochet and Tirole, 2003). It is applied to credit and debit card pricing arrangements (the interchange fee) as well as to other markets and forms the basis for recent theoretical work in the retail payments area (a survey is provided by Chakravorti, 2003). Many developments in this field are related to antitrust and competition issues that have dominated policy debates, regulatory actions, and legal cases. Papers on these topics can be found in the *Review of Network Economics* (2005 and 2006).

Differences in payment structure

Current payment instruments evolved from earlier forms (pepper corns, sea shells, precious metals) because they are easier to transfer and store, have an agreed upon stable nominal value, are more easily divisible, and safer to use. The current transition from paper (checks, paper giros) to electronic non-cash payments (cards, electronic giros) is mostly due to expanded convenience and lower costs. Generally, an electronic transaction costs only one-third to one-half as much as the all-in social or private cost of a paper-based one for the same purpose. The substitution of electronic- for paper-based payments has expanded considerably over the last fifteen years and, when finished, may save 0.5 percent to 0.7 percent of a nation's GDP annually. Some European countries have essentially eliminated checks and reduced cash in favor of cards while the US, which in the past used checks markedly to reduce cash, is now using cards to replace checks.

Use of cash is approximated by the ratio of cash in circulation to GDP (other methods are more involved). The cost of delivering cash acquisition services to depositors has fallen significantly as banks have progressively substituted ATM networks for networks of branch offices. Indeed, in many countries the number of expensive stand-alone offices has fallen absolutely. In terms of cash use, Table 28.1

Table 28.1. Payment instrument use, 2005

	Cash/GDP Ratio	Annual use per person			
		Check	Card	Giro/ACH	Total non-cash
US	2.6	112	145	43	300
UK	3.5	32	104	95	231
Canada	3.7	42	156	46	244
Euro area	7.3	16	46	92	154
Japan	16.7	1	39	11	51

indicates that European countries in the euro area use more than twice the value of cash than does the US, while Japan uses more than twice that of Europe (US data have been corrected for the approximately 60 percent of cash held overseas rather than used domestically). The six countries/areas shown in the table have been ranked according to their approximate cash use which, in general, is the reverse of the ranking of their total use of non-cash instruments at the point of sale and for bill payments and other disbursements (in the last column).

Estimates of the share of cash in point of sale payments are 20 percent for the US but can be more than three times higher in Europe (Humphrey, Snellman, and Vesala, 2001). This is consistent with a general aversion to consumer debt in Europe (hence their relatively low use of credit cards) and a history of relying on giro credit transfers for bill payments (where funds have to be in an account in order to make a transfer).

One area where cash is the clear instrument of choice is in tax evasion and other illegal transactions. Here large denomination notes are heavily used and in some countries the share of these notes in the value of cash in circulation is in rough agreement with statistical and survey-based estimates of the value of illegal activity. Estimates of such activities differ greatly but averages 19 percent of GDP for twelve European countries and 9 percent for the US (Schneider and Enste, 2000: Table 7).

Differences in payment instrument use across countries is perhaps best seen when non-cash payments are in terms of annual use per person. As illustrated in Table 28.1, 112 checks are written per person in the US (about evenly divided between consumers and businesses), 16 per person in the fifteen-country euro area, but only 1 per person in Japan. And, while there are 39 card transactions per person in Japan and 46 in the euro area, card transactions are three times higher (at 145 and 156) in the US and Canada. In contrast, the UK and countries in the euro area make twice as many giro/ACH payments per person than does the US or Canada and over eight times that for Japan. These comparisons illustrate the heavy reliance in Japan on cash for retail transactions, the focus of the US on checks and cards, and the emphasis of the UK and the fifteen countries in the euro area on giro/ACH payments.

These differences in payment instrument use across countries are associated with historical and institutional 'accidents' rather than any clear plan to shape payment use. Such accidents include the development of a postal giro system in Europe before 1900 but never in the US; restrictions on intrastate and interstate banking in a geographically large US fearful of the concentration of economic power, but far fewer branching or concentration restrictions in the geographically smaller countries of Europe; the cost of obtaining a banking charter over 100 years ago—low in the US, so have many small banks—but very high in Canada, which led to a nationwide and concentrated banking structure (similar to that of most European countries with a history of royal or state monopolies); and the fact that Japan is

considered a very safe country so the need to find a substitute for cash was and remains weak. Given the different institutional history, Europe was able to offer nationwide paper-based giro payments using postal banks (and now also commercial banks) while the US had no alternative but to rely on checks for transactions among thousands of small banks since no national payment supplier existed. Consequently, the ongoing shift to electronics for bill payments and disbursements has been more rapid in Europe than in the US.

Another 'accident' concerned the early US practice of discounting the face value of a check to cover the cost of collection (non-par checking). This resulted in payment delays that were at times inordinately long and disrupted commerce. One reason the Federal Reserve was established was to transport and process checks at face (par) value and at no charge to eliminate circuitous routing of checks which hindered commerce. Non-par checking was eliminated, but today, alone among central banks, the Federal Reserve processes about one-third of all checks and the majority of ACH payments (now for a fee). Europe, not having many (debit transfer) checks to begin with due to a reliance on cash and giro (credit transfer) payments, instead covered payment costs by debiting accounts prior to the value date to earn float revenues, and central banks are not involved in processing retail transactions in competition with banks.

A newer way to make a payment at the point of sale is with e-money, such as a card with a chip. This currently accounts for less than 1 percent of transactions but is viewed as a possible replacement for small-value cash transactions in the future (although adoption has been slow). Another recent development involves the use of the Internet in Europe to initiate giro bill payments (credit transfers). Online banking has lagged in the US since, unlike Europe, many payees are not set up to accept ACH transactions from individual consumers.

Payment costs

The full cost of a nation's payment system has been estimated to be 1 percent to 2 percent of GDP annually. The social cost of making a $50 or €50 payment is around 1 percent to 5 percent of the transaction value, depending on the instrument used. Banks know their payment production costs but per transaction fees are rarely assessed. Consumers choose different payment instruments for different applications (e.g., local point of sale transactions versus more distant bill payments) and these choices are largely driven by non-price influences such as convenience and perceived safety. The availability of terminals is a clear precondition for use of cards and cash flow considerations and reward programs are additional factors for credit cards. On the merchant side, some limit cash use due to counterfeiting concerns or limit card use to transaction values greater than a cut-off amount to cover better bank card fees.

On the supply side, cost considerations have induced banks to shift cash acquisition by consumers away from branch offices to cheaper ATMs and to cards away from checks and cash for the same reason. Averaging two European cost accounting studies (for the Netherlands and Belgium) suggests that the bank plus merchant average cost of a cash transaction at the point of sale is around €0.42 while a debit card is €0.52, with incremental costs of €0.12 and €0.20, respectively. Average credit card costs are €3.11, with incremental cost of €0.68. As the ratio of marginal to average cost reflects scale economies, these figures also suggest that cash and cards realize strong cost economies (from 0.29 to 0.38, respectively, for cash and debit cards, and lower still at 0.22 for credit cards). Similar payment-scale economy estimates for Europe have been obtained in statistical estimation using bank-based as well as payment-processor-based data (Bolt and Humphrey, 2007). Thus, consolidation of payment processors offers the opportunity for substantial reductions in payment costs in Europe, a result that will facilitate the emergence of a more competitive cross-country product market within Europe as envisioned by the Single Euro Payments Area policy promoted by the EC and the European Central Bank. Payment-scale economies have long been identified (and realized) for US checks, ACH, and wire transfers (using Federal Reserve data) and debit cards (private source). With scale economies, card and giro/ACH unit costs should continue to fall as volume expands, while for cash, and especially checks, unit costs should rise with falling volumes.

Direct vs. indirect pricing of payment services

Most pricing of consumer payments in the US and Europe is indirect and relies on fixed monthly account fees, minimum balance requirements, no or low interest paid on deposits, and/or payment float. Payment services, like deposits, are not viewed as a profit center but as a means to provide lower cost funding for bank loans compared to purchased funds or debt. The lack of price data or detailed information on deposit account non-price characteristics has limited the application of microeconomic theory to assess payment or other banking service demand and substitution relationships. However, some of these relationships have been inferred from US consumer payment survey data (Borzekowski, Kiser, and Ahmed, 2008) while in Europe the focus has been on Norway which has directly priced their payment services for over a decade (Norges Bank, various years). Per transaction pricing in Norway apparently speeded up the shift to electronic card payments from cash and checks by about 20 percent compared to the Netherlands which did not price (Bolt and Humphrey, 2008). Payment demand elasticities here are typically inelastic and non-price characteristics of electronic payments seem to be more important than pricing in explaining volume growth. In this regard, terminal availability is a good summary indicator for the non-price attributes of

card use since mere availability is quickly followed by use. For ATMs, convenience trumps pricing which is why non-banks offer ATM services to any bank depositor for a fee and customers of one bank use another bank's ATMs even though a surcharge applies (Hannan, Prager, and McAndrews, 2003).

Effect on monetary policy

The substitution of electronic payments for cash is slow enough not to have an important effect on monetary policy. The oft-predicted 'cashless society' is still a long way off, especially in Europe, which uses more than twice as much cash as the US. So far, population growth and inflation have typically offset the reduction in the share of cash in total payments so that its absolute value is still expanding or stable. Once the value of cash falls absolutely, however, tax revenues will have to be used for its redemption, and seigniorage revenues will be correspondingly reduced. Singapore has raised the possibility that a government could redeem currency and replace it with a government-issued card and account of equal value, effectively retaining seigniorage revenues as well saving tax revenues otherwise needed for currency redemption.

Fraud

US losses from payment fraud are $0.04 billion annually from counterfeit cash, $0.7 billion from check fraud, and $1.8 billion (and rising) from unauthorized use of debit and credit cards. These losses trade off with user convenience and illustrate the difficulty in restricting improper use, especially when card information is compromised through data breaches at merchants and payment processors. Solutions exist but are expensive: PIN numbers that are not observed or saved by merchants, chips that encrypt card information, and randomly varied transaction verification numbers after each payment.

Data availability

There is little more than fifteen years of annual data on the volume and value of non-cash payment instrument use by country (Bank for International Settlements and European Central Bank, various years). This is all at the national level. Only Norway has collected payment use, cost, and pricing data at the individual bank level. No country has time series information on the number and value of cash transactions and the value of cash outstanding is only an approximate indicator of its domestic use (especially for the US). European cash data are better in this regard

as are their check data. Card information is good but the split between debit and credit is often missing for Europe. The US does not have accurate time series check volume or value data, except perhaps after 2000 and two or three earlier time periods. Instead, there is considerable survey information on the demographic characteristics of people who use different payment instruments (cf., recurring University of Michigan surveys). This has been used to infer payment preferences (Hayashi and Klee, 2003) and sometimes price responsiveness. More recently, cost-accounting analysis has been applied to determine the unit cost of payment instruments at the bank and merchant level (costs only: Brits and Winder, 2005; costs and benefits: Garcia-Swartz, Hahn, and Layne-Farrar, 2004; and Shampine, 2007; literature survey: Koivuniemi and Kemppainen, 2007). Consumer benefits from using the various payment instruments, however, remain difficult to estimate accurately. Consequently, comparing calculated net benefits with observed relative usage may be an indicator of how inaccurate these calculations are and/or reflect the fact that most payment instruments are only indirectly priced (lost float, fixed monthly bank account fees, mostly free ATM cash withdrawal, hidden interchange fees, etc.) so relative use is not strongly based on actual relative resource cost.

Policy issues

Central banks are responsible for the safety and soundness of payment systems and profess a concern for its cost efficiency as well. Australia has a specific legislative mandate regarding payment efficiency, Europe is getting involved, but the US is much less active. All countries, however, have regulations, legislation, and case law that spell out the various rights and liabilities of consumer, merchant, and bank participants to a payment transaction.

One current policy example is the Single Euro Payments Area initiative noted above which seeks to lower the cost of within and cross-border payments in Europe. Another cost-related issue concerns debit and credit card interchange fees. The theory of two-sided markets outlines the logic behind interchange fees between a merchant acquiring bank and a card-issuing bank where the interchange fee compensates the issuing bank when the issuing bank's customer uses its card, since expenses are incurred by both banks. If the issuing bank is also the acquiring bank, the 'interchange fee' becomes an internal (and less obvious) transfer rather than an external one.

The interchange controversy exists because (a) merchant costs of accepting credit cards tend to be much higher than for other payment instruments and (b) interchange fees have often not fallen as volume has expanded even though scale economies are being realized. Merchant costs are higher still for cards that offer a reward when consumers use them and, since these costs are factored into what the merchant sells, users of cash debit cards, and checks effectively cross-subsidize credit card users'

reward programs. The Reserve Bank of Australia mandated a reduction in interchange fees (Reserve Bank of Australia, 2005) while competition authorities in Europe are putting pressure on banks to reduce and/or justify better the interchange fees they charge. As well, legal action by merchants is under way in the US to achieve the same end—with some success, since acceptance of a credit card no longer requires that the issuer's debit card—and interchange fee—also be accepted.

Other policy issues deal with money laundering using pre-paid payment cards (although industry initiatives have largely contained this problem) and use of government-issued cards to replace cash and checks for welfare and food stamp programs (Electronic Benefit Transfer in the US). Overall, the use of stored-value cards for general purposes (rather than narrow applications) has been unsuccessful in the US and is currently stalled at a low level of acceptance in Europe (Van Hove, 2006).

Wholesale payments

Interbank payments

In modern monetary economies, the value of non-cash payments dominates cash transactions and almost all retail and wholesale (large-value) non-cash transactions are made by the transfer of bank deposit liabilities. Banks therefore play a central role in the system of payments, both as providers and users of payments. Central banks in most countries provide a system of accounts to commercial banks, and an associated large-value payment system (LVPS), through which commercial, savings, or co-operative banks transfer account balances among themselves. Private sector LVPSs also exist in some countries. Banks typically use these LVPSs to make payments on behalf of bank customers, so transactions on LVPSs are a mixture of bank-to-bank payments and bank-customer-to-bank-customer payments.

The LVPSs transfer large amounts of value relative to GDP each year, averaging approximately seventy-five times the value of GDP in Europe and the US in 2005 (Bank for International Settlements, various years). These values are over twelve times larger than values transferred by retail payment networks and reflect the financial nature of most of the underlying transactions. These include the interbank (often overnight) market for funds or bank account balances, other interbank transactions, including the sales or syndications of loans, proceeds of sales of securities by large commercial and financial firms, settlement of foreign exchange trades, and other financial asset trades. The large values that flow through LVPS systems are much better protected from fraud and operational failure than are retail payment networks.

Wire transfers

Most LVPSs use wire transfers as the payment instrument. A wire transfer is a credit transfer in which the originating bank directs funds to be sent to a recipient bank. Wholesale, or interbank, wire transfer systems (to be distinguished from retail systems that are often used for cross-country remittances) limit access by requiring participant banks to use dedicated hardware to initiate payments which are transferred over dedicated communication links using encryption. Fedwire™ is the wire transfer system in the US, owned and managed by the Federal Reserve System, and TARGET is the wire transfer system owned and managed by the European Central Bank used in many countries in Europe.

The growth in wholesale wire transfers has been rapid over 1985 to 2005. In the US, payment values have expanded from thirty-five times annual GDP to seventy-five times GDP, while transaction volume growth has been more moderate, averaging 6 percent annually. Similar growth rates have been observed in other industrialized countries, with a general trend of more-moderate growth in both values and volumes after 2000.

Real-time gross settlement systems

The design of wholesale wire transfer systems has changed in the period 1980 to date from systems that net payments during the day, to systems that instead transfer funds in the full amount immediately during the day, so-called real-time gross settlement (RTGS) systems. Real-time gross settlement systems offer a number of advantages, including immediate final settlement, which means that even though the sending bank were to fail following the completion of the payment, the payment would not be reversed. In a net settlement system, by contrast, final settlement is pending during the operation of the system until the end of the period (typically a day) when participants fund their net obligations to other participants. The failure of a net debtor bank prior to settlement could, if not covered by posted collateral on a network, lead to an unwinding of that party's payment. This can negatively affect other participants in the system and lead to a failure to settle their payments as well (termed 'systemic risk'). Bankruptcy law has been changed to not require that funds sent by a failed bank be reversed on the day of failure, but contractual agreements banks have with a net settlement network can require a settlement unwind anyway, if posted collateral is insufficient. A comparative analysis of the settlement risk on gross and net systems is contained in Kahn, McAndrews, and Roberds (2003), while a comprehensive analysis of the risk, liquidity, and public policy issues associated with large-value payment networks is covered in Manning, Erlend, and Schanz (2008).

In 1985, only three central banks had adopted RTGS systems, while by 2006, ninety-three of the world's 174 central banks had adopted them. Real-time gross settlement systems require real-time monitoring of account balances and efficient and timely communication infrastructures to support their operations relative to the netting systems they displaced. Subsequent advances in computing power and communication links as well as consolidation of operating centers over time have allowed central banks to lower their costs and the fees they charge (Bech, Preisig, and Soramaki, 2008).

Liquidity and systemic risk

Because funds are transferred immediately in RTGS systems, and because the values transferred in the systems are so large, banks must make provisions to have funds available in their accounts for the payments to occur. To accommodate the demand for funds for payments, most central banks provide credit to banks during system-operating hours. This credit, often called 'daylight credit', is different from an overnight loan by the central bank through a discount window facility. Most often, central banks provide daylight credit at a zero interest rate against collateral provided by commercial banks in order to assure the funding of payments during the day. The US is an anomaly as it provides limited amounts of daylight credit at a small fee, but (generally) without the requirement that the credit be collateralized.

The fact that many out-payments are fully or partially offset later in the day with in-payments, netting these asynchronous payments is one way to economize on the need for holding large idle money balances or posted collateral required in RTGS systems. However, networks that net payments during or at the end of the day face the possibility of default by a participant on its obligation when it settles. The default by one participant can lead to the default of others, as they are deprived of the expected inflow of funds by the defaulting party. A simulation in the 1980s using actual payment data suggested that the unexpected failure to settle on a US net settlement network could lead to the unwinding of as much as one-third of that day's payments (Humphrey, 1986). However, simulations using data from other countries were far less extreme because their banking systems were more concentrated (generating more internal transfers among customer accounts within banks) so their bank-to-bank external daylight credit exposures were lower. More recent simulations (in the 2000s) show that a much-lower level of unwinding would occur given today's payments activities—probably because of increased levels of bank capital, more stringent net debit limits, and bank mergers which make previously external payments among two banks into internal account transfers of a merged institution.

Systemic risks exist in financial markets outside of large-value net settlement payment networks. Unfortunately, information on the web of cross-bank and non-bank financial institution exposures is currently either too limited or non-existent

(as in the case of credit default swaps during the credit crisis of 2008) to permit authorities to monitor continuously or simulate in near-real time the systemic effects of financial institution failure on the financial system.

Analyses of risks on LVPS networks have identified several sources of risk, including credit risk, liquidity risk, legal risk, time-gap risk, in addition to systemic risk already discussed. Credit risk is the possibility of loss from a bank not settling its net position on a net settlement network; liquidity risk is the increased cost that may be experienced due to an unexpected delay in receiving a payment; legal risk refers to the uncertain legal status of some payments in situations where banks are in different legal jurisdictions; time-gap risk is the credit risk incurred when there is a delay between when one part or 'leg' of a currency transaction is completed in a foreign exchange trade and when the other leg is completed, since two currencies are involved in the trade. These risks have been addressed and mitigated in a variety of ways.

The adoption of real-time gross settlement systems was influenced by the fact that daylight credit exposures on net settlement networks were large relative to banks' capital positions, were growing much faster than capital levels, and were shown to threaten a substantial percentage of a given day's network payment value if there was an unexpected failure to settle by a bank participant. Real-time gross settlement eliminates this systemic risk since each payment is individually settled with finality when made, rather than waiting until the end of the day to settle a net position. However, RTGS systems require either more idle daylight balances or posted collateral to cover what otherwise would be daylight credit exposures. Although central banks have addressed this issue by offering daylight credit at zero or low marginal costs, sometimes having sufficient collateral to post can be a constraint. Legal risks have been addressed through changes in bankruptcy laws (so payments made earlier in the day by a sending bank that fails are not reversed) and strengthening the legal underpinnings of netting (by novation) contracts through changes in statutory laws. Time-gap risk in the foreign exchange market has been addressed through the development of the CLS (continuous linked settlement) Bank, discussed below. Systemic risk has been addressed by the adoption of real-time gross settlement, real-time monitoring of payment accounts, improved bank supervision, and enhanced regulation of bank capital. Private sector net settlement systems have also adopted real-time monitoring of payment accounts, have placed limits on maximum net debit exposures, and established pools of collateral that can be used in the event a participant fails to settle (usually equal to the largest single net debit that can be incurred on the network).

It is easy to see how daylight overdrafts and systemic risk arose on early net-settlement networks. Indeed, they were purposely structured so that the net position of intraday exposures would be settled only at the end of day since this reduced the cost of holding idle intraday balances or posting collateral to cover and clear each payment as it was made. But, how did the daylight overdraft problem arise on central bank wire transfer networks which offered immediate settlement in good and final funds when payments are sent? First, at least for the US, the shift to

making wire transfers fully electronic during the late 1970s and early 1980s did not initially include real-time monitoring of reserve account positions. This was deemed to be too expensive since, when wire transfers required more manual intervention and balances were effectively monitored both during and at the end of the day, daylight overdrafts were relatively rare. Second, the combination of an unconcentrated banking system, the institutional practice favoring overnight (rather than one-week, or 'good til canceled') interbank borrowing, the practice of purchasing and accumulating securities during the day until the entire order could be delivered and paid for (as partial deliveries were not then allowed), along with a reduction in reserve balances (through lower reserve requirements and, later, sweep accounts) all contributed to the problem by increasing the demand for settlement balances while decreasing the supply. Other countries have more concentrated banking systems where internal transfers among customer accounts replace what would otherwise be external funding/payment transactions between accounts at different banks in a less concentrated environment. Finally, there are simply more large-value payment transactions in countries where money centers are located but balances to fund these payments did not expand in tandem.

Liquidity savings mechanisms

Although central banks offer daylight credit to meet payment needs, most limit the amount of credit provided to the level of acceptable posted collateral. If collateral is insufficient, banks may have to economize on their use of central bank-provided daylight credit. This has led to the development of liquidity savings mechanisms employed in conjunction with real-time gross settlement. One procedure allows banks to submit a payment for either immediate settlement or to place it in a queue of pending payment orders. The payment order remains pending until the bank receives offsetting funds (after which the payment is sent) or until the bank's queued payment order is met by an offsetting payment order in the queue of another bank. In this latter case, both payments settle simultaneously as funds are provided to settle one another's payment. These systems have been implemented in the German Bundesbank's RTGS system, and are planned for the LVPSs in Italy, Japan, and in the European Central Bank's Target 2 system (Bank for International Settlement, 2005). The adoption of these systems has reduced the liquidity/collateral costs associated with RTGS systems.

CLS Bank

The CLS Bank, headquartered in London, began operation in September 2002. It was created to provide a means for banks to settle both legs of a foreign exchange

trade simultaneously. It accomplished this first by establishing links with the LVPSs in different countries and it is open simultaneously in all the countries for approximately five hours during the day. Member banks each have a multicurrency account with the CLS Bank. To settle trades, banks make payments into CLS accounts; a trade is settled by simultaneously debiting the accounts of each bank in the currency being sold and crediting them in the currency being bought. Settlement of a trade takes place if, and only if, both sides of the trade successfully complete all the requirements of the payment, a method known as payment-versus-payment.

CLS Bank has grown rapidly, with daily average settlement values exceeding $3.5 trillion and volume exceeding 350,000 sides of foreign exchange trades. It settles trades in fifteen currencies, has more than fifty direct bank members, and over 700 indirect participants (Bank for International Settlements, 2007b). CLS Bank represents a significant innovation in the practice of large-value payment systems by linking the systems in multiple countries and eliminating the time-gap risk noted above.

Related systems

In addition to the LVPS operated by the central bank in a country or currency area, other wholesale payment or settlement systems usually operate. In particular, securities settlement systems are important complementary institutions to LVPSs. Security settlement systems are designed specifically to settle securities trades. In the US, the Depository Trust and Clearing Corporation provides settlement services for the trades of most stock and bond trades in the US along with commercial paper.

Securities settlement systems typically provide another source of demand for wholesale payment systems, as securities settlement system participants usually make payments to and receive payments from the security settlement system using the LVPS operated by the central bank. Security settlement systems employ a wide variety of designs in their settlement system, often using a technique that assures the delivery of the security if, and only if, the payment for the security occurs, a technique known as delivery-versus-payment (again eliminating a time-gap risk).

In many countries, LVPSs owned and operated by the private sector compete with a central-bank-operated LVPS. It is often the case that the private sector-operated LVPS has a narrower base of client banks and settles net (not gross) positions at the end of the day. Net settlement saves on holding idle balances and liquidity and this explains why it was originally adopted on private sector networks instead of RTGS.

Policy issues

In addition to the risk and liquidity cost issues that have been discussed, a related issue in LVPSs is the efficient use of collateral by systems that have bank participants in common. Developments in the clearing and settlement of securities and derivative securities that trade over the counter are recent important policy issues along this line.

As banks conduct business and participate in LVPSs in different countries, they can find themselves with idle collateral available in one country even though they have a binding collateral constraint in a different country in which they operate. This has led some central banks to expand their acceptance of foreign collateral against which they offer daylight credit. However, many technical and legal difficulties prevent banks from completely overcoming restrictions on use of collateral posted in one country for obtaining daylight credit in another country (Bank for International Settlements, 2006).

Security settlement systems which also process large-value transactions often are affiliated with a particular securities trading system, such as a stock exchange. For securities that are traded over the counter (e.g., through a bilateral or dealer trading system), it is often the case that the settlement of trades is performed on a bilateral basis. This lack of centralization and standardization in settlement can lead to less efficient mechanisms for securities settlement. With the growth in derivatives trading, many parties recognized that settlement of these trades was subject to large back-office backlogs and delays. These backlogs have been reduced, but verifying and reconciling these trades still lags behind the rapid growth in trading volume, creating risks for participants (Bank for International Settlements, 2007a).

Payment activities of non-banks

In many countries firms that are not banks also provide payment services. Non-bank firms provide payments for individual retail customers by providing travelers' checks, telegraph transfers, or other remittances, and by offering charge cards or bill aggregation. In addition, large credit cards firms such as Visa and Mastercard are non-banks. While not offering the full range of banking services, these firms provide valuable services to customers in narrow market segments. Called 'money transmitters' in the US, these firms can accept currency from customers and, using the services of a correspondent bank, allow the customer's designated recipient of a payment to withdraw cash from a bank or branch of the firm in a different city. Money transmitter payment services are priced to be profitable and earn revenue

through transaction fees. This stands in contrast to banks who typically charge low fees per transfer for retail customers, but encourage the customer to maintain large balances on account, on which the bank aims to earn a net interest margin.

Firms that offer charge cards or bill aggregation allow consumers to purchase goods from merchants and to pay for the goods only at the end of a billing cycle, such as a one-month period. The charge card firm makes the payment to the merchant earlier, and earns fees by collecting both from the merchant and the consumer. The consumer is expected to pay the charges in full at the end of the billing cycle; this is in contrast to credit cards, which offer the consumer the opportunity to incur a revolving balance loan. In addition, consumer prices in charge card arrangements are often made up of fixed fees, rather than variable interest charges.

Visa and Mastercard are both for-profit public firms but both began their institutional life as associations of banks. Their roles in the credit and debit card industries are far-reaching. They sponsor the systems of branding and acceptance of cards by banks (and bank customers, the cardholders) and merchants. In addition, they manage, together with many telecommunications and computer processing firms, the systems that route the transaction information from the merchant ultimately to the cardholder's bank. These firms and their activities have been of enormous importance and consequence for the worldwide acceptance of credit and debit cards.

The future

In retail payments, debit and credit cards will continue to replace cash (in Europe) and checks (in the US) and their expanded use will intensify efforts to reduce fraud. As card transactions are not anonymous, the continued replacement of cash will raise privacy issues, especially in countries with high sales taxes, where cash is still heavily used (facilitating tax evasion). Similar rapid growth and privacy concerns will probably be seen in electronic retail payment systems used for transportation services such as toll roads. The future of stored-value cards, a potentially lower cost replacement for small-value cash transactions, is uncertain and has not expanded beyond their currently narrow applications (subway rides, public phone calls, gift cards, and merchants close to college campuses). The focus for wholesale payments will be on ways to use excess collateral or liquidity on one LVPS network and on other networks within and across countries as well as reducing back-office delays in handling derivative trades. In general, achieving greater integration of back-office systems and uses of LVPSs will continue to be a goal of LVPS design. None of these

likely developments will affect monetary policy, but can reduce the bank cost of providing payment services. More controversial will be disputes related to the competitive effects of retail card pricing and efforts to price payment services to consumers on a per transaction basis (already done for higher-volume business users) in order to induce consumers to adopt more rapidly instruments with the lowest bank/merchant costs.

References

Bank for International Settlements (2005). *New Developments in Large-Value Payment Systems*. May, 67, Basel.

—— (2006). *Cross Border Collateral Arrangements*. January, 71, Basel.

—— (2007a). *New Developments in Clearing and Settlement of OTC Derivatives*. March, 77, Basel.

—— (2007b). *Progress in Reducing Foreign Exchange Settlement Risk: A Consultative Report*. July, 81, Basel.

—— (various years). *Statistics on Payment Systems in the Group of ten Countries*. Basel.

Bech, M., Preisig, C., and Soramaki, K. (2008). 'Global Trends in Large-Value Payment Systems', *Economic Policy Review*, 14: 59–81.

Bolt, W. and Humphrey, D. (2007). 'Payment Network Scale Economies, SEPA, and Cash Replacement', *Review of Network Economics*, 6: 453–73.

Bolt, W., and Humphrey, D. (2008). 'Transaction Pricing and the Adoption of Electronic Payments: A Cross-Country Analysis', *International Journal of Central Banking*, 4: 89–123.

Borzekowski, R., Kiser, E., and Ahmed, S. (2008). 'Consumers' Use of Debit Cards: Patterns, Preferences, and Price Response', *Journal of Money, Credit and Banking*, 40: 149–72.

Brits, H. and Winder, C. (2005). *Payments Are No Free Lunch*. Occasional Studies 3/2, De Nederlandsche Bank.

Chakravorti, S. (2003). 'Theory of Credit Card Networks: A Survey of the Literature', *Review of Network Economics*, 2: 50–68.

European Central Bank (various years). *Payment and Securities Settlement Systems in the European Union*. Frankfurt.

Garcia-Swartz, D., Hahn, R., and Layne-Farrar, A. (2004). 'The Move Toward a Cashless Society: A Closer Look at Payment Instrument Economics', *Review of Network Economics*, 5: 175–98.

Hannan, T., Prager, R., and McAndrews, J. (2003). 'To Surcharge or Not to Surcharge: An Empirical Investigation of ATM Pricing', *Review of Economics and Statistics*, 85: 990–1002.

Hayashi, F. and Klee, B. (2003). 'Technology Adoption and Consumer Payments: Evidence from Survey Data', *Review of Network Economics*, 2: 175–90.

Humphrey, D. (1986). 'Payments Finality and Risk of Settlement Failure', in *Technology and the Regulation of Financial Markets*, A. Saunders and L. White (eds.), New York: Lexington Books/Solomon Brothers Center Series on Financial Institutions and Markets, 97–120.

—— Snellman, J., and Vesala, J. (2001). 'Substitution of Noncash Payment Instruments for Cash in Europe', *Journal of Financial Services Research*, 19: 131–45.

KAHN, C., MCANDREWS, J., and ROBERDS, W. (2003). 'Settlement Risk under Gross and Net settlement', *Journal of Money, Credit, and Banking*, 35: 591–608.

—— and ROBERDS, W. (2007). 'Transferability, Finality, and Debt Settlement', *Journal of Monetary Economics*, 54: 955–78.

KOIVUNIEMI, E. and KEMPPAINEN, K. (2007). 'On Costs of Payment Methods: A Survey of Recent Studies', Working Paper, Bank of Finland.

MANNING, M., ERLEND, N., and SCHANZ, J. (eds.) (2008). *A Central Bank's Perspective on Large Value Payments and Settlement: Economic Theory and Policy Issues*, Oxford: Oxford University Press.

Norges Bank (various years). *Annual Report on Payment Systems*, Norway.

Reserve Bank of Australia (2005). 'Common Benchmark for the Setting of Credit Card Interchange Fees', November, http://www.rba.gov.au/.

ROCHET, J.-C. and TIROLE, J. (2003). 'Platform Competition in Two-Sided Markets', *Journal of the European Economic Association*, 1: 990–1029.

SCHNEIDER, F. and ENSTE, D. (2000). 'Shadow Economies: Size, Causes, and Consequences', *Journal of Economic Literature*, 38: 77–114.

SHAMPINE, A. (2007). 'Another Look at Payment Instrument Economics', *Review of Network Economics*, 6: 495–508.

VAN HOVE, L. (2006). 'Why Electronic Purses Should Be Promoted', *Banking and Information Technology* (Universität Regensburg, Germany), 2: 20–31.

CHAPTER 29

FINANCE AND ECONOMIC DEVELOPMENT

THE ROLE OF GOVERNMENT

ASLI DEMIRGÜÇ-KUNT[1]

Introduction

What is the role of the financial sector in economic development? Economists hold very different views. On the one hand, prominent researchers believe that the operation of the financial sector merely responds to economic development, adjusting to changing demands from the real sector and is therefore overemphasized (Robinson, 1952; and Lucas, 1988). On the other hand, equally prominent researchers believe that financial systems play a crucial role in alleviating market frictions and hence influencing savings rates, investment decisions, technological

[1] The author is grateful to Meghana Ayyagari, Thorsten Beck, Bob Cull, Patrick Honohan, Vojislav Maksimovic, and Sole Martinez for helpful comments and Edward Al-Hussainy for excellent research assistance. This chapter's findings, interpretations, and conclusions are entirely those of the author and do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

innovation, and therefore long-run growth rates. (Schumpeter, 1912; Gurley and Shaw, 1955; Goldsmith, 1969; McKinnon, 1973; and Miller 1998).[2]

As the financial crisis that started in the summer of 2007 continues to grow and spread all around the world, the potentially disastrous consequences of weak financial sector policies have moved to the forefront of policy debate once again. At its best, finance works quietly in the background, contributing to growth and poverty reduction; but when things go wrong, financial sector failures are painfully visible. Both success and failure have their origins largely in the policy environment; hence, getting the important policy decisions right has always been and continues to be one of the central development challenges.

Despite their inherent fragility, financial institutions underpin economic prosperity. Financial markets and institutions arise to mitigate the effects of information and transaction costs that prevent direct pooling and investment of society's savings. While some theoretical models stress the importance of different institutional forms financial systems can take, more important are the underlying functions that they perform (Levine, 1997; Levine, 2000; and Merton and Bodie, 2004). Financial systems help mobilize and pool savings, provide payments services that facilitate the exchange of goods and services, produce and process information about investors and investment projects to enable efficient allocation of funds, monitor investments and exert corporate governance after these funds are allocated, and help diversify, transform, and manage risk.

While still far from being conclusive, the bulk of the empirical literature on finance and development suggests that well-developed financial systems play an independent and causal role in promoting long-run economic growth. More recent evidence also points to the role of the sector in facilitating disproportionately rapid growth in the incomes of the poor, suggesting that financial development helps the poor catch up with the rest of the economy as it grows. These research findings have been instrumental in persuading developing countries to sharpen their policy focus on the financial sector. If finance is important for development, why do some countries have growth-promoting financial systems while others do not? What can governments do to develop their financial systems?

This chapter addresses these questions. The next section provides a brief review of the extensive empirical literature on finance and economic development and summarizes the main findings. Section three discusses the governments' role in building effective and inclusive financial systems. Finally, the last section concludes with a discussion of the implications of the still-unfolding financial crisis on financial sector policies going forward.

[2] Two famous quotes by Robinson and Schumpeter illustrate these different views. Joan Robinson (1952) argued: 'Where enterprise leads finance follows', whereas Joseph Schumpeter (1950) observed: 'The banker, therefore, is not so much primarily a middleman . . . He authorizes people in the name of society . . . (to innovate).'

Finance and economic development: Evidence

By now there is an ever-expanding body of evidence that suggests countries with better developed financial systems experience faster economic growth (Levine, 1997; and Levine, 2005). More recent evidence also suggests financial development not only promotes growth, but also improves the distribution of income. The following sections provide a brief review of this literature and its findings, also discussing the main criticisms—namely, issues of identification, problems associated with measurement and non-linearities, as well as potential counter-examples and outliers.

Finance and growth

It is by now well established that significant part of the differences in long-run economic growth across countries can be explained by differences in their financial development (King and Levine, 1993; and Levine and Zervos, 1998). The finding that better-developed banks and markets are associated with faster growth is also confirmed by panel and time series estimation techniques (Levine, Loazya, and Beck, 2000; Christopoulos and Tsionas, 2004; and Rousseau and Sylla, 1999). This research also indicates that financial sector development helps economic growth through more-efficient resource allocation and productivity growth rather than through the scale of investment or savings mobilization (Beck, Levine, and Loayza, 2000). Furthermore, cross-country time series studies also show that financial liberalization boosts economic growth by improving allocation of resources and the investment rate (Bekaert, Harvey, and Lundblad, 2005).

However, dealing with *identification issues* is always very difficult with aggregate data. Widespread problems include heterogeneity of effects across countries, measurement errors, omitting relevant explanatory variables, and endogeneity, all of which tend to bias the estimated effect of the included variables. Although the studies cited above have made plausible efforts to deal with these concerns relying on instruments and making use of dynamic panel estimation methodologies, questions still remain. Hence, researchers have used micro data and tried to exploit firm-level and sectoral differences to go beyond aggregates. These studies address causality issues by trying to identify firms or sectors that are more likely to suffer from limited access to finance and see how the growth of these firms and sectors is affected in countries with differing levels of financial development. Demirgüç-Kunt and Maksimovic (1998) and Rajan and Zingales (1998) are two early examples of this approach.

Both studies start by observing that if financial underdevelopment prevents firms (or industries) from investing in profitable growth opportunities, it will

not constrain all firms (or industries) equally. Firms that can finance themselves from retained earnings, or industries that technologically depend less on external finance will be minimally affected, whereas firms or industries whose financing needs exceed their internal resources may be severely constrained. Looking for evidence of a specific mechanism by which finance affects growth—that is, ability to raise external finance—allows both papers to provide a stronger test of causality.

Specifically, Demirgüç-Kunt and Maksimovic (1998) uses firm-level data from 8,500 large firms in thirty countries and a financial planning model to predict how fast those firms would have grown if they had no access to external finance. And they find that in each country the proportion of firms that grew faster than this rate was higher, the higher the country's financial development and quality of legal enforcement.

Rajan and Zingales (1998) instead use industry-level data across thirty-six sectors and forty-one countries and show that industries that are naturally heavy users of external finance benefit disproportionately more from greater financial development compared to other industries. Natural use of external finance is measured by the finance-intensity of US industries since the US financial system is relatively free of frictions, so each industry's use of external finance in the US is assumed to be a good proxy for its demand.

The additional information obtained by working with cross-country firm or industry-level data may not be adequate to satisfy the skeptics, however. For example, although the measure of external financing employed by Demirgüç-Kunt and Maksimovic does not require the assumption that external capital requirements in each industry are the same across countries as that of Rajan and Zingales, it is also more endogenous since it relies on firm characteristics. And although Rajan and Zingales' analysis looks at within-country, between-industry differences and is therefore less subject to criticism due to omitted variables, the main underlying assumption that industry external dependence is determined by technological differences may not be accurate. After all, two firms with the same capital-intensive technology may have very different financing needs, since their ability to generate internal cash flow would depend on the market power they have or the demand they face. Moreover, the level of competition faced by the firm may itself depend on the development of the financial system, introducing more endogeneity.

Beck, et al. (2006) use Rajan and Zingales (1998)'s approach to highlight a distributional effect: They find that industries that are naturally composed of small firms grow faster in financially developed economies, a result that provides additional evidence that financial development disproportionately promotes the growth of smaller firms. Beck, Demirgüç-Kunt, and Maksimovic (2005) also highlight the size effect, but using firm-survey data: they show that financial development eases the obstacles that firms face to growing faster, and that this effect is stronger particularly for smaller firms. More recent survey evidence also

suggests that access to finance is associated with faster rates of innovation and firm dynamism consistent with the cross-country finding that finance promotes growth through productivity increases (Ayyagari, Demirgüç-Kunt, and Maksimovic, 2007b).

Dropping the cross-country dimension and focusing on an individual country often increases the confidence in the results by reducing potential biases due to measurement error and reducing concerns about omitted variables and endogeneity. In a study of individual regions of Italy, Guiso, Sapienza, and Zingales (2002) use a household dataset and examine the effect of differences in local financial development on economic activity across different regions. They find that local financial development enhances the probability that an individual starts a business, increases industrial competition, and promotes growth of firms. And these results are stronger for smaller firms which cannot easily raise funds outside of the local area. Another example is Haber's (1997) historical comparison of industrial and capital-market development in Brazil, Mexico, and the US between 1830 and 1930. He uses firm-level data to illustrate that international differences in financial development significantly affected the rate of industrial expansion.

Perhaps one of the cleanest ways of dealing with identification problems is to focus on a particular policy change in a specific country and evaluate its impact. One example of this approach is Jayaratne and Strahan's (1996) investigation of the impact of bank branch reform in individual states of the US. Since the early 1970s, US states started relaxing impediments on their intrastate branching. Using a difference-in-difference methodology, Jayaratne and Strahan estimate the change in economic growth rates after branch reform relative to a control group of states that did not reform. They show that bank branch reform boosted bank lending quality and accelerated real per capita growth rates. In another study, Bertrand, Schoar, and Thesmar (2004) provide firm-level evidence from France that shows the impact of 1985 deregulation eliminating government intervention in bank lending decisions fostered greater competition in the credit market, inducing an increase in allocative efficiency across firms. Of course, focusing on individual country cases often raises the question how applicable the results are in different country settings. Nevertheless, these careful country-level analyses boost our confidence in the link between financial development and growth that is suggested by the cross-country studies.

Unfortunately, many potential causal factors of development interest do not vary much within a country, and exogenous policy changes do not occur often enough. For example, besides debates concerning the role of finance in economic development, economists have debated the relative importance of bank-based and market-based financial systems for a long time (Goldsmith, 1969; Boot and Thakor, 1997; Allen and Gale, 2000; and Demirgüç-Kunt and Levine, 2001). Research findings in this area have established that the debate matters much less than was

previously thought, and that it is the financial services themselves that matter more than the form of their delivery. Financial structure *does* change during development, with financial systems becoming more market-based as the countries develop (Demirgüç-Kunt and Levine, 1996). But, controlling for overall financial development, differences in financial structure per se do not help explain growth rates. Nevertheless, these studies do not necessarily imply that institutional structure is unimportant for growth, rather that there is not one optimal institutional structure suitable for all countries at all times. Growth-promoting mixture of markets and intermediaries is likely to be determined by the legal, regulatory, political, policy and other factors that have not been adequately incorporated into the analysis or the indicators used in the literature may not sufficiently capture the comparative roles of banks and markets.

Financial development has also been shown to play an important role in dampening the impact of external shocks on the domestic economy (Beck, Lundberg, and Majnoni, 2006; and Raddatz, 2006), although financial crises do occur in developed and developing countries alike (Demirgüç-Kunt and Detragiache, 1998; Demirgüç-Kunt and Detragiache, 1999; and Kaminsky and Reinhart, 1999). Indeed, deeper financial systems without the necessary institutional development has been shown to lead to a poor handling or even magnification of risk rather than its mitigation. For example, when banking systems grow too quickly, booms are inevitably followed by busts, in which case size and depth may actually reflect policy distortions rather than development, as in numerous country case studies discussed in Demirgüç-Kunt and Detragiache (2005).

Besides issues of identification, problems associated with *measurement* and *nonlinearities* also plague the literature. For example, below a certain level of development, small differences in financial development do not seem to help growth (Rioja and Valev, 2004). Distinguishing between short-run and long-run effects of financial development is also important. Loayza and Rancière (2005) estimate both effects using a pooled mean-group estimator. While they confirm a positive long-run effect, they also identify a negative short-run effect, where short-term surges in bank lending can actually signal the onset of financial crisis as discussed above. Also, financial development may boost income and allow developing countries to catch up, but not lead to an increase in the long-run growth rate. Aghion, Howit, and Mayer-Foulkes (2005) develop a model that predicts that low-income countries with low financial development will continue to fall behind the rest, whereas those reaching the higher level of financial development will converge. Their empirical results confirm that financial development helps an economy converge faster, but that there is no effect on steady-state growth.

Another challenge to the finance and growth literature comes in the form of individual country *outliers*. For example, China is often mentioned as a counterexample to the findings in finance and growth literature since, despite weaknesses in its formal banking system, China is one of the fastest-growing economies in the

world (Allen, Qian, and Qian, 2005). So, is the emphasis on formal financial system development misplaced? Can informal systems substitute for formal systems? Indeed, in China, inter-provincial differences in growth rates are highly correlated with banking debt, but negatively (Boyreau-Debray and Wei, 2005). This emphasizes the importance of focusing on allocation of credit to the private sector, as opposed to all bank intermediation. Hence, mobilizing and pouring funds into the declining parts of the Chinese state enterprise system, as the main Chinese banks were doing, has not been growth promoting. However, focusing on small and medium firms—which account for the most dynamic part of the Chinese economy—shows that those firms receiving bank credit in recent years did tend to grow more quickly compared to those receiving funds from informal sources (Ayyagari, Demirgüç-Kunt, and Maksimovic, 2007). This suggests that the ability of informal mechanisms to substitute for formal financial systems is likely to be exaggerated.

Finance, income distribution, and poverty

If finance promotes growth, over the long term financial development should also help reduce poverty by lifting the welfare of most households. But, do poor households benefit proportionately from financial development? Could there be a widening of income inequalities with the deepening of financial systems? And how important is direct access to financial services in this process?

Theory provides conflicting predictions in this area.[3] Some theories argue that financial development should have a disproportionately beneficial impact on the poor since informational asymmetries produce credit constraints that are particularly binding on the poor. Poor people find it particularly difficult to become entrepreneurs and fund their own investments, or invest in their education internally or externally since they lack resources, collateral, and political connections to access finance (see, e.g., Banerjee and Newman, 1993; Galor and Zeira, 1993; and Aghion and Bolton, 1997). More generally, some political economy theories also suggest that better-functioning financial systems make financial services available to a wider segment of the population, rather than restricting them to politically connected incumbents (Rajan and Zingales, 2003; and Morck, Wolfenzon, and Young, 2005). Yet, others argue that financial access, especially to credit, only benefits the rich and the connected, particularly at early stages of economic development, and therefore, while financial development may promote growth, its impact on income distribution is not clear (Lamoreaux, 1994; and Haber, 2005).

Finally, if access to credit improves with aggregate economic growth and more people can afford to join the formal financial system, the relationship between

[3] See Demirgüç-Kunt and Levine (2007) for an extensive review of the theoretical literature in this area.

financial development and income distribution may be non-linear, with adverse effects at early stages, but a positive impact after a certain point (Greenwood and Jovanovic, 1990). Hence, at the outset, expanding access to finance may actually increase inequality, as new entrepreneurs who manage to finance their investments will experience a surge in their incomes. Only after labor and product market effects start becoming significant, increasing employment opportunities and wages of the poor, we would see a reduction in income inequality. This is, indeed, what Gine and Townsend (2004) find when they build a general equilibrium model of Thai growth and use household data over the 1976–96 period to estimate some of the model's parameters and calibrate others. Their simulations suggest net welfare benefits of financial development to be substantial, though they are initially disproportionately concentrated on a small group of talented, low-income individuals who were unable to become entrepreneurs without access to credit. But, eventually, the greatest impact of financial deepening on income inequality and poverty comes through indirect effects, as more people enter the labor market and the wages increase. Although these calibrated theoretical models illuminate important aspects of the financial development process, their results need to be interpreted with care since, despite their complexity, it is very difficult to model all relevant aspects of the growth and inequality processes.

There is also considerable empirical work on the impact of access to finance on the poor from the microfinance literature (see Armendariz de Aghion and Morduch, 2005). Although success stories of microfinance are well documented in the practitioner literature, a rigorous evaluation requires careful distinction between those changes that can clearly be attributed to financial access from those that might have happened anyway or are due to other changes in the environment in which microfinance clients operate. In other words, identification issues again complicate the analysis. The debate surrounding the most famous microfinance institution, Bagladesh's Grameen Bank, illustrates how difficult this task has been. While Pitt and Khandker (1998) found a significant effect of use of finance on household welfare, more-careful analyses and greater attention to identification issues by Morduch (1998) and Khandker (2003) found insignificant or much smaller effects. There is quite a bit of ongoing research in this area and this research using randomized experiments to address identification issues will probably shed more light on the issue of impact (see World Bank, 2007). However, it is fair to say that, at present, the large body of empirical research evidence on the benefits of microfinance is not conclusive (see Cull, Demirgüç-Kunt, and Morduch, 2008).

But, to evaluate the impact of finance on poverty and income distribution one needs to look beyond the direct impact on the households anyway, since the theoretical models discussed above suggest the spillover effects of financial development through labor and product markets are likely to be significant. Given that these effects cannot be analyzed through micro studies, a more macro approach helps complete the picture.

For example, in cross-country regressions, Beck, Demirgüç-Kunt, and Levine (2007) investigate the relationship between financial depth and changes in both income distribution and absolute poverty. Looking at the 1960–2005 period, they find that not only does a deeper financial system accelerate national growth, but it is associated with a faster increase in the income share of the poorest group. They also find a negative relationship between financial development and the growth rate of the Gini coefficient, suggesting that finance reduces income inequality.[4] These findings are not only robust to controlling for other country characteristics associated with economic growth and changes in income inequality, but the authors make an attempt to control for potential reverse causality using instrumental variables, as well as using panel techniques that control for omitted variable and endogeneity bias.

Although they are able to capture spillover effects, these results obtained in cross-country regressions are subject to caveats given the difficulty of resolving identification issues as discussed above. But, these results are also consistent with the findings of the general equilibrium models which suggest that, in the long run, financial development is associated with reductions in income inequality.

If financial development promotes growth and improves income inequality, it should also reduce poverty. Beck, Demirgüç-Kunt, and Levine (2007) also estimate the change in the share of each country's population below international poverty lines resulting from financial deepening. Again, they find a positive effect of finance on poverty reduction. Countries with higher levels of financial development experience faster reductions in the share of population living on less than a dollar a day over the 1980s and 1990s. Investigating levels rather than growth rates, Honohan (2004) also shows that even at the same average income, economies with deeper financial systems have fewer poor people.

As in the case of finance and growth literature, here, too, further evidence comes from case studies that investigate the impact of specific policy changes better to deal with identification issues. Following the Jayaratne and Strahan (1996) approach discussed above, Beck, Levine, and Levkov (2007) exploit the same policy change to assess the effect of US branch deregulation, this time on income inequality. They find that states see their Gini coefficient decrease by a small but statistically significant amount in the years after deregulation relative to other states, and relative to before the deregulation. They also find that the main decrease on income inequality comes not from enhancing entrepreneurship, but rather through indirect effects of higher labor demand and higher wages.

Another study looks at the branching restrictions policy imposed by the Indian Government between 1977 and 1990, which allowed new branching in a district that already had bank presence, only if the bank opened four branches in districts

[4] Looking at levels, rather than growth rates, Clarke Xu and Zhou (2003) provide further evidence that financial development is associated with lower levels of inequality.

without bank presence. This led to the opening of 30,000 new rural branches over this period. Burgess and Pande (2005) find that this branch expansion during the policy period accounted for 60 percent of rural poverty reduction, largely through an increase in non-agricultural activities and especially through an increase in unregistered or informal manufacturing activities. Although the poverty impact is striking, there were also large losses incurred by the banks due to subsidized interest rates and high loan losses suggesting significant long-term costs.

Although a large body of evidence suggests that financial development reduces income inequality and poverty, we are still far from understanding the channels through which this effect operates. For example, how important is direct provision of finance to the poor? Is it more important to improve the functioning of the financial system so that it expands access to existing firms and households or it is more important to broaden access to the underserved (including the non-poor who are often excluded in many developing countries)? Of course, efficiency and access dimensions of finance are also likely to be linked; in many countries improving efficiency would have to entail broader access beyond concentrated incumbents. Much more empirical research using micro datasets and different methodologies will be necessary better to understand the mechanisms through which finance affects income distribution and poverty.

Qualifications and caveats notwithstanding, taken as a whole, the empirical evidence reviewed in this section suggests that countries with better-developed financial systems grow faster and that this growth disproportionately benefits the poorer segments of the society. Hence, for policymakers, making financial development a priority makes good sense. Yet, financial system development differs widely across countries. What makes some countries develop growth-promoting financial systems, while others cannot? If finance is crucial for economic development, what can governments do to ensure well-functioning financial systems? I turn to these questions next.

Policy choices in finance: Government's role in making finance work

Although finance thrives on market discipline and fails to contribute to development process effectively in the presence of interventionist policies, governments do have a very important role to play in promoting well-functioning financial systems. Below, I discuss different government policies and, where applicable, the evidence on pros and cons of these policies.

Political and macroeconomic environment

Even if historical factors are favorable to financial development, political turmoil may lead to macroeconomic instability and deterioration in business conditions.[5] Civil strife and war destroys capital and infrastructure, and expropriations may follow military takeovers. Corruption and crime thrive in such environments, increasing cost of doing business and creating uncertainty about property rights. Detragiache, Gupta, and Tressel (2005) show that for low-income countries political instability and corruption have a detrimental effect on financial development. Investigating the business environment for eighty countries using firm-level survey data, Ayyagari, Demirgüç-Kunt, and Maksimovic (2005) find that political instability and crime are important obstacles to firm growth, particularly in African and transition countries. Further, Beck, Demirgüç-Kunt, and Maksimovic (2005) show that the negative impact of corruption on firm growth is most pronounced for smaller firms.

Given a stable political system, well-functioning financial systems also require fiscal discipline and stable macroeconomic policies on the part of governments. Monetary and fiscal policies affect the taxation of financial intermediaries and provision of financial services (Bencivenga and Smith, 1992; and Roubini and Sala-i-Martin, 1995). Often, large financing requirements of governments crowd out private investment by increasing the required returns on government securities and absorbing the bulk of the savings mobilized by the financial system. Bank profitability does not necessarily suffer given the high yields on these securities, but the ability of the financial system to allocate resources efficiently is severely curtailed. Empirical studies have also shown that countries with lower and more-stable inflation rates experience higher levels of banking and stock market development (Boyd, Levine, and Smith, 2001) and high inflation and real interest rates are associated with higher probability of systemic banking crises (Demirgüç-Kunt and Detragiache, 1998; and Demirgüç-Kunt and Detragiache, 2005).

Legal and information infrastructure

Financial systems also require developed legal and information infrastructures to function well. Firms' ability to raise external finance in the formal financial system is quite limited if the rights of outside investors are not protected. Outside investors are reluctant to invest in companies if they will not be able to exert corporate governance and protect their investment from controlling shareholders/owners or the management of the companies. Thus, protection of property rights

[5] There is also a large literature that discusses the historical determinants of financial development—such as legal origin, religion and culture, ethnic diversity, and initial geographic endowments. See Ayyagari, Demirgüç-Kunt, and Maksimovic (2006); and Ayyagari, Demirgüç-Kunt, and Maksimovic (2008) for a discussion and evaluation of these theories.

and effective enforcement of contracts are critical elements in financial system development.

Empirical evidence shows firms are able to access external finance in countries where legal enforcement is stronger (La Porta et al., 1997; Demirgüç-Kunt and Maksimovic, 1998; and Beck, Demirgüç-Kunt, and Maksimovic, 2005), and that better creditor protection increases credit to the private sector (Djankov, McLiesh, and Shleifer, 2007). More effective legal systems allow more flexible and adaptable conflict resolution, increasing firms' access to finance (Djankov et al., 2007; and Beck, Demirgüç-Kunt, and Levine, 2005). In countries where legal systems are more effective, financial systems have lower interest rate spreads and are more efficient (Demirgüç-Kunt, Laeven, and Levine, 2004).

Timely availability of good quality information is equally important, since this helps reduce information asymmetries between borrowers and lenders. The collection, processing, and use of borrowing history and other information relevant to household and small business lending—credit registries—have been rapidly growing in both the public and private sectors (see Miller, 2003, for an overview). Computer technology has also greatly improved the amount of information that can be analyzed to assess creditworthiness, such as through credit scoring techniques. Governments can play an important role in this process, and while establishment of public credit registries may discourage private entry, in several cases it has actually encouraged private registries to enter in order to provide a wider and deeper range of services. Governments are also important in creating and supporting the legal system needed for conflict resolution and contract enforcement, and strengthening accounting infrastructures to enable financial development.

Empirical results show that the volume of bank credit is significantly higher in countries with more information sharing (Jappelli and Pagano, 2002; and Djankov, McLeish, and Shleifer, 2007). Firms also report lower financing obstacles with better credit information (Love and Mylenko, 2003). Detragiache, Gupta, and Tressel (2005) find that better access to information and speedier enforcement of contracts are associated with deeper financial systems even in low-income countries. Indeed, compared to high-income countries, in lower-income countries it is credit information more than legal enforcement that matters (Djankov et al., 2007).

Regulation and supervision

For as long as there have been banks, there have also been governments regulating them. While most economists agree that there is a role for government in the regulation and supervision of financial systems, the extent of this involvement is an issue of active debate (Barth, Caprio, and Levine, 2006). One extreme view is the laissez-faire or invisible-hand approach, where there is no role for government in the financial system, and markets are expected to monitor and discipline financial

institutions. This approach has been criticized for ignoring market failures, as depositors, particularly small depositors, often find it too costly to be effective monitors.

On the other extreme is the complete interventionist approach, where government regulation is seen as the solution to market failures (Stigler, 1971). According to this view, powerful supervisors are expected to ensure stability of the financial system and guide banks in their business decisions through regulation and supervision. To the extent that officials generally have limited knowledge and expertise in making business decisions and can be subject to political and regulatory capture, this approach may not be effective (Becker and Stigler, 1974; and Haber, et al., 2003).

Between the two extremes lies the private empowerment view of financial regulation. This view simultaneously recognizes the potential importance of market failures which motivate government intervention, and political/regulatory failures, which suggest that supervisory agencies do not necessarily have incentives to ease market failures. The focus is on enabling markets, where there is an important role for governments in enhancing the ability and incentives of private agents to overcome information and transaction costs, so that private investors can exert effective governance over banks. Consequently, the private empowerment view seeks to provide supervisors with the responsibility and authority to induce banks to disclose accurate information to the public, so that private agents can more effectively monitor banks (Barth, Caprio, and Levine, 2006).

Empirical evidence overwhelmingly supports the private empowerment view. While there is little evidence that empowering regulators enhances bank stability, there is evidence that regulations and supervisory practices that force accurate information disclosure and promote private sector monitoring boost the overall level of banking sector and stock market development (Barth, Caprio, and Levine, 2006).

Beck, Demirgüç-Kunt, and Levine (2006) show that bank supervisory practices that force accurate information disclosure ease external financing constraints of firms, while countries that empower their official supervisors actually make external financing constraints more severe by increasing the degree of corruption in bank lending. Consistent with these findings, Demirgüç-Kunt, Detragiache, and Tressel (2008) investigate compliance with Basel Core Principles of regulation and supervision and show that only information disclosure rules have a significant impact on bank soundness. Finally, Detragiache, Gupta, and Tressel (2005) find little significant impact of regulatory and supervisory practices on financial development of low-income countries. Where there is significance, greater supervisory powers seem to be negatively associated with financial depth.

Related to the debate on different approaches for regulation and supervision is the important debate on whether prudential regulation and safety nets designed for developed countries can be successfully transplanted to developing countries. For

developing countries, these results have important implications for which aspects of the Basel II accord (which was designed for and by regulators in advanced economies) to adopt and over what time period. In particular, the complicated rules and procedures for determining bank capital adequacy presuppose expertise and governance conditions which simply do not exist in most low-income countries. Caprio, Demirgüç-Kunt, and Kane (2008) discuss how the recent financial crisis exposed fundamental flaws in the Basel approach and argue that true reform of regulation and supervision must go beyond improving transparency but address incentive conflicts and increase accountability in government and industry alike.

Similarly, research has questioned safety net design, particularly adoption of deposit insurance in developing countries by highlighting the potential costs of explicit schemes–lower market discipline, higher financial fragility, and lower financial development—in countries where complementary institutions are not strong enough to keep these costs under control (Demirgüç-Kunt and Kane, 2002; Demirgüç-Kunt and Detragiache, 2002; Demirgüç-Kunt and Huizinga, 2004; and Cull, Senbet, and Sorge, 2005). These findings are particularly important for lower-income countries with underdeveloped institutions. For example, Detragiache, Gupta, and Tressel (2005) also find that presence of an explicit deposit insurance system does not lead to more deposit mobilization in low-income countries; to the contrary, it is associated with lower levels of deposits. Demirgüç-Kunt, Kane and Laeven (2008) summarize the cross-country evidence on the impact of deposit insurance and assess the policy complications that emerge in developing countries by reviewing individual country experiences with DI: including issues raised by the European Union's Deposit Insurance directive, banking reform in Russia, and policy efforts to protect depositors in China.

Contestability and efficiency

Policymakers around the world frequently express concern about whether their countries' bank competition policies are appropriately designed to produce well-functioning and stable banks. Globalization and the resulting consolidation in banking have further spurred interest in this issue, leading to an active public policy debate. Competition policies in banking may involve difficult trade-offs. While greater competition may enhance the efficiency of banks with positive implications for economic growth, greater competition may also destabilize banks with costly repercussions for the economy.

Recent research has shown that contrary to conventional wisdom, the trade-offs are exaggerated when it comes to bank competition. Greater competition—as captured by lower entry barriers, fewer regulatory restrictions on bank activities, greater banking freedom, and better overall institutional development—is good for efficiency, good for stability, and good for firms' access to finance (see Berger, et al.,

2004). Indeed, regulations that interfere with competition make banks less efficient, more fragile, and reduce firms' access to finance. Thus, it seems to be a good idea for governments to encourage competition in banking by reducing the unnecessary impediments to entry and activity restrictions. Similarly, improving the institutional environment and allowing greater freedoms in banking and economy in general would lead to desirable outcomes.

Government ownership of financial institutions

Ownership is another important dimension of competition in banking. Policymakers in many countries have felt the need to retain public ownership of banks. However, research has shown that government ownership of banks everywhere, but especially in developing countries, leads to lower levels of financial development, more concentrated lending and lower economic growth, and greater systemic fragility (La Porta, et al., 2002). The inefficient allocation of credit by state-owned banks to politically favored and commercially unviable projects frequently necessitates costly recapitalizations (Cole, 2005; and Dinc, 2005). Even in the area of access to financial services, recent evidence suggests that bank customers face higher barriers to credit services in banking systems which are predominantly government-owned (Beck, Demirgüç-Kunt, and Martinez-Peria, 2007). More recently a handful of government financial institutions have moved away from credit, and evolved into providers of more complex financial services, entering into public–private partnership to overcome coordination failures and first-mover disincentives (De la Torre, Gozzi, and Schmukler, 2008). Ultimately, however, without a large presence of state institutions these initiatives could have been undertaken by the private sector, but the state had a useful role in jump-starting these initiatives. Overall, a large body of empirical evidence suggests that the ownership of financial firms is an area where the public sector tends not to have a comparative advantage; such ownership weakens the financial system and the economy.

Nevertheless, privatization also entails risks and needs careful design. Studies of privatization processes suggest the preferred strategy is moving slowly but deliberately with bank privatization, while preparing state banks for sale and addressing weaknesses in the overall incentive environment. On average, bank privatization tends to improve performance over continued state ownership, there are advantages to full rather than partial privatizations, and in weak institutional environments selling to a strategic investor and inviting foreign interests to participate in the process increase the benefits (see Clarke, Cull, and Shirley, 2005, for an overview). Privatization, however, is not a panacea, and privatizing banks without addressing weaknesses in the underlying incentive environment and market structure will not lead to a deeper and more efficient financial system.

Financial liberalization

In comparison with the scale of global finance, financial systems in individual developing countries are often very small. Small financial systems underperform because they suffer from concentration of risks, cannot exploit economies of scale, and are thus more vulnerable to external shocks. Theoretically, these countries fall short of minimum efficient scale and have much to gain by liberalizing and sourcing some of their financial services from abroad.

There is a very large literature on macroeconomic and international financial issues which is outside the scope of this chapter. In this section, I limit my discussion to a brief review of the impact of financial liberalization on financial development and the importance of sequencing liberalization and institutional reforms; and the impact of foreign entry on financial development.

Financial liberalization, financial development, and the sequencing of reforms

Many countries have liberalized their financial systems in the 1980s and 1990s with mixed results. Liberalization, including deregulation of interest rates and more-relaxed entry policies, often led to significant financial development, particularly in countries where there was significant repression, but the enthusiasm with which financial liberalization was adopted in some countries in the absence of or slow implementation of institutional development also left many financial systems vulnerable to systemic crises (Demirgüç-Kunt and Detragiache, 1999). Poor sequencing of financial liberalization in a poorly prepared contractual and supervisory environment contributed to bank insolvencies as banks protected by implicit and explicit government guarantees aggressively took advantage of new opportunities to increase risk, without the necessary lending skills. Banking crises in Argentina, Chile, Mexico, and Turkey in the 1980s and 1990s have been attributed to these factors (Demirgüç-Kunt and Detragiache, 2005).

On the other hand, many Sub-Saharan African countries that have also liberalized their interest rates and credit allocation and privatized their institutions by allowing entry of reputable foreign banks did not suffer instability but from lower intermediation and in some cases lower access to financial services. Some of this was due to the absence of an effective contractual and informational framework (Honohan and Beck, 2007). This has also resulted in claims of failed liberalizations in these countries and calls for greater government intervention in the financial sector. Both of these experiences with financial liberalization underline the importance of sequencing liberalization and institutional improvements.

Impact of foreign entry

With financial liberalization, more and more developing economies also allow entry of foreign financial institutions. While governments have worried about

whether allowing foreign banks to take a large ownership share in the banking system could damage financial and economic performance, the bulk of the empirical research in this area, particularly drawing on the experience of Latin American and Eastern European countries, suggests that facilitating entry of reputable foreign institutions to the local market should be welcomed. Arrival or expansion of foreign banks can also be disruptive as the Indian experience shows evidence of cream-skimming by foreign banks (Gormley, 2004). Even there, however, in the years following entry, foreign banks have started expanding their clientele base. Overall, a large body of evidence suggests that over time foreign bank entry brings competition, improves efficiency, lifts the quality of the financial infrastructure, and expands access (Claessens, Demirgüç-Kunt, and Huizinga, 2001; and Clarke, Cull, and Martinez-Peria, 2001).

However, as the African experience discussed above illustrates, foreign bank entry cannot guarantee rapid financial development in the absence of sound contractual and informational weaknesses. Such weaknesses can prevent low-income countries from reaping full benefits of opening their markets to foreign providers of financial services, and can potentially explain the finding that greater foreign bank penetration is associated with lower levels of financial development (Detragiache, Tressel, and Gupta, 2006). For example, while in some countries (such as Pakistan), foreign banks have been shown to lend less to smaller more-opaque borrowers because they rely on hard information (Mian, 2006), evidence from Eastern Europe has shown that foreign banks eventually go down-market, increasing small business lending (De Haas and Naaborg, 2005). Overall, addressing institutional weaknesses is likely to allow foreign banks to act as an important catalyst for the sort of financial development that promotes growth.

Facilitating access

Access to financial services has increasingly been receiving greater emphasis over the recent years, becoming a focal part of the overall development agenda. One reason is that modern development theory sees the lack of access to finance as a critical mechanism for generating persistent income inequality, as well as slower growth. Another is the observation that small enterprises and poor households face much greater obstacles in their ability to access finance all around the world, but particularly in developing countries.

What does access to finance mean? Broad access to financial services implies an absence of price and non-price barriers. It is difficult to define and measure because there are many dimensions of access, including availability, cost, and range and quality of services being offered. While there is much data on financial sector development more broadly, until recently there was very little data on usage and access to finance, for both households and firms. Hence, there is also very

limited analysis on the impact of access to finance on economic development. Research using firm-level survey data suggests that financing obstacles are the most constraining among different barriers to growth (Ayyagari, Demirgüç-Kunt, and Maksimovic, 2005). Financing obstacles are also found to be highest and most constraining for the growth of smaller firms (Beck, Demirgüç-Kunt, and Maksimovic, 2005). At the household level, lack of access to credit is shown to perpetuate poverty because poor households reduce their children's education (Jacoby and Skoufias, 1997). Similarly, Dehejia, and Gatti (2003) find that child labor rates are higher in countries with under developed financial systems, while Beegle, et al. (2007) show that transitory income shocks to greater increases in child labor in countries with poorly functioning financial systems. A better understanding of what the chief obstacles to improving access are, and access to which type of financial services has the greater impact on reducing poverty and promoting growth, will need to wait for availability of better data and analysis in this area (see World Bank, 2007 for a discussion).

There are many different reasons why the poor do not have access to finance—loans, savings accounts, insurance services. Social and physical distance from the formal financial system may matter. The poor may not have anybody in their social network who knows the various services that are available to them. Lack of education may make it difficult for them to overcome problems with filling out loan applications, and the small number of transactions they are likely to undertake may make the loan officers think it is not worthwhile to help them. As financial institutions are likely to be in richer neighborhoods, physical distance may also matter: banks simply may not be near the poor. Specifically for access to credit services, there are two important problems. First, the poor have no collateral, and cannot borrow against their future income because they tend not to have steady jobs or income streams to keep track of. Second, dealing with small transactions is costly for the financial institutions. Ceilings on the rates financial institutions can charge backfire and limit access to the poor even more.

Microfinance—specialized institutions that serve the poor—tries to overcome these problems in innovative ways. Loan officers come from similar social status as the borrowers and go to the poor instead of waiting for the poor to come to them. Microcredit also involves education as much as it provides credit. Group lending schemes not only improve repayment incentives and monitoring through peer pressure, but they are also a way of building support networks and educating borrowers.

Has microfinance fulfilled its promise? Microfinance allows poor people to have more direct access, but development of microfinance around the world has been very non-uniform, with significant penetration rates only in a few countries like Bangladesh, Indonesia, and Thailand (Honohan, 2004). Group lending is very costly since labor cost per dollar of transactions needs to be high by design. The most controversial aspect of microfinance, however, has been the extent of subsidy

required to provide this access. Overall, the microfinance sector remains heavily grant- and subsidy-dependent. Skeptics question whether microfinance is the best way to provide those subsidies and point out that development of mainstream finance is a more promising way to reach the poor and alleviate poverty in significant ways.

There are also good political economy reasons why we should not focus on the poor and ask how we can make microfinance more viable, but instead ask how financial services can be made available for all (Rajan, 2006). The poor lack the political clout to demand better services, and subsidies may spoil the 'credit culture'. By defining the issue more broadly to include the middle class, who often also lack access, would make it more likely that promotion of financial access will be made a priority.

What can governments do to promote access? Many of the policies recommended above to enhance the overall development of the financial sector will also help increase access. However, the overlap is not perfect, and explicit prioritization of access is therefore important. For example, certain regulations aimed at financial stability or 'combating terrorism' can restrict access of small firms and poor households. Or focusing on development of offshore financial centers to export wholesale financial services may lead to the neglect of onshore financial infrastructures necessary for access of small firms and individuals. Also, it is important to set realistic goals; not all potential borrowers are creditworthy, and many banking crises were precipitated by overly relaxed credit policies, including the latest crisis of structured securitization. These tensions between improving access without increasing vulnerabilities are discussed in World Bank (2007).

First and foremost, governments can further access by making and encouraging infrastructure improvements. However, prioritizing different reform efforts is important and recent research also suggests that in low-income countries improving information infrastructures seems to yield more immediate access benefits than legal reforms (Djankov, et al., 2007). But legal reforms are also important, and among those there is evidence that while protection of property rights against the state is more important for financial development generally, other aspects of contract enforcement (such as institutions relating to collateral) may be more important for access (Haselmann, Pistor, and Vig, 2006).

Institutional reform is a long-term process and specific policy actions can help boost access sooner. There are a wide range of such measures, ranging from specific legislation to underpin non-blank intermediation, including leasing and factoring; technologies based on the Internet and mobile phones; development of credit registries; protection against money laundering and 'anti-terrorist' financing, without jeopardizing household access and others.

For example, at the household level, giving each individual a national identification number and creating credit registries where lenders share information about their clients' repayment records would help since all borrowers could then borrow

using their future access to credit as collateral (Rajan, 2006). Reducing costs of registering and repossessing collateral is also crucial. In Brazil, for example, inability to repossess property has contributed to the cost of the housing- finance program, keeping the mortgage rates too high to be affordable for the poor. Governments can also be instrumental in facilitating innovative technologies to improve access. For example in Mexico, a program developed by Nafin, a government development bank, allows many small suppliers to use their receivables from large credit-worthy buyers to receive working capital financing (Klapper, 2006). This type of trade finance is called 'reverse factoring' and effectively allows small firms to borrow based on the creditworthiness of their buyers, allowing them to borrow more at cheaper rates.

Government regulation can also help. Removal of interest ceilings, or usury laws, would allow institutions to charge the rates that they need to be profitable and improve access. These regulations end up hurting the very poor they are trying to protect as the supply of these services completely dry up. Anti-predatory lending or truth-in-lending requirements are also very important since households may also be forced into over-borrowing by unscrupulous lenders, as the latest subprime mortgage crisis amply illustrates. Anti-discrimination policies may also help against cases of active or passive discrimination against the poor or different ethnic groups.

It is also important to ensure that other complex regulations—such as Basel II regulations that are intended to help banks minimize costly bank failures—do not inadvertently penalize small borrowers and hurt access by failing to make full allowance for the potential for a portfolio of small and medium-size enterprise (SME) loans to achieve risk pooling. Financial regulations can also prevent the emergence of institutions better suited to the needs of lower-income households or smaller firms. Rigid chartering rules, high capital adequacy requirements, or very strict accounting requirements may reduce the ability of institutions to serve the poorer segments of the society. As many households are interested in savings services but not in credit services, considering and regulating savings mobilization separately from credit services may be helpful (Claessens, 2005). For example, in South Africa, extension of bank regulation and supervision to microfinance institutions reduced their capacity to offer their services profitably.

Governments can also opt to stimulate access more directly. The US Treasury's electronic transfer accounts (ETAs) to increase use of bank accounts, the US Community Reinvestment Act (CRA) to improve access to credit services, and legal measures adopted by the UK, France, Sweden, and Ireland, among others, are such examples. However, there is little consensus on the success of those schemes (Claessens, 2005) and whether they can be replicated in developing countries. The experiences with credit extensions, especially to improve the maturity structure of debt and reach the SMEs, are extensive in both developed and developing countries. However, both the rationale for and effectiveness of those interventions

are much more doubtful (see Caprio and Demirgüç-Kunt, 1997; and Beck and Demirgüç-Kunt, 2006). As already discussed above, interventions through ownership of government institutions have also not been successful, overall.

Last, but perhaps most importantly, governments can improve access by increasing competition in the financial sector. As financial institutions find their traditional business coming under competition they seek out new lines of profitable opportunities, including lending to the SMEs and the poor. Given the right incentives, the private sector can develop and make use of new technologies—like credit scoring—to reach the underserved segments. As already discussed above, foreign banks' role in improving the competition environment and improving access is important. There is accumulating evidence that, over time, foreign banks can enhance access. Indeed, multinational banks have been leading the way in expanding access all around the world.

Conclusions

As this chapter was still being written, the financial turmoil that started as a meltdown in structured securitization instruments in the summer of 2007 in the US and UK quickly spread and has become a full-blown financial crisis. In an effort to contain the crisis from spreading, the authorities in the US and many European governments have taken unprecedented steps of providing extensive liquidity, giving assurances to bank depositors and creditors that include blanket guarantees, structuring bailout programs that include taking large ownership stakes in financial institutions, in addition to establishing programs for direct provision of credit to non-financial institutions. These policy responses to the crisis have shaken the confidence of developed and developing countries alike in the very blueprint of financial sector policies that underlie western capitalist systems.

Demirgüç-Kunt and Serven (2008) draw on a large body of econometric evidence and country experience to argue that the financial sector policy advice provided in this chapter is still valid. For the most part, the confusion arises from not being able to recognize incentive conflicts and trade-offs inherent in short-term and long-term responses to a systemic crisis. Policies employed to contain a crisis—often in a haste to re-establish confidence and with inadequate consideration of long-term costs—should not be interpreted as permanent deviations from well-established policy positions. The fact that governments may end up providing blanket guarantees or owning large stakes in the financial sector in an effort to contain and deal with the crisis does not negate the fact that generous guarantees over the long term are likely to backfire or that government officials make poor bankers.

In conclusion, should all countries follow the recommendations outlined in this chapter? While the general messages will not be dissimilar, the directions in which the financial sector needs improvement in different countries will be based on their initial conditions (World Bank, 2001; and World Bank, 2007). Furthermore, good policymaking draws inputs from many sources, and research is only one such input. Implementation of policy requires complementing the results of research analysis with practitioner experience, Hence, tempering and tailoring this advice to individual country circumstances. In general, these reforms are likely to be most challenging for low-income countries, where the legacy of financial repression and state ownership has generally hampered the development of a vigorous private financial system, where the underlying legal and information infrastructure is weak, and achieving minimum efficient scale will be difficult.

Despite their inherent fragility, financial systems underpin economic development. The challenge of financial sector policies is to align private incentives with public interest without taxing or subsidizing private risk taking. The task is becoming increasingly complex for all countries in an ever more integrated and globalized financial system.

References

Aghion, P. and Bolton, P. (1997). 'A Trickle-Down Theory of Growth and Development with Debt Overhang', *Review of Economic Studies,* 64: 151–72.

—— Howitt, P., and Mayer-Foulkes, D. (2005). 'The Effect of Financial Development on Convergence: Theory and Evidence', *Quarterly Journal of Economics,* 120: 173–222.

Allen, F. and Gale, D. (2000). *Comparing Financial Systems.* Cambridge, MA: MIT Press.

—— Qian, J., and Qian, M. (2005). 'Law, Finance, and Economic Growth in China', *Journal of Financial Economics,* 77: 57–116.

Armendariz de Aghion, B. and Morduch, J. (2005). *The Economics of Microfinance.* Cambridge, MA: MIT Press.

Ayyagari, M., Demirgüç-Kunt, A., and Maksimovic, M. (2005). 'How Important are Financing Constraints? The Role of Finance in the Business Environment', Mimeo, World Bank.

—— —— —— (2006). 'What Determines Protection of Property rights? An Analysis of Direct and Indirect Effects', Mimeo, World Bank.

—— —— —— (2007a). 'Formal versus Informal Finance: Evidence from China', Mimeo, World Bank.

—— —— —— (2007b). 'Firm Innovation in Emerging Markets: Role of Finance'. Mimeo, World Bank.

—— —— —— (2008). 'How Well Do Institutional Theories Explain Firms' Perceptions of Property Rights?', *The Review of Financial Studies,* 21: 1833–71.

Banerjee, A. and Newman, A. (1993). 'Occupational Choice and the Process of Development', *Journal of Political Economy,* 101: 274–98.

Barth, J., Caprio Jr., G., and Levine, R. (2006). *Rethinking Bank Regulation: 'Till Angels Govern.* New York: Cambridge University Press.

Beck, T. and Demirgüç-Kunt, A. (2006). 'Small and Medium-Sized Enterprises: Access to Finance as a Growth Constraint', *Journal of Banking & Finance*, 30: 2931–43.

——— Laeven, L., and Levine, R. (2004). 'Finance, Firm Size, and Growth', World Bank Policy Research Working Paper 3485.

——— and Levine, R. (2003). 'Law, Endowments, and Finance', *Journal of Financial Economics,* 70: 137–81.

——— ——— (2005). 'Law and Firms' Access to Finance', *American Law and Economics Review*, 7: 211–52.

——— ——— (2006). 'Bank Supervision and Corruption in Lending', *Journal of Monetary Economics*, 53: 2131–63.

——— ——— (2007). 'Finance, Inequality and the Poor', *Journal of Economic Growth*, 12: 27–49.

——— and Maksimovic, V. (2005). 'Financial and Legal Constraints to Firm Growth: Does size matter?', *Journal of Finance,* 60: 137–77.

——— ——— (2006). 'Financing Patterns around the World: Are Small Firms Different?', Mimeo, World Bank.

——— and Martinez-Peria, M. S. (2007). 'Banking Services for Everyone? Barriers to Bank Access and Use around the World', World Bank Policy Research Working Paper 4079.

——— Levine, R., and Levkov, A. (2007). 'Financial Regulation and Income Distribution: Evidence from Branch Deregulation', Mimeo, Brown University.

——— and Loayza, N. (2000). 'Finance and the Sources of Growth', *Journal of Financial Economics*, 58: 261–300.

——— Lundberg, M., and Majnoni, G. (2006). 'Financial Intermediary Development and Growth Volatility: Do Intermediaries Dampen or Magnify Shocks?', *Journal of International Money and Finance*, 25: 1146–67.

Becker, G. and Stigler, G. (1974). 'Law Enforcement, Malfeasance, and the Compensation of Enforcers', *Journal of Legal Studies*, 3: 1–18.

Beegle, K., Dehejia, R., and Gatti, R. (2007). 'Child Labor and Agricultural Shocks'. *Journal of Development Economics,* 81: 80–96.

Bekaert, G., Harvey, C. R., and Lundblad, C. (2005). 'Does Financial Liberalization Spur Growth?', *Journal of Financial Economics*, 77: 3–55.

Bencivenga, V. R. and Smith, B. D. (1992). 'Deficits, Inflation and the Banking System in Developing Countries: The Optimal Degree of Financial Repression', *Oxford Economic Papers* 44: 767–90.

Berger, A., Demirgüç-Kunt, A., Haubrich, J., and Levine, R. (2004). 'Introduction: Bank Concentration and Competition: An Evolution in the Making', *Journal of Money, Credit, and Banking*, 36: 433–53.

Bertrand, M., Schoar, A. S., and Thesmar, D. (2004). 'Banking Deregulation and Industry Structure: Evidence from the French Banking Reforms of 1985', Discussion Paper 4488, Center for Economic Policy Research.

Boot, A. W. A. and Thakor, A. (1997). 'Financial System Architecture', *Review of Financial Studies*, 10: 693–733.

Boyd, J. H., Levine, R., and Smith, B. D. (2001). 'The Impact of Inflation on Financial Sector Performance', *Journal of Monetary Economics*, 47: 221–48.

Boyreau-Debray, G. and Wei, S.-J. (2005). 'Pitfalls of a State-Dominated Financial System: The Case of China', National Bureau of Economic Research Working Paper 11214.

Burgess, R. and Pande, R. (2005). 'Can Rural Banks Reduce Poverty? Evidence from the Indian Social Banking Experiment', *American Economic Review*, 95: 780–95.

Caprio, G. and Demirgüç-Kunt, A. (1997). 'The Role of Long-Term Finance: Theory and Evidence'. *World Bank Economic Review*, 10: 291–321.

—— —— and Edward Kane (2008), 'The 2007 Meltdown in Structured Securitization: Searching for Lessons not Scapegoats', World Bank Working Paper.

Christopoulos, D. K. and Tsionas, E. G. (2004). 'Financial Development and Economic Growth: Evidence from Panel Unit Root and Cointegration Tests', *Journal of Development Economics*, 73: 55–74.

Claessens, S. (2005). 'Access to Financial Services: A Review of the Issues and Public Policy Objectives', World Bank Policy Research Working Paper 3589.

—— Demirgüç-Kunt, A., and Huizinga, H. (2001). 'How Does Foreign Entry Affect Domestic Banking Markets?', *Journal of Banking & Finance*, 25: 891–911.

Clarke, G., Cull, R., and Martinez-Peria, M. S. (2001). 'Does Foreign Bank Penetration Reduce Access to Credit in Developing Countries: Evidence from Asking Borrowers', World Bank Policy Research Working Paper 2716.

—— —— and Shirley, M. (2005). 'Bank Privatization in Developing Countries: A Summary of Lessons and Findings', *Journal of Banking & Finance*, 29: 1905–30.

—— Xu, L. C. and Zhou, H. (2003). 'Finance and Income Inequality: Test of Alternative Theories', World Bank Policy Research Working Paper 2984.

Cole, S. (2005). 'Fixing Market Failures or Fixing Elections? Elections, Banks, and Agricultural Lending in India', Mimeo, Harvard Business School.

Cull, R., Demirgüç-Kunt, A., and Morduch, J. (2008). Microfinance: The Next Capitalist Revolution?, *Journal of Economic Perspectives*, <http://wagner.nyu.edn/faculty/cv/morduch_cv.pdf>.

—— Senbet, L., and Sorge, M. (2005). 'Deposit Insurance and Financial Development', *Journal of Money, Credit, and Banking*, 37: 43–82.

De Haas, R. and Naaborg, I. (2005). 'Does Foreign Bank Entry Reduce Small Firms' Access to Credit? Evidence from European Transition Economies', Dutch National Bank Working Paper 50.

De la Torre, A., Gozzi, J. C. and Schmukler, S. (2008). *Innovative Experiences in Access to Finance: Market Friendly Roles for the Visible Hand?*, Stanford, CA: Stanford University Press and the World Bank.

Demirgüç-Kunt, A. and Detragiache, E. (1998). 'The Determinants of Banking Crises: Evidence from Developing and Developed Countries', IMF Staff Papers 45, 81–109.

—— —— (1999). 'Financial Liberalization and Financial Fragility', in B. Pleskovic and J. Stiglitz (eds.), *Proceedings of the Annual World Bank Conference on Development Economics,* The World Bank, Washington, DC, 332–4.

—— —— (2002). 'Does Deposit Insurance Increase Banking System Stability? An Empirical Investigation', *Journal of Monetary Economics*, 49: 1373–406.

—— —— (2005). 'Cross-Country Empirical Studies of Systemic Bank Distress: A Survey', *National Institute Economic Review*, 192: 68–83.

—— —— and Tressel, T. (2008). 'Banking on the Principles: Compliance with Basel Core Principles and Bank Soundness', *Journal of Financial Intermediation*, 17: 511–42.

——and Huizinga, H. (2004). 'Market Discipline and Deposit insurance', *Journal of Monetary Economics*, 51: 375–99.
——and Kane, E. (2002). 'Deposit Insurance around the Globe: Where Does it Work?', *Journal of Economic Perspectives*, 16: 175–96.
————and Laeven, L. (2008). *Deposit Insurance around the World: Issues of Design and Implementation*. Cambridge, MA: MIT Press.
——Laeven, L. and Levine, R. (2004). 'Regulations, Market Structure, Institutions, and the Cost of Financial Intermediation', *Journal of Money, Credit, and Banking*, 36: 593–622.
——and Levine, R. (1996). 'Stock Market Development and Financial Intermediaries: Stylized Facts', *World Bank Economic Review*, 10: 291–322.
————(2001). *Financial Structure and Economic Growth: A Cross-Country Comparison of Banks, Markets, and Development*. Cambridge, MA: MIT Press.
————(2007). 'Finance and Economic Opportunity', World Bank Working Paper.
——and Maksimovic, V. (1998). 'Law, Finance, and Firm Growth', *Journal of Finance*, 53: 2107–37.
——and Serven, L. (2008). 'Are All the Sacred Cows Dead? Implications of the Financial Crisis for Macro and Financial Policies', Mimeo, World Bank.
Detragiache, E., Gupta, P., and Tressel, T. (2005). 'Finance in Lower-Income Countries: An Empirical Exploration', International Monetary Fund Working Paper 05/167.
Dinc, S. (2005). 'Politicians and Banks: Political Influences on Government-Owned Banks in Emerging Countries', *Journal of Financial Economics*, 77: 453–79.
Djankov, S., McLiesh, C., and Shleifer, A. (2007). 'Private Credit in 129 Countries', *Journal of Financial Economics*, 84: 299–329.
Galor, O. and Zeira, J. (1993). 'Income Distribution and Macroeconomics', *Review of Economic Studies*, 60: 35–52.
Gine, X. and Townsend, R. (2004). 'Evaluation of Financial Liberalization: A General Equilibrium Model with Constrained Occupation Choice', *Journal of Development Economics*, 74: 269–307.
Goldsmith, R. W. (1969). *Financial Structure and Development*. New Haven, CT: Yale University Press.
Gormley, T. A. (2004). 'Banking Competition in Developing Countries: Does Foreign Bank Entry Improve Credit Access?', Working Paper, Washington University, John M. Olin School of Business, St. Louis.
Greenwood, J. and Jovanovic, B. (1990). 'Financial Development, Growth, and the Distribution of Income', *Journal of Political Economy*, 98: 1076–9107.
Guiso, L., Sapienza, P., Zingales, L. (2002). 'Does Local Financial Development Matter?', National Bureau of Economic Research Working Paper 8922.
Haber, S. H. (1997). 'Financial Markets and Industrial Development: A Comparative Study of Governmental Regulation, Financial Innovation and Industrial Structure in Brazil and Mexico, 1840–1940', in S. Haber (ed.), *How Latin America Fell Behind?* Stanford, CA: Stanford University Press, 146–78.
——(2005). 'Mexico's Experiments with Bank Privatization and Liberalization, 1991–2003', *Journal of Banking & Finance*, 29: 2325–50.
——Maurer, N., and Razo, A. (2003). *The Politics of Property Rights: Political Instability, Credible Commitments, and Economic Growth in Mexico, 1876–1929*. New York: Cambridge University Press.

Haselmann, R. F. H., Pistor, K., and Vig, V. (2006). 'How Law Affects Lending', Columbia Law and Economics Working Paper 285.

Honohan, P. (2004). 'Financial Development, Growth, and Poverty: How Close are the Links?', in C. Goodhart (ed.), *Financial Development and Economic Growth: Explaining the Links*. London: Palgrave.

—— and Beck, T. (2007). *Making Finance Work for Africa*. Washington, DC: World Bank.

Jacoby, H. G. and Skoufias, E. (1997). 'Risk, Financial Markets, and Human Capital in a Developing Country', *The Review of Economic Studies*, 643: 311–35.

Jappelli, T. and Pagano, M. (2002). 'Information Sharing, Lending and Defaults: Cross-Country Evidence', *Journal of Banking & Finance*, 26: 2017–45.

Jayaratne, J. and Strahan, P. E. (1996). 'The Finance-Growth Nexus: Evidence from Bank Branch Deregulation', *Quarterly Journal of Economics*, 111: 639–70.

Kaminsky, G. and Reinhart, C. M. (1999). 'The Twin Crises: The Causes of Banking and Balance of Payments Problems', *American Economic Review*, 89: 473–500.

—— and Schmukler, S. (2003). 'Short-Term Pain, Long-Term Gain: The Effect of Financial Liberalization', National Bureau of Economic Research Working Paper 9787.

Khandker, S. R. (2003). 'Microfinance and Poverty: Evidence Using Panel Data from Bangladesh', World Bank Policy Research Working Paper 2945.

King, R. G. and Levine, R. (1993). 'Finance and Growth: Schumpeter Might Be Right', *Quarterly Journal of Economics*, 108: 717–38.

Klapper, L. (2006). 'The Role of Factoring for Financing Small and Medium Enterprises', *Journal of Banking & Finance*, 30: 3111–30.

Lamoreaux, N. (1994). *Insider Lending: Banks, Personal Connections, and Economic Development in Industrial New England*. New York: Cambridge University Press.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. (2002). 'Government Ownership of Commercial Banks', *Journal of Finance*, 57: 265–301.

—— —— —— (2006). 'What Works in Securities Laws?', *Journal of Finance*, 61: 1–32.

—— —— ——, and Vishny, R. W. (1997). 'Legal Determinants of External Finance', *Journal of Finance*, 52: 1131–50.

Levine, R. (1997). 'Financial Development and Economic Growth: Views and Agenda', *Journal of Economic Literature*, 35: 688–726.

—— (2005). 'Finance and Growth: Theory and Evidence', in P. Aghion and S. Durlaff (eds.), *Handbook of Economic Growth*. The Netherlands: Elsevier Science.

—— Loayza, N., and Beck, T. (2000). 'Financial Intermediation and Growth: Causality and Causes', *Journal of Monetary Economics*, 46: 31–77.

—— and Zervos, S. (1998). 'Stock Markets, Banks, and Economic Growth', *American Economic Review*, 88: 537–58.

Loayza, N. and Rancière, R. (2005). 'Financial Development, Financial Fragility, and Growth', IMF Working Paper 05/170.

Love, I. and Mylenko, N. (2004). 'Credit Reporting and Financing Constraints', World Bank Policy Research Working Paper 3142.

Lucas, R. E. (1988). 'On the Mechanics of Economic Development', *Journal of Monetary Economics*, 22: 3–42.

Merton, R. C. and Bodie, Z. (2004). 'The Design of Financial Systems: Towards a Synthesis of Function and Structure', National Bureau of Economic Research Working Paper 10620.

Mian, A. (2006). 'Distance Constraints: The Limits of Foreign Lending in Poor Economies', *Journal of Finance* 61: 1465–505.

MILLER, M. (2003). *Credit Reporting Systems and the International Economy*. Cambridge, MA: MIT Press.

MORCK, R., WOLFENZON, D., and YEUNG, B. (2005). 'Corporate Governance, Economic Entrenchment, and Growth', *Journal of Economic Literature*, 43: 655–720.

MORDUCH, J. (1998). 'Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh', Mimeo, Princeton University.

PITT, M. M. and KHANDKER, S. R. (1998). 'The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?', *Journal of Political Economy*, 106: 958–96.

RADDATZ, C. (2006). 'Liquidity Needs and Vulnerability to Financial Underdevelopment', *Journal of Financial Economics*, 80: 677–722.

RAJAN, R. G. (2006). 'Separate and Unequal', *Finance and Development*, 43: 56–7.

—— and ZINGALES, L. (1998). 'Financial Dependence and Growth', *American Economic Review*, 88: 559–86.

—— —— (2003). *Saving Capitalism from the Capitalists*. New York: Random House.

RIOJA, F. and VALEV, N. (2004). 'Does One Size Fit All? A Reexamination of the Finance and Growth Relationship', *Journal of Development Economics*, 74: 429–47.

ROBINSON, J. (1952). *The Rate of Interest and Other Essays*. London: MacMillan (Chapter 'The Generalization of the General Theory').

ROUBINI. N. and SALA-I-MARTIN, X. (1995). 'A Growth Model of Inflation, Tax Evasion, and Financial Repression', *Journal of Monetary Economics*, 35: 275–301.

ROUSSEAU, P. L. and SYLLA, R. (1999). 'Emerging Financial Markets and Early U.S. Growth', National Bureau of Economic Research Working Paper 7448.

STIGLER, G. (1971). 'The Theory of Economic Regulation', *Bell Journal of Economics and Management Science*, 2: 3–21.

World Bank (2001). *Finance for Growth: Policy Choices in a Volatile World*. A World Bank Policy Research Report, Washington, DC: World Bank.

—— (2007). *Finance for All: Policies and Pitfalls in Expanding Access*. A World Bank Policy Research Report, Washington, DC: World Bank.

CHAPTER 30

BANKING AND REAL ECONOMIC ACTIVITY

NICOLA CETORELLI

Introduction

On 12 December A.D. 56, Lucius Caecilius Jucundus recorded a transaction related to a loan for 11,039 sesterces, which he had extended for the completion of an auction sale that took place in the city of Pompeii.[1] Wax tablets recording this and sixteen other similar loan contracts were found, charred but still legible, in an archeological excavation of his house, partially destroyed after the Vesuvius eruption on 24–5 August 79.

L. Caecilius Jucundus was a very wealthy banker in Pompeii, the son of a freedman, who had become a banker himself. They were bankers in the basic definition of the term, in that they would accept deposits from clients and extend loans using part of the received deposits. The standard terms for such loan contracts implied a commission plus an interest rate that was normally about 2 percent a month, with a typical duration of up to one year but normally no more than just a few months.

[1] As a term of reference, in that same period, a laborer's wage was 2–4 sesterces per day, and the average price for the purchase of a slave was 2,000 sesterces (Stambaugh, 1988).

Their main role, and the one documented in the above mentioned tablet, was to provide credit at auctions for the sale of property, harvests, and slaves. In many cases they would arrange the sale of the very same collateral that had been pledged on past due loans that could not be repaid. They would also act as assayers of coins, provide foreign exchange services, extend types of loans other than those related to auction transactions, and engage in activities resembling what would currently be defined as trust management (Andreau, 1999: 36).

While there is significant written evidence of banking activity in the first and second centuries A.D in Rome, professional bankers were already in operation in Athens back in the fifth century B.C. and were found at about the same time in Egypt and Palestine (Andreau, 1999: 30–2).

Banks have thus been present even in the earliest instances of pre-modern, pre-capitalist societies, their role so pervasive and ingrained in the basic functioning of markets and economies that one almost wonders about the need to discuss the importance of banking institutions for the real economy. However, a debate on the basic determinants of the process of economic development has been alive and kicking at least throughout the entire twentieth century. Within this debate, the role of banks, and the financial sector in general, has been either readily dismissed (see, e.g., Robinson, 1952; and Lucas, 1988) or alternatively recognized as 'too obvious for serious discussion' (Miller, 1998: 14). The start of the modern analysis of this issue is normally associated with the work of Joseph Schumpeter, who synthesized the idea that credit, especially bank credit, does create real value, in his *Theory of Economic Development* (1911). However, Schumpeter himself drew on an even older debate when he challenged, for instance, the view of Ricardo that 'banking operations cannot increase a country's wealth' (Schumpeter, 1911: 98). This chapter first illustrates the reasons why the debate went on for so long, and goes on to make the claim that perhaps scholars have finally reached a consensus. Second, it presents the most current directions of research on this topic.

The causality debate

Perhaps the main reason for the persistence of this debate is that it has been very difficult to pin down the issue of causality. Anecdotal evidence from historical case studies, or even broader informal observations from cross-country data, will normally show a strong, positive correlation between any standard measure of real economic activity—per capita output growth, per capita capital growth, and productivity growth—with standard measures of development of financial markets. For instance, in what is widely recognized as the first contribution to reignite

the most current interest in this debate, King and Levine (1993) drew on data from seventy-seven countries from 1960 to 1989 to show that a basic measure of the 'depth' of the financial system—the aggregate value of currency demand and interest-bearing liabilities of banks and non-bank intermediaries—had a positive and economically very significant association with real economic activity.

The underlying idea behind this and related measures of the size of the financial sector is that a broader, deeper financial sector increasingly facilitates firms' access to capital. This is the basic 'financial engine' of growth. More precisely, King and Levine show that if a country could increase the size of its financial sector from the bottom quartile to the top quartile of the distribution, the resulting superior access to capital would be reflected in an increase in per capita income growth by almost 1 percent per year. Given that the difference in income growth between countries in the top and the bottom quartile of the distribution for the countries in this dataset over this sample period was about 5 percent, the change in financial depth would contribute to an impressive 20 percent reduction of such a gap (King and Levine, 1993: Table 7).

In subsequent work, a more specific link between just *bank* credit and real economic activity was also confirmed using similar data and empirical methodology. Levine and Zervos (1998) show with a dataset for forty-two countries between 1976 and 1993 that an increase in bank credit by one standard deviation results in an increase in real per capita income growth of 0.7 percentage points per year (Levine and Zervos, 1998: Table 3). The basic message of such studies is that the economic magnitude of increasing the overall scale of the banking industry is potentially very significant. However, despite the robustness and the strength of such results, skeptics in the underlying debate have always maintained that while this empirical evidence clearly indicates an important correlation between finance and real activity, it cannot address the fundamental issue at stake—namely, whether banking activity—and by extension activity of the financial sector—is somewhat exogenously determined and, if it is, whether it exerts an independent impulse on real economic sectors. Critics of the role of finance in real economic activity have always argued that characteristics of financial markets are *endogenously* determined—that is, the existence and the development of anything financial is simply a reflection of real economic activity. The empirical evidence mentioned above cannot disprove the argument that financial markets simply develop simultaneously to accommodate the expanding needs of a growing economy, or even that measures to deepen financial activity could be undertaken in anticipation of predicted future economic growth.

Much subsequent work, most of it by Levine and coauthors, has addressed specifically this issue of endogeneity and causality. It has done so by departing from basic cross-sectional analysis and embracing the more sophisticated econometric tools of dynamic panel estimation techniques (see, e.g., Levine, Loayza, and Beck, 2000; and Beck, Levine, and Loayza, 2000) and instrumental variables

(Levine, 1998; Levine, 1999; and Levine, Loayza, and Beck, 2000). In essence, the basic strategy involves trying to identify an exogenous component of financial development. This is achieved assuming that the level of development of the financial sector in a country is very much a reflection of the quality of the basic institutional setting that has developed in that country over time. In turn, such institutional setting (reflected in the degree of protection of property rights, in the quality of the legal enforcement system, in the overall level of trust, in the degree of corruption, etc.) is found to be highly determined by the legal origin of that country (see La Porta, et al., 1998). More specifically, the nature and quality of basic institutions appear to be highly correlated with whether the legal system of a country has roots in the British, German, French, or Scandinavian traditions of rule of law. Since the basic activity of financial markets relies on the possibility of writing well-defined contracts describing transactions based on promises of future payments, financial markets will be more or less developed to the extent to which the legal system allows protection and enforcement of such contracts. And, since the establishment of a given legal system in a country is to a large extent the result of past events, such as experiences of colonization, it is plausible to consider this feature as exogenously determined.

Thus, by either augmenting basic cross-sectional studies with instrumental variable analysis, or by using instrumental variables in dynamic panel models, the studies mentioned above came to conclusions that were remarkably similar to the earlier ones that relied on simpler identification techniques. That is (the exogenous component of) financial development has a substantial economic impact on the real economy.

While this represented an important step in addressing the causality issue, questions could still be raised regarding the quality of the instruments and, perhaps most important, the fact that both the quality of the financial sector and that of other important institutions could all be determined by still other omitted factors. Hence, the doubt remains, following this approach, that an observed positive effect of financial sector variables could in fact be the reflection of something else affecting simultaneously the financial and the real side (see Zingales, 2003).

No less important, however, is the interpretation that we can give to studies that have used the measures of depth illustrated earlier to capture the importance of financial markets for the real economy. Depth, or size, is really an *outcome* measure, meaning that whatever it is that is done or could be done to improve the financial industry is then reflected in its relative size, which is what is observed. Yet, by focusing on this end-result variable, we are at least one step removed from addressing causality in that we do not directly investigate *how* banks or other parts of the financial industry generate an independent impact on real economic activity. This leaves unanswered the above-mentioned criticisms that financial markets evolve hand in hand with other economic variables and that those are the ones actually responsible for real sector growth. What is more, by maintaining

the focus on the depth variables, the analysis is also much constrained in terms of quality of its normative content: if it is probably the case that going from the bottom to the top quartile in the banking size distribution is associated with considerably higher income growth, these studies are not able to prescribe exactly how deepening can be achieved.

Banks matter

Both the issue of causality and of normative content were directly addressed in another highly influential study, Jayaratne and Strahan, 1996, which represented another significant leap forward in the quest for the ultimate word on the role of banking for the real sector. The authors narrowed the focus of analysis down from a cross-country perspective to a country-specific case study, that of the US. While the choice may seem deficient with respect to the broader cross-country variability of previous studies, it actually comes with a tremendous payoff for a study on the role of banking: as a result of decades of regulatory restrictions preventing or limiting bank expansion within and/or across states, up to the mid-1970s the US featured what was in effect fifty separate banking markets (Morgan, Rime, and Strahan, 2004), with state lines demarcating the boundaries of each individual market. Hence, studies limited to just US banking still allows substantial cross-sectional variation. At the same time, the narrower focus on one country also reduces potentially important sources of unobservable heterogeneity that are more likely to plague multi-country data. Moreover, and most importantly, the end of the 1970s marks the beginning of an intense process of deregulation, in which individual states—*at different points in time*—removed regulatory barriers that had prevented bank entry. By the mid-1990s the process had concluded, allowing banks originally headquartered anywhere from that point on to expand potentially anywhere else without restriction.

As a result of this process of deregulation, banking markets have become increasingly more competitive and efficient. This should in turn translate into more credit availability, a clear and direct effect on the real economy. This is exactly what Jayaratne and Strahan (1996) test. The simultaneous existence of cross-sectional and over-time variation concerning individual states' timing of deregulation represents a unique opportunity to conduct analysis in conditions that approximate those of a 'natural experiment', a scenario notoriously hard to achieve in social science inquiry. More precisely, it is possible to measure the impact of bank deregulation—and the associated changes in competition and efficiency—comparing state-specific real economic variables *before and after* deregulation. In the language

of natural experiment analysis, the control group is represented by observations in a state before deregulation, including observations in other states that have not deregulated yet, while the treatment group is represented by all observations in years following deregulation. Because deregulation is not implemented at the same time in all states, unobserved state-specific omitted factors and time-specific events common across states that could explain the dependent variable can be absorbed by state and time indicator variables, still leaving sufficient variation to identify the specific effect of deregulation.

This identification strategy makes an important step forward in dealing with causality for at least two reasons: first, it does not capture developments in the financial sector, banking in particular, by looking at an *ex post* outcome performance such as credit size, capturing instead the effect of a specific event, bank deregulation, that is supposed to generate developments in the sector. And, because theory would suggest that the resulting improvement in competition and efficiency should be associated with more and better allocated capital, the causal link is now much more direct. Second, the event in question can be plausibly considered to be exogenous and occurring independently of current or expected developments in the real economy. For example, studies have indicated that small banks had been very influential in establishing tight restrictions to expansion as early as the 1930s and that their influence remained strong through the early 1980s (Economides, Hubbard, and Palia, 1996; and White, 1998). Also, the extensive failure of thrifts in the 1980s has been considered another cause of deregulation, as large, better-diversified banks were allowed to acquire the failing banks (Kane, 1996). Finally, Kroszner and Strahan (1999) find that technological changes in both deposit and lending activity were among the leading factors behind deregulation.

With these premises, Jayaratne and Strahan (1996) were able to find evidence confirming a causal link between banking deregulation and state income growth. In particular, using a panel from 1972 to 1992, they find that income growth in a state was more than half a percentage point higher, per year, after deregulation of its banking industry. Their contribution should be recognized both methodologically, for the important tightening in the strategy to address causality, but also because it focused on a specific characteristic of the banking industry, thereby bringing the data closer to theory and at the same time enhancing the normative content of the analysis. Their evidence represents, in my opinion, the closest to a nail in the coffin of the causality debate. After this paper, it has become very difficult to counter Schumpeter's assertion that 'bank credit does create value', or at the very least the burden of proof has shifted squarely to the other side of the debate.[2]

[2] A parallel paper that should also be considered as a turning point, although focusing on the broader relationship between overall financial development and economic growth, is Rajan and Zingales (1998). Their contribution is described in detail in the previous chapter.

How do banks matter?

From this point on, the research frontier advances forward. No longer is it necessary to expend effort making the point that banks are important for the real economy. Taking that as a given, scholars can now focus on the perhaps richer and more satisfying quest of fully understanding the *mechanisms* through which banks can affect the real economy. The operative questions at stake are what specific characteristics of banks and of the banking industry are likely to matter the most for real output variables, such as income or productivity growth? And, similarly, from the other end, what specific elements or features of the real sectors of the economy are really affected by banks' activity, so that ultimately such activity is reflected in an impact on real output? Delving deeper into the *micro* details governing the banking–real-economy relationship, it is now possible really to put to the test specific theories of banking. Moreover, and as mentioned earlier, the normative value of the newest studies of banking and the real economy increases tangibly. As economists fine tune what works the most, policymakers are increasingly able to navigate the sometimes turbulent waters of banking regulatory activity.

The role of bank competition

In the decade following Jayaratne and Strahan (1996), research work in this field has evolved in multiple directions, in which emphasis was directed at the explicit features and characteristics of the banking industry. A relatively large amount of work, in particular, has been—and continues to be—dedicated to studying the role played by bank *competition* for the real economy. The reasons for the attention paid to this characteristic of the industry are twofold. First, in contrast to most industries, where the default is that market structure and competitive conduct evolve endogenously, the banking industry has historically been heavily regulated, and for the most varying reasons, both in the US and in other countries. Hence, it is plausible to make the case that this is an exogenously determined characteristic of the industry when studying the impact on the real economy. Second, and not less important, there is a fascinating contrast of theoretical conjectures that can be formulated about the effects of bank competition. Petersen and Rajan (1995) expressed very clearly the essence of this contrast. The authors challenged the conventional view that enhancing bank competition necessarily leads to better loan terms and better access to credit. The theoretical argument is that in fact banks need at least some degree of market power to have the right incentives to undertake the proper investments in screening and monitoring necessary to resolve uncertainty about the quality of new entrepreneurs. The intuition is that in the absence

of some ability to 'capture' the client firm over time, a bank anticipates that an entrepreneur that turns out to be successful has the possibility in future periods to seek better terms from competing banks that would not need to incur any additional cost of screening and monitoring (or would spend just a fraction of what the original bank had to). Hence, in a highly competitive banking environment, banks would be required to charge loan terms reflecting the high intrinsic risk of the entrepreneurs. A bank with market power could instead offer better initial terms knowing that any upfront cost in starting such a lending relationship could be recuperated at later stages. The unconventional prediction that follows is that firms, especially young ones, might have better access to credit if they operate in more concentrated banking markets.[3] Petersen and Rajan (1995) test this prediction using US data for more than 3,000 respondent firms to the 1988 National Survey of Small Business Finance, matched by geographic location using FDIC Summary of Deposits information on the location of bank offices. The bank-specific data allowed them to construct measures of local market concentration, which they used as a proxy for market power.

As they specified in the paper, there can be multiple sources of market power, but local market concentration seeks to capture market power derived from spatial location. The idea is that firms, especially the small ones they focus on, are bound to have very local relationships with a bank and therefore their proximity to banking centers, and the related density of such centers in the firm's location, give an idea of how 'captive' the firm can be (Petersen and Rajan, 1995: 417–18). The authors find that young firms operating in markets with high bank concentration are more likely to access credit. They also find that loan terms are more advantageous (lower lending rates) than in less concentrated markets. This term differential disappears and in fact reverses as firms become more mature.

Another paper to focus on the role of bank competition for the real economy is Cetorelli and Gambera, 2001. This paper explored the empirical relevance of the market structure of the banking sector for industrial growth. The authors took the basic cross-sectional work initiated by King and Levine (1993) and then asked: if it is agreed that the size of the banking industry is important to capital accumulation, does it matter whether the underlying industry structure is unconcentrated, thus approximating perfectly competitive conditions, or whether instead market power is concentrated among few banking institutions? From a theoretical standpoint, Cetorelli and Gambera played with the same antagonism of conjectures presented in Petersen and Rajan (1995). Their methodological approach built on the contribution of Rajan and Zingales (1998), in that they used a cross-country dataset but

[3] The basic ideas behind the Petersen and Rajan contribution were already present in Schumpeter, 1911 and formulated in Mayer, 1988. For additional theoretical work see, e.g., Rajan (1992), Pagano (1993), Shaffer (1998), Manove, Padilla, and Pagano (2001), Marquez (2002), Dell'Ariccia and Marquez (2004), Hauswald and Marquez (2006), Boot and Thakor (2000), and Boyd and De Nicolo (2005).

looked at the differential impact of bank concentration in a country *across* industrial sectors that for their own idiosyncratic reasons display varying degree of dependence on external sources of finance for capital investment. The identification strategy is then based upon the intuition that if bank competition has a role, it should matter more for firms in sectors that are highly dependent on external finance availability. As in Rajan and Zingales (1998), by seeking such a differential effect, the identification strategy raises considerably the bar for potential objections on ground of endogeneity, omitted variable biases and reverse causality. The findings suggest a non-trivial impact of bank concentration on industrial growth and in fact simultaneous support for both sides of the theoretical controversy. First, there is evidence that bank concentration has a first-order negative effect on growth. This finding is consistent with the theoretical prediction that higher bank concentration results in a lower amount of credit available in the economy *as a whole*. Regardless of their external financial dependence, this effect is common to all industrial sectors. However, the paper also finds evidence that bank concentration has a *heterogeneous* effect across industries. In particular, sectors where young firms are more dependent on external finance enjoy a beneficial effect from a concentrated banking sector, which could actually more than compensate the first-order, negative effect. This finding supports the basic argument in Petersen and Rajan (1995) predicting that concentration of market power in banking facilitates the development of lending relationships, which have in turn an enhancing effect on firms' growth.

Banks and industry dynamics in product markets

As the research agenda on bank competition and the real economy picked up momentum, it also became more ambitious. Much current work has gone, for example, into the understanding of how bank competition can actually affect the life-cycle dynamics of industrial sectors of production. For instance, does more bank competition mean more entry in non-financial industries? And what is the related impact for *incumbent* firms? Would changes in bank competition lead to *structural* changes in other industries, such as affecting average firm size, or the whole firm size distribution?

As recounted in Cetorelli (2004: 545–6), the role of bank competition on these characteristics of non-financial industries had not really been explored before, at least in the mainstream economic literature. Scattered evidence is found in the work of history scholars. For example, in his study of Italian industrialization in the late nineteenth century, Cohen (1967) describes how a quasi-monopolistic banking industry 'led to the emergence of concentration of ownership and control in the new and rapidly growing sectors of the industrial structure'. Capie and Rodrik-Bali (1982) note that the intense process of consolidation and increase in concentration

that characterized British banking in the early 1890s preceded that experienced later on by manufacturing industrial sectors. Similarly, Haber (1991) reports over a century of Mexican history, between 1830 and 1930, a very close connection between bank and industry concentration. The general impression from historical studies that bank concentration should be associated with concentrated industries is finally expressed by Cameron in his renowned 1967 study on banking in the early stages of industrialization, where he states that

> Competition in banking is related to the question of competition in industry. In general the two flourish—and decline—together. Whether this phenomenon is a joint by-product of other circumstances, or whether it results from the decline or restriction of competition among banks, is a matter worthy of further research. It is a striking coincidence, in any case, that industrial structure—competitive, oligopolistic, or monopolistic—tends to mirror financial structure. (Cameron, 1967: 313)

At the same time, while important as analyses of countries' economic development, the empirical evidence presented in these studies is limited by their focus on specific countries, periods, and socio-institutional circumstances. Lacking in the literature had been broader empirical analyses apt to make general assertions about the role of banking market structure on industries' market structure.

A lack of systematic empirical evidence on this relationship is also accompanied by scattered formal theoretical modeling to guide the implementation of empirical identification strategies. Nevertheless, we can delve into the existing literature on the economic role of banking market structure to formulate alternative theoretical conjectures. To this end, the framework proposed by Petersen and Rajan (1995) described above represents a good foundation from which to ponder the role of bank concentration on industry concentration. In their reasoning, banks with market power fund young firms with the expectation that they will be capable of extracting future rents once those firms eventually become profitable. Consequently, one could then argue that banks with market power, following their goal of profit maximization, should always attempt to select the best available pool of entrepreneurs, thus favoring new entrants along the entire life-cycle of an industry. This is because new entrants are potentially endowed with higher-return projects and more innovative technologies that would guarantee ever-increasing profit-sharing opportunities for the banks. Thus, according to this logic, bank concentration should continuously foster entry and therefore contribute to enhance industry competition. This is a testable hypothesis. Yet, maintaining the same premises in the Petersen and Rajan model, it is also legitimate to envision a completely different set of economic forces at play that could lead to the opposite conclusion. Consider, for example, a nascent industry where a bank with market power may indeed facilitate credit access to young firms. Once lending relationships are established, however, at later stages the bank may have an incentive to preserve its ties with the older clients—now industry incumbents—and constrain

access to credit to newer entrants. The argument is that by increasing market competition the newcomers would undermine the market power and therefore the profitability of industry incumbents and consequently the profitability of the bank itself. In essence, this argument follows from the recognition that market power gives banks an implicit equity stake in the firms they are already financing, thus potentially distorting their incentive to extend credit in product markets. This theoretical argument would then suggest that bank concentration should enhance industry concentration. Judging by the formulation of these alternative conjectures, the effect of bank concentration on industry market structure is therefore theoretically ambiguous.

These alternative conjectures have been brought to the data. A first example is represented by Black and Strahan (2002). Regulatory action that removes entry barriers and leads to a more efficient and competitive banking industry should be reflected in a direct effect on entrepreneurship. More precisely, the entry rates of business entrepreneurships should be higher following the kind of deregulation illustrated above. Using data on new business incorporations between 1976 and 1996, Black and Strahan compare the number of new incorporations, and the growth rate of new incorporations before and after deregulation, using the same identification methodology in Jayaratne and Strahan (1996). The authors find that after deregulation that first allowed banks freely to branch within a state, the number of new businesses increased by almost 10 percent, and its growth rate between 3 and 4 percent. Subsequent to deregulation that also allowed banks to expand across state lines, the number of new businesses increased further by about 6 percent, while there was no significant effect on the growth rate. The economic impact on new business creation of opening up banking markets is therefore very large. Moreover, allowing more businesses to be in operation may be reflected in increasing output accumulation. Hence, analyzing the relationship between banking deregulation and business formation puts substance in the original goal of understanding the mechanism through which banking activity can affect real economic activity and ultimately long-term economic growth.

At the same time, going back to the theoretical conjectures outlined earlier on, while it is certainly consistent with theory that bank competition enhances business entry, it may not be inconsistent to speculate that banks may also have, as noted earlier, diminished incentives to screen and monitor entrepreneurs. Consequently, more entry may also be associated with higher mortality of young firms. A first take on this issue was presented in Cetorelli (2003). The author used a public version of the US Bureau of the Census dataset on business establishments that contains information on age categories to test—among other things—the impact of banking deregulation on entry, but also on the persistence rates of young businesses, as measured by rates of job destruction. The evidence was consistent with that presented by Black and Strahan (2002) on entry. Moreover, it suggested that the persistence of younger businesses was actually higher after deregulation (hence,

lower job destruction). In contrast, using the more detailed, confidential version of the census data, Kerr and Nanda (forthcoming) find that the failure rates (complete business shutdown) of enterprises three years and younger are actually higher after deregulation. Since the authors also find evidence consistent with more entry, they interpret the combination of results as suggesting that improvements in bank competition favor a 'democratization' of the entry process. That is, potential entrepreneurs have a greater chance to start a business but they do not necessarily have a greater chance of surviving and remaining in business.

More work has been conducted related to the effects of bank competition on life-cycle dynamics. In the above-mentioned paper, Cetorelli (2003) also looks at the growth rates of incumbent enterprises and at their own persistence rates. The evidence there suggests that *less* bank competition meant *delayed exit* of incumbent firms, a finding consistent with the idea exposed before that monopolistic banks may have distorted incentives in favoring their older clients, thus preventing a healthier process of creative destruction.

Still with a focus on life-cycle dynamics, Cetorelli and Strahan (2006) attempted a broad analysis of the impact of more bank competition on a complete set of metrics of the market structure of non-financial industries. The authors use US census data between 1977 and 1994 on the number of business establishments and their size—measured by employment level—located in the different states and operating in any of the twenty, two-digit SIC manufacturing sectors. Consistent with previous findings, they show evidence that more vigorous competition leads to more firms in operation. In addition, they also find that the average firm size decreases as banks become more competitive. Lower average firm size is consistent with the finding of more firms in operation, and both add strength to the idea that more bank competition favors entry and allows entry at a smaller scale.

Finally, they find that the whole firm size distribution is shifted, with an increase in mass toward smaller-size firms. The additional evidence on the size distribution adds content to the conjecture that the impact of improved competition may be felt differently by different firms—young or old, small or large—and in different sectors. More firms in operation and smaller average size could reflect entry by very small establishments. If that were simply the case, one would expect an increase in mass at the smallest end of the size distribution and declines in mass elsewhere in the distribution. If better bank competition also helps the existing small firms grow (due to an overall increased supply of financial resources), then we ought to see a greater proportion not only of the smallest but also of medium-sized establishments as well. Moreover, testing for shifts in the whole size distribution allows us to compare how the shares of small and medium-sized (presumably bank-dependent) establishments behave relative to another sort of control group—namely, the share of the very largest establishments. These establishments (those with 1,000 or more employees) should not be affected by banking conditions because very large firms have access to nationwide (and competitive) securities

markets. Thus, their fortunes should not vary with local credit conditions (Cetorelli and Strahan, 2006: 455).

Finally, and still on the relationship of bank competition to industry life-cycle, Cetorelli (2009) investigates whether the process of banking deregulation could be so important as to render firms after deregulation intrinsically different from those that started their operations prior to deregulation. This idea, commonly referred to in corporate demography as 'imprinting' effect, posits that in a time when external capital was relatively harder to obtain, prospective firms needed a set of organizational and managerial characteristics that would increase their chance not only to obtain (scarcer) financing, but also to survive in the event of constraints to obtaining additional credit. Conversely, after deregulation, and the removal of important frictions to credit supply, new firms may not need to develop that set of characteristics that were previously required, thus resulting in a group of intrinsically more fragile units. Cetorelli (2009) finds preliminary evidence consistent with this conjecture. Firms born prior to deregulation seem to have a 'thicker skin' than firms with similar characteristics born after deregulation.

Analyzing the role of bank competition for the real economy has thus proven to be a quite rich investment. One last contribution that followed a similar thread, but with a significantly different focus is Morgan, Rime, and Strahan (2004), which tackled the very relevant issue of the impact of bank deregulation for macroeconomic stability. The issue is potentially debatable. On the one hand, the opening up of markets should have a beneficial effect on market stability as it enhances opportunities for risk sharing across markets. On the other hand, as mentioned earlier, banks in a more competitive environment may be less apt to efficiently allocate credit. Also, as suggested by Keeley (1990), increased competition may have a perverse effect on banks' risk taking behavior because competition reduces banks' franchise value. The authors inquire about this issue and test the response of state-specific measures of business cycle volatility to the deregulation events. They find that volatility drops substantially, between 30 and 40 percent, after deregulation.

The material reviewed in this chapter has been developed using datasets drawn almost exclusively from US banks and US sectors of production. A legitimate criticism could be made about whether these conclusions have broader, international validity. Recent work based on specific country studies or cross-sections of non-US countries seems to confirm the findings examined so far. A very ingenious paper drawing extensively on the comparison of many decades of economic history is Haber (1991). The author essentially presents a horse race between Mexico, Brazil, and the US starting from the earliest stages of industrialization and then following the evolution of industrial structure in response to developments in capital markets. Haber focuses his attention on the evolution of the textile industry between approximately the 1840s and the 1930s. The reason to

focus on such an industry is that, aside from the fact that it had a much higher relevance in the past, it also had specific characteristics (including low entry barriers, divisibility of capital, and scale efficiencies exhausted at small size) such that the industry was not naturally prone to encourage concentration. The only substantial barrier to entry was the ability to access external capital. Focusing on such industry represents a good case study to understand the finance–real-economy relationship, because any change in the financial sector that may have occurred in the three countries could convincingly be argued not to be endogenously determined by events in the textile industry.

Haber documents very persuasively the relationship between efficiency, regulatory reforms—or lack thereof—in the financial sector, and the structure of the textile industry. Specifically, he shows that the US adopted important reforms of both the banking industry and capital markets early on. In particular, the National Banking Act of 1863 produced important entry in the banking industry that facilitated access to credit in the production sector. As Haber remarks, 'by the end of World War I the textile industry was awash in finance and many companies took advantage of the swollen credit markets to float numerous securities issue' (Haber, 1991: 564).

The experience of Mexico was just the opposite. The textile industries, and production in general, had no access either to bank finance or to capital markets throughout a large part of the nineteenth century. When banks eventually emerged, they actually developed as institutions tightly connected to a limited number of entrepreneurs, who were themselves connected to government officials. Hence, external finance remained severely limited and differentially available in the market. The result was a textile industry that evolved as highly concentrated. Brazil had a very similar experience and developmental trajectory as Mexico, at least initially. Financial markets were extremely underdeveloped, but in 1890 important reforms sowed the seeds for significant expansions of the banking sector and capital markets. As a result, the structure of the textile industry in Brazil in the first decades of the twentieth century looked a lot more like that of the US than that of Mexico.

Another example of a non-US country study is the experience of France in response to the reform of the banking industry of 1985, analyzed by Bertrand, Schoar, and Thesmar (2007). The French reform reduced significantly government intervention in banks' lending activity. This reform lead naturally to a boost in efficiency in private banks and enhanced competition in the credit market. The reforms to the banking industry, the authors document, generated important changes in the microeconomic behavior of firms and had a related strong impact on the structure of industries. In relation to the topics analyzed in this chapter, the authors show that the regulatory reform of the banking sector in France had a positive effect on firm entry and exit rates and a negative effect on product market concentration.

Still on baking reforms, but with a focus on Europe, Cetorelli (2004) investigated the effect of the implementation of the Second Banking Directive of the European Union—a regulatory reform that essentially created the conditions for an integrated European banking market—on industrial structure. Cetorelli used a panel of manufacturing industries in twenty-nine Organization for Economic Cooperation and Development (OECD) countries, both European Union and non-European Union members. The evidence showed that enhanced competition in European Union banking markets lead to markets in non-financial sectors characterized by lower average firm size. This conclusion is consistent with the findings in Cetorelli and Strahan (2004) and points at a beneficial effect of bank competition on credit access to young firms.

Finally, Beck, Demirgüç-Kunt, and Maksimovic (2004) use data from a World Bank survey of firms in a large cross-section of countries. One of the questions the survey asked was whether firms had difficulty in obtaining credit. Matching this and other information with the specific market structure of the banking sector where those firms were located, the authors found that higher bank concentration was associated with more financing obstacles, especially for smaller firms.

Two important points can be made regarding the role of bank competition in enhancing real economic activity. First, there does not seem to be a Pareto-dominant policy regarding the optimal banking market structure: competition in banking does not necessarily dominate monopoly, and vice versa. Second, regulation of the financial industry is intimately related to industrial policy. Depending on the level of concentration of the banking industry, *ceteris paribus*, individual sectors will grow at different speeds. Therefore, banking market structure plays an important role in shaping the cross-industry size distribution within a country.

Importance of ownership structure

As mentioned earlier, by delving deeper into the mechanics of the banking–real-economy relationship, scholars have been able to broaden significantly the policy implications associated with their findings. In addition to the numerous, separate threads of research related to bank competition, a good deal of emphasis has also been placed on the characteristics of banks' ownership structure and on the ownership connections between banks and firms. A well-established view in corporate finance stresses the value of close ties between banks and borrowers. Bankers that are represented on the borrower's board of directors and are able to follow closely the day-to-day operations of the borrowers are also able to perform careful assessments of the firm's investment strategies, growth opportunities, and overall risk exposure. Access to such information allows the bank to make efficient credit

allocation decisions. An alternative view stresses the possibility that related lending may be subject to important conflicts of interests, such that a bank may have incentives to allocate credit to a firm with which it shares close relations not necessarily on the basis of standard risk–return considerations.

Empirical evidence has offered support for both views. Gorton and Schmid (2000) found a positive effect on performance for German firms with close bank ties. A positive effect on liquidity needs and resolution of financial distress is found for Japanese firms (Hoshi, Kashyap, and Stein, 1990). However, a negative effect on firm performance, firm growth, and cost of capital is found on Japanese firms by Weinstein and Yafeh (1998). A negative effect on default rates is found for Mexican firms by La Porta, et al. (2003) and a negative effect on growth and on capital allocation is found by Maurer and Haber (2004) using nineteenth-century records for Mexican firms. Finally, Krozner and Strahan (2001) find that US firms with close bank ties are not treated differently from similar firms without bank ties.

Conclusions

It is not by chance that Lucius Caecilius Jucundus established himself and prospered in the city of Pompeii. Pompeii was a well-developed center with close proximity to the sea and where markets were held on a regular basis. It is well recognized that bankers like him were instrumental in facilitating and developing commercial activity. And, while the contours of entrepreneurship in ancient Rome may not fit a modern profile, it is clear that Jucundus and others still played a role in assisting productive activity (Andreau, 1999: 145–52).

The tale of Lucius Caecilius Jucundus serves as a good example to illustrate the crux of the debate around the role of banks for the real economy. It is certainly the case that banks—and financial activity—follow where real activity goes. The direction of causality from economics to finance, in other words, has never been seriously questioned. Much harder to prove is that banking can develop independently of what goes on in the real economy and that developments in the banking industry can in fact alter economic activity.

The impulse to map these dynamics has inspired a lively body of literature, one that reflects both the intrinsic intellectual interest on the issues at stake but also their vast policy implications. After all, the pervasive nature of policy control of the banking industry rests on certain assumptions about banks' fundamental role in the real economy. After more than a decade of rigorous research we are probably now in a position to assert with a significant degree of confidence that banking

does matter for real economic activity. We have not only learned that banking activity has a large impact on various measures of output growth but have also made important progress in understanding exactly how that happens. Developments in the way banks operate, as reflected, for example, in their competitive conduct or on specific ownership structures, bring with them far-reaching implications for economic activity. Pushing research further in this direction is expected to continue to yield significant results with no diminishing returns in sight yet.

References

Andreau, J. (1999). *Banking and Business in the Roman World (Key Themes in Ancient History)*, Trans. Janet Lloyd, Cambridge: Cambridge University Press.

Beck, T., Demirgüç-Kunt, A., and Maksimovic, V. (2004). 'Bank Competition and Access to Finance: International Evidence', *Journal of Money, Credit, and Banking*, 36: 627–48.

—— Levine, R., and Loayza, N. (2000). 'Finance and the Sources of Growth', *Journal of Financial Economics,* 58: 261–300.

Bertrand, M., Schoar, A., and Thesmar, D. (2007). 'Banking Deregulation and Industry Structure: Evidence from the French Banking Reforms of 1985', *The Journal of Finance*, 62: 597–628.

Black, S. E. and Strahan, P. E. (2002). 'Entrepreneurship and Bank Credit Availability', *Journal of Finance*, 57: 2807–33.

Boyd, J. and De Nicolo', G. (2005), 'The Theory of Bank Risk-Taking and Competition Revisited', *Journal of Finance*, 60: 1329–43.

Cameron, R. (1967). *Banking in the Early Stages of Industrialization,* New York, NY: Oxford University Press.

Cao, M. and Shi, S. (2001). 'Screening, Bidding, and the Loan Market Tightness', *European Finance Review*, 5: 1–2.

Capie, F. and Rodrik-Bali, G. (1982). 'Concentration in British Banking, 1870–1920', *Business History*, 24: 280–92.

Cetorelli, N. (2003). 'Life-Cycle Dynamics in Industrial Sectors. The Role of Banking Market Structure', *Federal Reserve Bank of St. Louis Review*, 85: 135–47.

—— (2004). 'Real Effects of Bank Competition', *Journal of Money, Credit and Banking*, 36: 543–58.

—— (2009). 'Credit Market Competition and the Nature of Firms', Federal Reserve Bank of New York Staff Report 366.

—— and Gambera, M. (2001). 'Banking Market Structure, Financial Dependence and Growth: International Evidence from Industry Data', *Journal of Finance*, 56: 617–48.

—— and Peretto, P. F. (2009). 'Credit Quantity and Credit Quality. Bank Competition and Capital Accumulation', Federal Reserve Bank of New York Staff Report 375.

—— and Strahan, P. (2006). 'Finance as a Barrier to Entry: Bank Competition and Industry Structure in Local U.S. Markets', *Journal of Finance*, 61: 437–61.

Cohen, J. (1967). 'Financing Industrialization in Italy, 1894–1914', *Journal of Economic History* 27: 363–82.

Dell'Ariccia, G. (2000). 'Learning by Lending, Competition, and Screening Incentives in the Banking Industry', Working Paper, IMF.

—— and Marquez, R. (2004), 'Information and Bank Credit Allocation', *Journal of Financial Economics*, 72: 185–214.

Economides, N., Hubbard, G. R., and Palia, D. (1996). 'The Political Economy of Branching Restrictions and Deposit Insurance: A Model of Monopolistic Competition among Small and Large Banks', *Journal of Law & Economics*, 39: 667–704.

Gorton, G. and Schmid, F. A. (2000). 'Universal Banking and the Performance of German firms', *Journal of Financial Economics,* 58: 29–80.

Guzman, M. (2000). 'Bank Structure, Capital Accumulation, and Growth: A Simple Macroeconomic Model', *Economic Theory*, 16: 421–55.

Haber, S. (1991). 'Industrial Concentration and the Capital Markets: A Comparative Study of Brazil, Mexico, and the United States, 1830–1930', *Journal of Economic History* 51, 559–80.

Hauswald, R. and Marquez, R. (2006). 'Competition and Strategic Information Acquisition in Credit Markets', *Review of Financial Studies*, 19: 967–1000.

Hoshi, T., Kashyap, A., and Scharfstein, D. (1990). 'The Role of Banks in Reducing the Costs of Financial Distress in Japan', *Journal of Financial Economics*, 27: 67–88.

Jayaratne, J., and Strahan, P. E. (1996). 'The Finance-Growth Nexus: Evidence from Bank Branch Deregulation', *Quarterly Journal of Economics*, 111: 639–70.

Kane, E. J. (1996). 'De Jure Interstate Banking: Why Only Now?', *Journal of Money, Credit, and Banking*, 28: 141–61.

Keeley, M. C. (1990). 'Deposit Insurance, Risk, and Market Power in Banking', *American Economic Review*, 80: 1183–200.

Kerr, W. and Nanda, C. (forthcoming). 'Democratizing Entry: Banking Deregulation, Financing Constraints and Entrepreneurship', *Journal of Financial Economics*.

King, R. G. and Ross, L. (1993). 'Finance and Growth: Schumpeter Might Be Right', *Quarterly Journal of Economics*, 108: 717–37.

Kroszner, R. S. and Strahan, P. E. (1999). 'What Drives Deregulation? Economics and Politics of the Relaxation of Bank Branching Restrictions', *Quarterly Journal of Economics*, 114: 1437–67.

—— —— (2001). 'Bankers on Boards: Monitoring, Conflicts of Interest, and Lender Liability', *Journal of Financial Economics*, 62: 415–52.

La Porta, R., Lopez-de-Silanes, F., and Shleifer, A. (2002). 'Government Ownership of Banks', *Journal of Finance*, 57: 265–301.

—— —— —— and Vishny, R. W. (1998). 'Law and Finance', *Journal of Political Economy*, 106: 1113–55.

—— —— —— and Zamarrita, G. (2003). 'Related Lending', *Quarterly Journal of Economics*, 118: 231–67.

Levine, R. (1998). 'The Legal Environment, Banks, and Long Run Economic Growth', *Journal of Money, Credit, and Banking*, 30: 596–613.

—— (1999). 'Law, Finance and Economic Growth', *Journal of Financial Intermediation*, 8: 8–35.

—— Loayza, N., and Beck, T. (2002) 'Financial Intermediation and Growth: Causality and Causes', *Journal of Monetary Economics*, 46: 31–77.

—— and Zervos, S. (1998). 'Stock Markets, Banks, and Economic Growth', *American Economic Review*, 88: 537–58.

Lucas, R. (1998). 'On the Mechanics of Economic Development', *Journal of Monetary Economics*, 22: 3–42.

Manove, M., Padilla, J., and Pagano, M. (2001). 'Collateral vs. Project Screening: A Model of Lazy Banks', RAND *Journal of Economics*, 32.

Marquez, R. (2002). 'Competition, Adverse Selection, and Information Dispersion in the Banking Industry', *Review of Financial Studies*, 15: 901–26.

Maurer, N. and Haber, S. (2002). 'Institutional Change and Economic Growth: Banks, Financial Markets, and Mexican Industrialization, 1878–1913', in *Mexican Economy, 1870–1930*, J. L. Bortz and S. Haber (eds.), Stanford, CA: Stanford University Press.

——— (2007). 'Related Lending and Economic Performance: Evidence from Mexico', *Journal of Economic History*, 67: 551–81.

Mayer, C. (1988). 'New Issues in Corporate Finance', *European Economic Review*, 32: 1167–83.

Miller, M. H. (1998). 'Financial Markets and Economic Growth', *Journal of Applied Corporate Finance*, 11: 8–14.

Morgan, D. P., Rime, B., and Strahan, P. E. (2004). 'Bank Integration and State Business Cycle', *Quarterly Journal of Economics*, 119: 1555–85.

Pagano, M. (1993). 'Financial Markets and Growth. An Overview', *European Economic Review*, 37: 613–22.

Petersen, M. A. and Rajan, R. G. (1995). 'The Effect of Credit Market Competition on Lending Relationships', *Quarterly Journal of Economics*, 110: 407–43.

Rajan, R. G. (1992), 'Insiders and Outsiders: The Choice between Informed and Arm's-Length Debt', *Journal of Finance*, 47: 1367–400.

——— and Zingales, L. (1998). 'Financial Dependence and Growth', *American Economic Review*, 88: 559–86.

Robinson, J. (1952). *The Rate of Interest and Other Essays*. London: Macmillan.

Schumpeter, J. (1911). *The Theory of Economic Development: An Inquiry into Profits, Capital, Interest, and the Business Cycle*, Translated from German by Redverse Opie, Brunswick, NJ: Transaction Books, 1983.

Shaffer, S. (1998). 'The Winner's Curse in Banking', *Journal of Financial Intermediation*, 7: 359–92.

Stambaugh, J. E. (1988). *The Ancient Roman City*, Baltimore, Md: Johns Hopkins University Press.

Weinstein, D. E. and Yafeh, Y. (1998). 'On the Costs of a Bank-Centered Financial System: Evidence from the Changing Main Bank Relations in Japan', *Journal of Finance*, 53: 635–72.

White, E. (1998). 'The Legacy of Deposit Insurance: The Growth, Spread, and Cost of Insuring Financial Intermediaries', in M. Bordo, C. Goldin, and E. N. White (eds.), *The Defining Moment: The Great Depression and the American Economy in the Twentieth Century*. Chicago: University of Chicago Press.

Zingales, L. (2003). 'Commentaries to More on Finance and Growth: More Finance More Growth? by Ross Levine', *Federal Reserve Bank of St. Louis Review*, 85.

PART V

INTERNATIONAL DIFFERENCES IN BANKING STRUCTURES AND ENVIRONMENTS

This page intentionally left blank

CHAPTER 31

BANKING IN THE UNITED STATES

ROBERT DEYOUNG

Introduction

In the US, the concept of 'banking' has meant different things at different times. For most of the twentieth century, 'banking' services were produced and delivered by a potpourri of financial institutions in separate industry segments—for example, commercial banks, investment banks, thrift institutions, insurance companies, finance companies—which, due to relatively rigid financial regulations, offered largely separate lines of business and did not compete with each other. Among these institutions, commercial banks and other depository institutions were traditionally the largest and offered the greatest number of financial services. For example, as shown in Table 31.1, depositories in the US (commercial banks, thrift institutions, and credit unions) held over half of all US financial assets in 1970, compared with just 17 percent for insurance companies, the next largest category of financial institution. Although their role in financial markets has changed somewhat over time, commercial banks and other depository institutions continue to be important if not dominant providers of financial services for many US businesses and households today. This chapter will focus primarily on commercial banking, and will view other US banking sectors and institutions through the prism of commercial banking.

Over the past quarter-century, change has been the most salient characteristic of US banking markets. Commercial banks, thrifts, and life insurers have lost market

Table 31.1. Distribution of assets at US financial intermediaries in 1970 and 2007

	1970 (%)	2007 (%)
Depository institutions (banks, thrifts, credit unions)	54.4	22.8
Insurance companies	17.4	10.5
Pension funds (public and private)	14.6	16.9
Finance companies	4.9	3.2
Mortgage finance companies and funds *	3.9	20.2
Mutual funds (stock, bond, money market)	3.7	18.5
Securities firms (brokers, dealers, funding corporations)	1.1	7.9
Total	100.0	100.0

Source: Federal Reserve System Flow of Funds Accounts.

* Includes government-sponsored enterprises (GSEs) and pools they sponsor, private mortgage securitizers and pools they sponsor, mortgage banks, and real estate investment trusts (REITs).

share, owing to the combined impacts of new information technologies, new financial instruments and markets, financial deregulation, and substantial increases in competition within and across industry segments. By 2007, the share of industry assets held by depository institutions had fallen by more than half to around 23 percent. Where are the majority of US financial assets held today? Mutual funds (stock, bond, and money market funds) increased their share of US financial assets from less than 4 percent in 1970 to more than 18 percent in 2007; mortgage finance companies (including off-balance sheet mortgage investment pools financed by asset-backed securities) increased their holdings from just 4 percent to 20 percent; and securities firms (brokers, dealers, and funding corporations) increased their holdings from just 1 percent to nearly 8 percent. In addition, banks' financial fortunes have become suddenly more volatile than in the recent past, with wide swings in profits and losses making the US banking industry a far different place from the stable and strictly regulated industry of just a generation ago. To understand the US banking industry today, one must first understand how and why this evolution occurred.

The evolution of the US banking industry

In the US, a commercial bank can operate under either a national bank charter or a state bank charter. Federal law gives the Office of the Comptroller of the Currency

(OCC), a bureau of the US Treasury Department, the authority to grant national bank charters and serve as the primary regulator and supervisor of national banks. State laws give each of the fifty state governments in the US the authority to grant state banking charters, and the banking commissions in each state share supervisory and regulatory authority over these banks with the Federal Deposit Insurance Corporation (FDIC) and the Federal Reserve (the Fed). These two federal agencies have other bank regulatory duties as well; the FDIC insures the deposits of both national banks and state banks (and recently increased its coverage from $100,000 to $250,000 per deposit account) while the Fed has additional regulatory and supervisory authority for banking firms organized as bank holding companies (BHCs) or financial holding companies (FHCs).

This web of separate chartering, regulatory, supervisory, and deposit insurance institutions reflect both the federal structure of US government in which power is shared by the various states and the national government, and the pragmatic introduction of banking regulations to solve macroeconomic challenges as they occurred over time. The OCC was founded in 1863 to administer a new network of nationally chartered banks that issued a single, unified national currency backed by US Treasury securities, thus providing the federal government with funds necessary to fight the Civil War. The Fed was founded in 1913 to stabilize the economy during economic panics by providing a source of liquidity for commercial banks. The FDIC was established in 1933 to prevent bank runs by insuring the deposits of households and small businesses. These three federal institutions still exist and they continue to play their original bank regulatory and supervisory roles today. Apart from these basic fundamentals, however, virtually all other banking laws and bank regulations in the US have changed since the 1970s.

Restrictive government regulations

During the 1970s, and indeed during the entire post-war period leading up to the 1970s, US commercial banking was a protected industry. Government regulations shielded banks from geographic competition, from product competition, and to a great extent from price competition. The McFadden Act of 1927 protected banks from competitors outside their home states by prohibiting interstate branch banking. Although the Act did permit banks to enter other states by organizing multibank holding companies, these organizational structures required state approval which the states typically did not grant. In addition to these interstate restrictions, most states imposed partial or blanket restrictions on intrastate branching.

Product and pricing competition were also restricted by regulation. The Glass-Steagall Act of 1933 effectively isolated commercial banking as a separate and highly regulated financial sector and thus insulated commercial banks from competition with investment banks, insurance companies, and brokerage firms. Moreover,

depository institutions such as savings and loans and credit unions were not permitted to compete with banks by making commercial loans. And Federal Reserve Regulation Q imposed interest rate ceilings on most deposit accounts, effectively prohibiting price competition between banks for deposit accounts.

In this highly protected environment, the number of commercial banks in the US remained relatively unchanged throughout the 1960s, 1970s, and early 1980s, at about 14,000 federally or state-chartered commercial banks (see Figure 31.1). Over 95 percent of these commercial banks were so-called 'community banks' that held less than $1 billion of assets (2006 dollars); collectively, these small banks accounted for about one-third of the industry's total assets. The regulatory limitations on interstate banking and intrastate branching insulated community banks from large bank competition, and gave them a competitive advantage in lending and deposit taking at the local level. These advantages also extended to the payments' system, which in the US at that time was based largely on paper checks. A paper-based payments system requires not just that payors and payees have deposit accounts upon which to write checks and deposit checks, but also that depository institutions have safe and convenient physical locations for processing those checks. In a world before electronic payments infrastructure (e.g., automated teller machine, credit card networks, Internet banking), the physical brick-and-mortar infrastructures of community banks was fundamental in explaining their disproportionate presence in the industry.

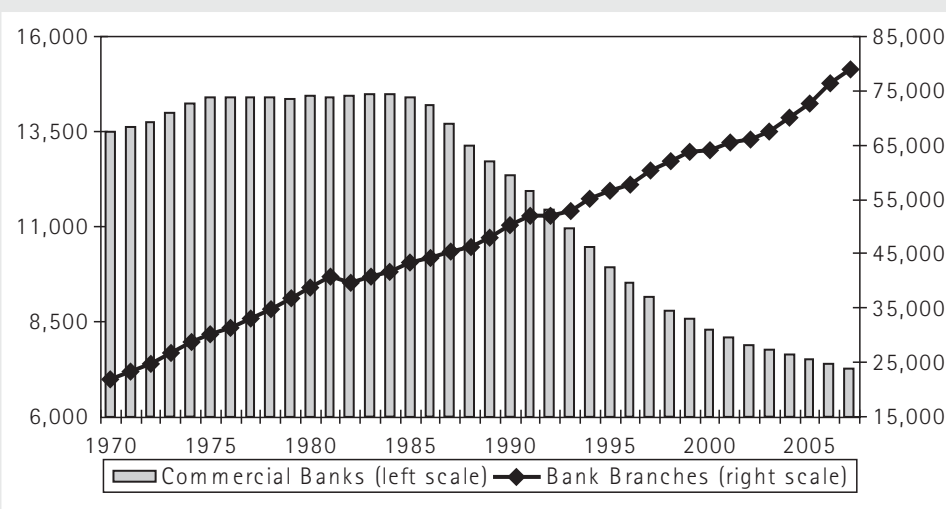

Fig. 31.1. Number of commercial banks and commercial bank branch offices in the US between 1970 and 2007

Source: Federal Deposit Insurance Corporation.

Because mutual funds were not yet well-established in the 1970s, banks were a leading investment vehicle for consumers, many of whom held their investments in savings accounts and time deposit accounts. Similarly, because the information technology necessary for modern mortgage banking had not yet emerged, banks combined with thrift institutions to dominate residential mortgage markets. In 1983 (the first year these data were available from the Federal Reserve's Survey of Consumer Finance), US households allocated approximately 23 percent of their assets to depository institutions, and obtained approximately 60 percent of their mortgage and consumer debt from depository institutions.

As the name implies, commercial banks were also the main supplier of loans to US businesses during the 1970s. Large commercial banks made loans to business firms of all sizes: they were the major source of both long-term and short-term financing to large businesses, and they made long-term loans to small businesses for purchasing fixed assets such as equipment and real estate (Carey, et al., 1993). Smaller community banks were a primary source of credit for small business enterprises, allocating between 20 and 30 percent of their loan portfolios to commercial lending during the 1970s (DeYoung, Hunter, and Udell, 2004).

Of the five basic financial needs of a typical household—credit, investments, transactions, safekeeping, and insurance—commercial banks were the dominant providers during the 1970s of all but insurance products (which they were prohibited from underwriting and severely limited in their abilities to market). Moreover, the prevailing technological and regulatory conditions allowed small banks to compete on equal footing with large banks in providing most of these services, especially in small and mid-sized cities and towns.

Innovation and technological change

A parade of financial and technological innovations during the 1970s and 1980s eroded the deposit-based funding advantages of US commercial banks, transformed their brick-and-mortar distribution networks, and reduced their traditional reliance on interest income. The first of these innovations was the money market mutual fund (MMMF), introduced in 1971. MMMFs transform large-denomination money market instruments (i.e., commercial paper, negotiable CDs, Treasury securities) into smaller-denomination investments affordable to the average household, allow investors (limited) check-writing privileges, and were not subject to Regulation Q. MMMFs grew dramatically in the late 1970s when the Federal Reserve's tight monetary policy pushed money market interest rates as much as 10 percentage points above the Regulation Q ceiling on deposit interest rates. Household funds flowed out of bank deposit accounts and into MMMFs, a process known as 'disintermediation'.

The automated teller machine (ATM), which was also introduced during the 1970s, had an equally powerful impact on retail banking. The ATM: (a) improved service quality by providing greater convenience for retail customers; (b) enhanced revenues by charging transactions fees to customers of other banks; and (c) increased the efficiency of bank branches by (as its name implies) serving as a substitute for more expensive human tellers. Indeed, the data suggest that the average banking office in the US has become more productive—assets, operating income, and the number transactions per banking office have all increased since the 1980s (DeYoung, Hunter, and Udell 2004)—which helps explain the large increase in the number of bank branches since the 1970s (see Figure 31.1).

In contrast to most developed economies, approximately two-thirds of US payments transactions were still conducted using checks and cash at the close of the twentieth century. However, electronic payments technologies—less expensive for banks to produce, and typically more convenient for their customers—are rapidly replacing paper-based payments in the US. The number of checks paid in the US was declining by about 3 percent annually during the late 1990s, while payments made with credit cards and debit cards were increasing by 7.3 percent and 35.6 percent per year, respectively (Gerdes and Walton, 2002; and Humphrey, 2002). Similarly, the volume of automated clearing house (ACH) transactions handled by the Federal Reserve—such as automatic payment of recurring monthly bills, and automatic deposit of wage and salary payments—increased at a 14.2 percent annual rate from 1990 to 2000 (Berger, 2003). Because the dispersal and receipt dates of electronic payments are more predictable than for check-based payments, US consumers now hold smaller precautionary balances: the fraction of household financial assets held in transactions accounts fell from 7.3 percent in 1983 to 4.6 percent in 2001 (Federal Reserve Survey of Consumer Finance, 2004).

Internet banking has further diminished the importance of geography and reduced the cost of producing the most basic banking services. The variable cost of producing a basic Internet banking transaction is very low, and as such there appear to be economies of scale associated with delivery channel—however, there is also some evidence that offering Internet banking services can also enhance the profitability of small banks (DeYoung, 2005; and DeYoung, Lang, and Nolle, 2007). Because US banks do not report detailed data on the throughput of their various delivery channels, a complete understanding has yet to develop. The predominant Internet banking strategy is the 'click-and-mortar' model that combines a transactional Internet site with networks of traditional brick-and-mortar offices and ATMs; no more than two dozen US banks offer their services exclusively over the Internet.

Among all of the financial innovations that have developed since the 1970s, securitized lending has perhaps left the biggest imprint on the structure and performance of the US banking industry. This lending technology—in which banks originate loans but do not finance them—has yielded large production

and financing efficiencies for banks that use it, and has increased access to credit for millions of households and small businesses. But the failings of this lending technology were one of several key causal factors for the disruptions in world financial markets during 2008 and 2009.

A loan securitization is a trust that purchases existing home mortgage loans (or auto loans, or credit card receivables) from banks, using funds raised by selling 'mortgage-backed securities' (MBS) to third-party investors. The MBS yield returns based on the performance of the mortgage loans held in the trust. This process allows banks to sell their otherwise illiquid loans to the securitization, and use the proceeds of these sales to fund additional loans or make other alternative investments. (A growing secondary market in the US for syndicated loans—loans made to large firms by 'syndicates' of large banks—has provided similar liquidity benefits and reduced the cost of loans to large firms. See Berlin, 2007.) Community banks have been able better to diversify their locally concentrated loan portfolios by purchasing MBS from securitizations of mortgages from other areas of the country. In contrast, many large retail banks have transformed themselves from traditional 'originate-and-hold' mortgage lenders to 'originate-and-securitize' mortgage bankers, relying less on traditional interest-based income and increasingly more on non-interest income from loan origination fees, loan securitization fees, and loan servicing fees. As discussed below, the scale economies associated with loan securitization have greatly influenced the industry's strategic profile.

The growth in securitized mortgage lending was facilitated in large part by two government-sponsored enterprises (GSEs). The Federal National Mortgage Association (Fannie Mae, founded in 1938) and the Federal Home Loan Mortgage Corporation (Freddie Mac, founded in 1970) are the dominant forces in US residential mortgage markets: approximately half of total existing residential mortgage debt in the US has either been securitized by, or is held in the portfolios of, these two institutions. These dominant market positions were achieved in large part as a result of the GSEs' lines of credit at the US Department of the Treasury—these credit lines created the perception in financial markets that Fannie Mae and Freddie Mac were 'too big to fail', which gave them a funding advantage over private sector mortgage securitizers. As outlined in detail by Frame and Wall (2002), the sheer size of these two companies raised fears about the systemic macroeconomic consequences should one or both ever become insolvent. Indeed, when Fannie Mae and Freddie Mac suffered losses in their huge portfolios of subprime MBS in 2008 and were on the verge of insolvency, the Treasury Department made good on the 'implicit government guarantee' by injecting equity funding and nationalizing ownership of the two GSEs.

Loan securitization rests on another financial innovation, credit scoring, which transforms quantitative information about individual borrowers (such as income, employment, or payment history) into a single numerical 'credit score'. Lenders use credit scores when analyzing loan applications; investment banks use credit scores

to construct pools of loans into which to be securitized; and bond rating companies use credit scores to assign risk ratings to asset-backed securities. First introduced in the 1950s, credit scoring is now widely used in consumer, mortgage, and micro small business lending (Mester, 1997). Although some (mostly larger) banks have developed their own credit scoring formulas, most lenders rely on standardized credit scores such as the 'FICO score'[1] acquired from third-party credit bureaus (e.g., Equifax, Experian, or TransUnion) to solicit and pre-screen loan applicants. Because credit scoring has significantly reduced the unit cost of underwriting individual loans, it has also increased the minimum efficient scale of consumer loan underwriting operations—hence, credit scoring has expanded lenders' incentives to make additional credit available (Berger, Frame, and Miller, 2005; and Frame, Srinivasan, and Woosley, 2001). During the mid-2000s, 'subprime' mortgage loans to households with low credit scores comprised a substantial portion of this credit expansion; many of these borrowers defaulted on their loans, causing large investment losses for US banks that held these mortgages as well as for other financial institutions that invested in securities backed by these mortgages.

Deregulation

By the 1980s, technological change and rapidly evolving conditions in financial markets had made the old regulatory regime untenable. Portions of the old regime were quickly dismantled. The disintermediation of household savings out of bank deposits and into higher yielding MMMFs and other non-bank investments forced the Federal Reserve to loosen and eventually remove almost entirely the interest rate restrictions imposed by Regulation Q. The Garn-St Germain Depository Institutions Act of 1982 authorized banks and thrifts to offer money market deposit accounts (MMDAs)—transaction accounts with no interest rate ceiling—which allowed them to compete directly with MMMFs. The Act also permitted thrift institutions to make commercial loans and thus compete more directly with community banks.

Other parts of the old regime took longer to fall away. Between 1980 and 1994, thirty-two states gradually liberalized geographic restrictions on banking and branching within state borders. Various states also circumvented the federal McFadden Act by entering into bilateral and multilateral agreements that allowed cross-border bank ownership through multi-bank holding companies; by the end of the decade, all but six states allowed some sort of interstate banking. In 1987, the Federal Reserve allowed commercial bank holding companies to operate 'Section 20' subsidiaries to underwrite corporate securities in limited amounts, and in 1989

[1] FICO is a registered trademark of Fair Isaac Corporation.

began relaxing restrictions in the Glass-Steagall Act that had banned commercial banks from themselves underwriting corporate securities.

The 1990s witnessed two crowning deregulatory acts. In 1994, the US Congress passed the Riegle-Neal Interstate Banking and Branching Efficiency Act, which effectively repealed the McFadden Act at the national level and harmonized the patchwork of state-by-state banking and branching rules. This Act limits interstate expansion of commercial banking companies only to the extent that they are prohibited from acquiring other commercial banks should their national deposit market share exceed 10 percent. (Bank of America is the only US banking company currently constrained by this law.) And, in 1999, Congress passed the Gramm-Leach-Bliley Financial Services Modernization Act, which effectively repealed the Glass-Steagall Act by granting broad-based securities and insurance powers to commercial banking companies. These deregulatory acts ratified the decades-long deregulation movement that began in the 1970s, and helped accelerate the adoption of new financial processes and information technologies by US banks.

The most noticeable industry response to deregulation was a historic wave of commercial bank mergers and acquisitions. Not including acquisitions arranged by the FDIC to resolve failing institutions, there were approximately 3,500 bank mergers during the 1980s, a further 5,000 during the 1990s, and over 2,000 more between 2000 and 2006. These combinations increased the size and geographic footprints of US commercial banks of all sizes, but were especially important for creating large, multi-state banking companies for the first time in the history of the US. While enormous in scope by any measure, this consolidation was slow to produce truly nationwide retail banking franchises. Predicting the speed and extent of this industry consolidation has been difficult: Berger, Kashyap, and Scalise (1995) made a comprehensive study of industry consolidation during the middle of the merger wave, and estimated a future path for consolidation that overshot the mark *ex post*; armed with the advantage of a decade of additional information, Jones (2006) suggested that consolidation was still in process but may finally be showing some signs of slowing down.

Large multi-state banks have been quicker than smaller community banks to adopt new financial and information technologies, including various forms of electronic payments, credit scoring and loan securitization, financial derivatives, and other off-balance sheet activities. The more scalable of these technologies have also disseminated rapidly at smaller banks, albeit with a few years lag, due to the declining costs of delivering these technologies and a highly competitive sector of third-party technology vendors (Frame and White, 2004). For example, imaging technology allows banks of all sizes to transmit checks as electronic images, saving the substantial transportation and handling expenses associated with paper checks; the Check Clearing for the Twenty-first Century Act of 2003 (Check 21) facilitated these efficiencies by recognizing an electronic image as a legal substitute for a paper check.

A STYLIZED VIEW OF BANKING STRATEGIES

Technological change and industry deregulation left US commercial banks at a strategic crossroads. DeYoung, Hunter, and Udell (2004) provide a simple but powerful model of the new strategic landscape, how banks responded, and the new industry equilibrium that resulted. Bank size, and the scale economies that can result from increased bank size, are key to the analysis. Research suggests that banks practicing traditional 'originate-and-hold' banking can operate without substantial unit cost disadvantages once they have about $500 million in assets, which is still very small by today's standards (DeYoung and Rice, 2004a). However, small banks attempting to practice 'originate-and-securitize' banking will face enormous cost disadvantages (Hughes, et al., 1996; and Rossi, 1998), chiefly because of scale economies associated with the collection and analysis of the 'hard', quantifiable borrower information essential to the asset securitization process (Stein, 2002). There is evidence that large banks and small banks have comparative advantages in lending to large firms and small firms, respectively (Berger, et al., 2005). The large-scale, impersonal retail delivery channels favored by large banks (e.g., ATM networks, electronic payments) also thrive on hard information. Because their rival banks usually have access to the same information, large banks face intense price competition for the non-differentiated financial products they sell (e.g., credit cards, mortgage loans, transactions services).

In contrast, small banks augment their stores of hard information with 'soft', non-quantifiable information collected over time via personal interaction with their borrowers, depositors, and the local community (Scott, 2004). This 'relationship-based' approach to banking allows small banks to serve local businesses that are unable to access public capital markets and households who require in-person financial services. Because these small banks differentiate themselves from their larger rivals by offering personalized products and services (e.g., small business loans, financial planning), and because they have a store of customer information to which their small local competitors do not have access, they can charge higher prices.

The key implication of the model is that the banking industry will naturally settle into a dichotomous structural equilibrium in which both large banks and small banks are profitable. Large bank size is a prerequisite for attaining the operating scale necessary profitably to exploit the technological and financial advantages of the transactions banking business model, just as relatively small size is a prerequisite for maintaining the local focus necessary for profitable community banking and relationship lending. The 'financial commodity' strategy practiced by large banks sacrifices personalized service and high prices in exchange for high sales volume, standardized products, and low unit costs. The 'relationship banking' strategy practiced by small banks sacrifices sales volume and lower unit

costs in exchange for local economic focus, differentiated products, and high prices. Banks of middling size will be unable to take profitable advantage of either strategy and will exit the industry, probably via merger and acquisition.

There is considerable empirical evidence consistent with this strategic framework. Table 31.2 compares various financial ratio averages for two groups of US commercial banks in 2006. The 'small bank' group in the first column is comprised of 434 banks each having between $500 million and $2 billion in assets. These banks are arguably large enough to capture enough scale economies to operate at a reasonable level of efficiency, but still small enough to practice traditional relationship-based community banking. The 'large bank' group in the second column is comprised of fifty-six banks each having at least $10 billion in assets. These banks far exceed the upper size limit attributed to community banks, and are

Table 31.2. Selected financial ratios (mean values) for 490 US commercial banks in 2006

	'small' banks	'large' banks
Number of banks	434	56
Asset range	$500 million to $2 billion	over $10 billion
Small business loans (% of loans)	8.55	4.46
Securitized loans (% of consumer loans)	0.14	15.59
Core deposits (% of assets)	63.07	50.70
Purchased federal funds (% of assets)	3.01	7.85
Net interest income (% of assets)	3.63	2.82
Interest income (% of assets)	6.18	5.33
Interest expense (% of assets)	2.54	2.51
Standby financial letters of credit (% of assets)	0.59	3.98
Non-interest income (% of operating income)	20.28	38.66
Deposit service charges (% of operating income)	7.94	8.39
Fiduciary income (% of operating income)	1.71	8.58
Trading income (% of operating income)	0.02	2.03
Investment banking income (% of operating income)	0.60	2.02
Insurance income (% of operating income)	0.73	1.29
Loan servicing income (% of operating income)	0.01	0.88
Fees from mutual fund sales (% of operating income)	0.54	1.16
Other non-interest income (% of operating income)	8.19	13.15
Return on assets	1.12	1.23
Return on equity	12.81	13.70

* Each of the banks in this analysis operated with either a state or a federal commercial banking charter. If a bank was affiliated with a bank holding company, it was included only if it was the largest bank (i.e., the 'lead bank') in their organization. Banks less than ten years old were excluded to insure that all banks in the analysis were financially mature (DeYoung and Hasan, 1998). Banks investing more than 10 percent of their assets in either agricultural loans or credit card loans were also excluded, as these banks tend to face idiosyncratic market conditions and/or use more specialized production functions.

Source: Federal Deposit Insurance Corporation.

large enough to pursue a 'financial commodity' production process and business strategy.

These data suggest a fundamentally different approach to lending for the two groups of banks. The large banks are more likely to practice large-scale 'originate-and-securitize' approaches to consumer lending and less likely to engage in relationship-based small business lending, while the small banks are more likely to 'originate-and-hold' consumer loans and to engage in relationship-based small business lending. For example, the average large bank sold off and securitized roughly a dollar of consumer (automobile, home mortgage, home equity, or credit card) loans for every eight dollars of consumer loans on its balance sheet (15.61 per cent), while the average small bank held nearly all of its consumer loans as portfolio investments (0.14 percent). These 'consumer-loan turnover' ratios are averages, and they belie an even starker comparison not included in the table: this ratio ranged as high as 200 percent for the large banks, but no higher than 33 percent for the smaller banks. In contrast, the average large bank invested only half as many of its assets (4.46 percent) in small business loans compared to the average small bank (8.55 percent). Moreover, this comparison probably understates the relationship lending gap between large and small banks, because a substantial portion of the small business loans made by large banks are underwritten based on the personal credit score of the proprietor and hence are more like hard-information-based credit card loans than soft-information-based relationship loans.

The difference in the funding sources used by these banks provides further evidence of two different strategic approaches to banking. The typical large bank funded nearly one dollar of every twelve dollars of assets with funds that it purchased overnight from other banks (7.85 percent), while the typical small bank relied less than half as much (3.01 percent) on purchased short-run financing. In contrast, the average small bank funded over 60 percent of its deposits using 'core deposits'—that is, stable deposit balances with long durations, including transactions deposits, small savings deposits, and certificates of deposit less than $100,000—compared to only about 50 percent for the average large bank. For the small banks, these data are more consistent with a traditional banking approach in which stable deposits fund loan portfolios, with profitable bank–customer relationships being forged on both sides of the balance sheet. For the large banks, these data are more consistent with a transactional approach to banking in which standardized loans are originated and sold, thus requiring a more flexible and shorter duration mix of funding.

These differences in lending and funding strategies are reflected in the interest margins earned by the two sets of banks. The net interest margin averaged only 2.82 percent for the large banks, substantially less than the 3.63 percent average for the smaller banks—and note that this 81 basis point difference is driven completely by higher interest income for the smaller banks, not by any small bank advantage in

interest expense. This remarkable finding is consistent with the differences in returns to hard-information lending and soft-information lending, as suggested above. The retail consumer and mortgage loans in which many large banks specialize are high-volume, financial commodities sold in highly competitive markets, resulting in downward pressure on loan interest rates. In contrast, the relationship-based small business loans in which many small banks specialize are low-volume, idiosyncratic credits made to informationally opaque borrowers in less competitive markets, which allows for higher lending margins.

While interest margins tend to be lower for the large banks, non-interest income tends to be higher. On average, non-interest income accounts for nearly 40 percent of operating income (net interest income plus non-interest income) at the large banks, or roughly twice the amount generated by the small banks. The large banks earn greater amounts of fee income both from traditional banking services—such as deposit account fees, fees for providing fiduciary services (e.g., managing trusts and investment accounts), and gains from trading bank-eligible investment securities—as well as from non-traditional banking services such as investment banking and insurance activities. And the large banks nearly double the earnings of small banks in the catch-all 'other non-interest income' category, which includes fees earned from providing loan commitments and letters of credit.

In terms of overall profitability, large banks' deficiencies in net interest income are more than offset by their ability to generate large amounts of non-interest income. On average, compared to the small banks the large banks earn both higher returns of assets (1.23 percent versus 1.12 percent) and higher returns on equity (13.70 percent versus 12.81 percent). These returns are not adjusted for risk, however, and it may be that the large banks need to generate higher-equity returns to reward their owners for higher risk. As discussed in detail below, there is growing evidence that fee-based banking activities generate riskier earnings than do margin-based banking activities.

To be sure, this highly stylized analysis oversimplifies the array of strategic choices available to commercial banks. Many large banks offer customized, relationship-based services to clients with idiosyncratic financial needs, such as corporate investment banking clients and high-net-worth 'private banking' customers. Similarly, most small banks have relied to some extent on hard information, such as pledged collateral and audited financial statements, to underwrite business loans. Thus, there is ground between the two polar strategies proposed here, and some (though clearly not all) of that strategic landscape is profitable. Large retail banks attempt to access that ground via marketing: while these banks encourage their retail customers to purchase standardized deposit and loan products (e.g., automated credit-scored loans) through impersonal banking channels (e.g., online banking), they also attempt to differentiate these products and services from those of their competitors with image-based advertising campaigns. As shown in Figure 31.2, the largest US banking companies spend proportionately

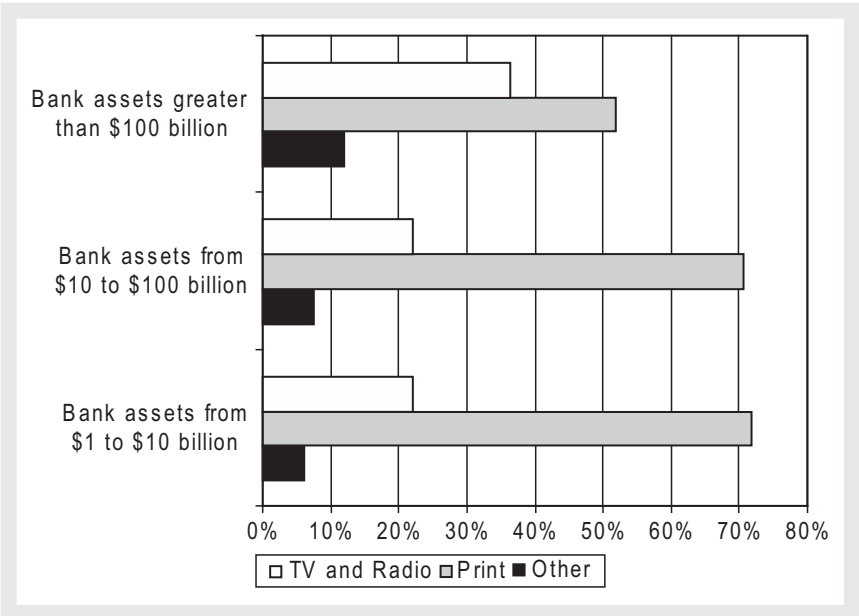

Fig. 31.2. Average advertising expenditures for fifty-one commercial banking companies in the US in 2006, expressed as a percentage of companies' total spending on advertising

Note: * The figure includes the largest 51 US banking companies for which data was available.

Source: *American Banker*.

more on television and commercials—media well-suited for delivering qualitative, image-based messages—and proportionately less on print advertising, which is better suited for delivering quantitative information about actual product characteristics.

There is very little research on the use and effects of advertising in US banking markets—largely because US regulators only recently required commercial banks to report even crude data on marketing expenditures. Berger and Dick (2007) have shown that banks with strong 'brand images' (defined as multi-market presence, presumably bolstered by large marketing expenditures) are able more quickly to expand the banks that they acquire. Some researchers have used marketing data from US thrift institutions: Hasan, Hunter, and Mathis (2000) found that thrifts facing more intense competition spend more on advertising, while the empirical tests performed by DeYoung and Ors (2004) suggest that thrifts use advertising both to communicate high interest rates offered on certificates of deposit as well as create brand images that allow them to pay lower interest rates on checking accounts.

Industry structure

As discussed above, geographic and product market regulations constrained the growth of US banking companies for many decades—historically, the top US banking companies have been small relative to the largest Japanese and European banks. When deregulation released these constraints, US commercial banks grew rapidly, chiefly by acquiring other US banks. On average, about 350 commercial banks were acquired each year during the 1980s, about 500 each year during the 1990s, and about 300 each year during the first half of the 2000s—in all, over 10,000 bank charters were merged out of existence since 1980 (see Figure 31.3). These acquisitions have substantially altered the structure of the US banking industry. The number of commercial banking charters in the US has dropped by almost half, from a plateau of approximately 14,000 banks that had remained remarkably stable since the 1950s, to a low of about 7,500 banks by 2005 (and still falling).

A wave of bank failures contributed further to the decline in the number of US commercial banks. Regulators shut down over 1,500 insolvent banks during the late 1980s and early 1990s, the largest number of bank failures in the US since the Great Depression. There were two primary causes of these insolvencies: an unexpected increase in interest rates that destroyed the profitability of banks that had

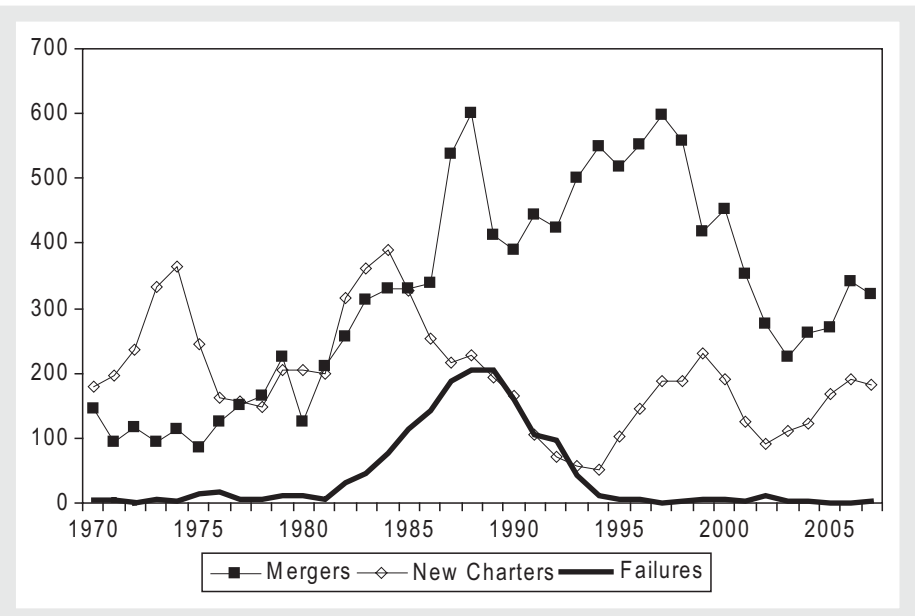

Fig. 31.3. Changes in the number of commercial bank charters in the US from 1970 to 2005 due to mergers, failures, and new entry

Source: Federal Deposit Insurance Corporation.

financed long-term fixed rate loans with short-term deposits, and sustained regional declines in real estate values, largely in the southwest oil-producing states and in the New England states, that destroyed the loan quality of local banks with non-diversified portfolios of real estate loans. (These same phenomena resulted in even more dramatic failures of US thrift institutions; for example, the overall population of federally chartered savings institutions declined by one-third, from about 3,600 to 2,400 institutions, between 1986 and 1992.)

As will be discussed below, changes in government supervision and industry risk mitigation practices were made during the 1990s and 2000s in hopes of reducing the chance of future bank failure waves. These improvements notwithstanding, there were twenty-five commercial bank failures in the US in 2008—the most in over a decade—and at year end an additional seventy-six banks were on the FDIC's list of financially troubled banks. While the 1980s–1990s bank failure wave was caused by unwise exposures to interest rate risk and geographic loan concentrations, these more recent bank failures reflect substantial investments in (geographically diversified) mortgage-backed securities coupled with a nationwide downturn in housing markets. Policy actions taken by the US Treasury (temporary capital injections) and the Federal Reserve (making short-term liquidity available) probably reduced the number of banks that would have otherwise failed during 2008 and going forward.

While bank mergers and bank failures were reducing the number of US commercial banks, over 7,000 new banking charters were granted by state and federal banking authorities between 1970 and 2005. New bank start-ups are rare outside of the US, and this large volume of 'de novo' banks in the US is made possible by the competition between the federal banking authority (the OCC) and the fifty separate state banking authorities, all of which can grant banking charters. The surge in new bank charters simultaneous with the merger-driven industry consolidation was no coincidence: when large, out-of-state banking companies acquire small, locally focused banks, some portion of the acquired bank depositors, borrowers, and employees will inevitably be unhappy with changes in post-acquisition policies and will want to change banks. Combining these three essential banking inputs—deposits, loans, and skilled banking employees—with a relatively small amount of investment capital (in most cases, US banking authorities require less than $20 million in start-up capital) is a simple recipe for a new bank. Indeed, studies have shown that new banks are more likely to start up in local markets immediately after established banks are acquired in mergers (Keeton, 2000; and Berger, et al., 2004). Many of these new banks have grown rapidly and have become very profitable, demonstrating that strong customer demand exists for small, locally focused banks.

As the number of US commercial banks declined, the size distribution of banks also changed. As shown in Figure 31.4, the net reduction in the number of banks occurred wholly among banks with less than $500 million of assets (2006 dollars),

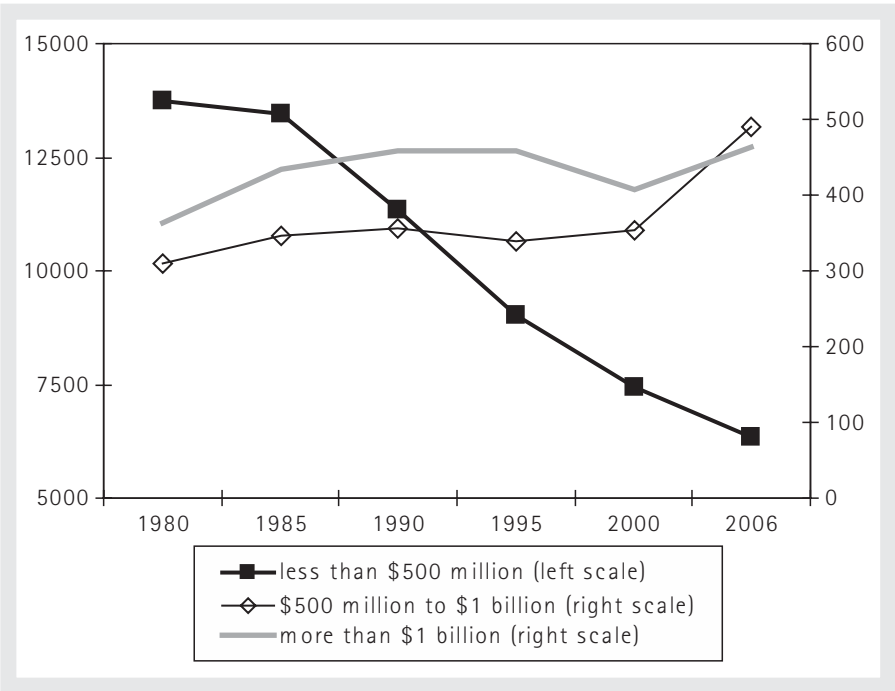

Fig. 31.4. The changing distribution of US commercial banks by size (measured in 2006 dollars) between 1980 and 2006

Source: Federal Deposit Insurance Corporation.

from nearly 14,000 banks in 1980 to a little over 6,000 banks today. The majority of banks that failed and banks that were acquired since 1980 were in this size group, while other small banks grew up and out of this group by acquiring other small banks. In sharp contrast, the number of banks with more than $1 billion in assets has remained relatively stable at between 300 and 500 since 1980, as has the number of banks with between $500 million and $1 billion in assets. An implication of these data is that banks can capture meaningful scale economies by growing up to $500 million in assets, but that growing beyond $500 million yields less substantial gains. Consistent with this implication, DeYoung and Rice (2004a) found that increases in the size of US commercial banks up to about $500 million unambiguously improved the risk–return trade-off—that is, expected returns increased while the variability of these returns declined—while increases in bank size beyond $500 million are associated with the choice of a less traditional business strategy that yielded increased returns but also increased risk.

These findings fit well with the large literature on scale economies at US commercial banks (e.g., Mester, 1987; Clark, 1988; Evanoff and Israilevich, 1991; and Berger and Mester, 1997). Studies that used banking data from the 1970s and

1980s, when banks of all sizes were using traditional banking models, typically found that scale economies were either fully or substantially exhausted by relatively small banks, with minimum efficient scale estimated to be substantially less than $1 billion of assets. But, studies that used data from the 1990s and early 2000s yielded different insights, and often concluded that additional scale economies exist for large regional banks and perhaps even for nationwide banks. While changes in estimation methodologies may be responsible for some of the differences in these two sets of studies, the change in banking production technologies and the proliferation of different banking strategies over time are arguably the more important developments. Rossi (1998) showed that economies of scale exist for even the largest mortgage banking companies, which employ a very basic transactions approach to banking. Hughes, et al. (1996) concluded that even the largest commercial bank holding companies—where sales volume is often dominated by transactions banking activities—also exhibit increasing returns to scale. And DeYoung (2005) argues that Internet-only banks—again, a pure transactions banking strategy—exhibit larger scale economies than do similar-sized banks that use a branch delivery system.

As shown in Table 31.3, the speed with which the scale of large US banks has increased has been staggering. During the mid-1980s, only the largest US commercial banking company (Citibank) had more than $100 billion in assets, but by the mid-2000s nearly twenty US banking companies had more than $100 billion, and three exceeded $1 trillion. Numerous storied US banking franchises (e.g., Chemical Bank, Manufacturers Hanover, Bankers Trust) disappeared over the past two decades as part of so-called 'mega-mergers' that fueled this rapid expansion. But, most of the merger-related growth has been geographic expansion, as banks in one city, state, or region took advantage of deregulation by acquiring banks in other cities, states, or regions. This has had very little effect on the structure of local banking markets—the ownership of a local bank changes when it is acquired by a bank from outside its local market, but its local market share is left unaffected—however, the nature of competitive rivalry in local markets has been affected. The cost efficiency of local banks tends to improve after one of their local peers is acquired by a large out-of-market bank, presumably because of increased competitive pressure (DeYoung, Hasan, and Kirchhoff 1998; and Evanoff and Ors, forthcoming). Part of this pressure comes from large multi-market banks' more intensive use of ATMs, online banking, credit scoring, and other information technologies to provide more convenience for retail customers (Berger, et al., 2007). And as mentioned above, banks with strong 'brand images' are better able to expand the local market shares of the banks they acquire (Berger and Dick, 2007), suggesting that perceived (as opposed to actual) differentiation can be an effective tool for large banks selling financial commodity products.

By definition, geographic expansion increases the distances within banking organizations, and this can create challenges for bank management. Berger and

Table 31.3. Ten largest US commercial banks in 1988, 1997, and 2007

	June 1988		December 1997		June 2007	
1	Citicorp	$194,600	Chase Manhattan Corp.	$365,531	Citigroup Inc.	$2,220,866
2	Chase Manhattan Corp.	$98,860	Citicorp	$262,159	Bank of America Corp.	$1,535,684
3	BankAmerica	$96,923	NationsBank Corp.	$260,159	JPMorgan Chase & Co.	$1,458,042
4	Chemical Banking Company	$78,410	JPMorgan & Co.	$157,274	Wachovia Corp.	$719,922
5	JPMorgan & Co.	$74,681	Bankamerica Corp.	$140,102	Deutsche Bank	$579,062
6	Maunfacturers Hanover Corp	$73,826	First Union Corp.	$116,182	Wells Fargo & Co.	$539,865
7	Security Pacific Corp.	$64,714	Bankers Trust New York Corp.	$140,102	Washington Mutual Inc.	$349,140
8	Bankers Trust New York Corp.	$54,700	Banc One Corp.	$116,182	U.S. Bancorp	$222,530
9	First Interstate Bancorp	$51,790	First Chicago NBD Corp.	$114,096	SunTrust Banks Inc.	$180,314
10	Wells Fargo & Co.	$44,721	Wells Fargo & Co.	$97,456	National City Corp. Cleveland	$140,648

Source: American Banker.

DeYoung (2001; 2006) found that the operational efficiency of bank holding company affiliates declined as they were located further away from their headquarter banks. While advances in communications and information technologies have helped mitigate these long-distance management problems, the very existence of these inefficiencies points to a competitive advantage for small, locally focused banks. Distances between banks and their loan clientele have also increased over time. This phenomenon is mainly technology-driven: automated, credit-scored lending models allow banks to make consumer, mortgage, credit card, and even some small business loans to borrowers they have never met in person, and asset securitization and credit derivatives allow banks to manage the risk associated with this type of lending (Petersen and Rajan, 2002; and DeYoung, Glennon, and Nigro 2008).

The geographic expansion of US banking companies has not been limited to domestic markets. As shown in Table 31.4, six of the largest fifty banking companies in the world in 2006 were US banking companies: Citigroup (ranked fourth in terms of assets), Bank of America (tenth), JPMorgan Chase (eleventh), Wachovia (twenty-seventh), Wells Fargo (fortieth), and Washington Mutual (fiftieth). This marks an important change from the past, when, despite operating in the world's largest economy, US banking companies were constrained from expanding outside their home states and were largely limited to providing commercial banking services, and as a result were smaller than the top universal banking companies from Japan, Germany, Switzerland, the Netherlands, and the UK. Through rapid geographic growth and expansion into investment banking services, US commercial banking companies now rank among the world's largest in terms of syndicated lending, debt underwriting, and equities underwriting, as displayed in Table 31.5.

Table 31.4. Largest banking companies in the world in December 2006 (dollars of assets)

1	UBS AG Zurich	$1,961,327
2	Barclays PLC London	$1,949,167
3	BNP Paribas Paris	$1,898,186
4	Citigroup Inc. New York, N.Y.	$1,884,318
5	HSBC Holdings PLC London	$1,857,520
6	Royal Bank of Scotland Group PLC Edinburgh	$1,705,044
7	Crédit Agricole SA Paris	$1,662,600
8	Mitsubishi UFJ Financial Group Tokyo	$1,585,767
9	Deutsche Bank AG Frankfurt	$1,480,984
10	Bank of America Corp. Charlotte	$1,459,737
11	JPMorgan Chase & Co. New York, N.Y.	$1,351,520
27	Wachovia Corp. Charlotte	$707,121
40	Wells Fargo & Co. San Francisco	$481,996
50	Washington Mutual Inc. Seattle	$346,288

Source: *American Banker.*

Table 31.5. Ten largest debt underwriters, equity underwriters, and loan syndicators in the world in 2007 (by proceeds in billions of dollars)

	Global Debt Underwriting		Global Equity Underwriting		Syndicated Loans	
1	*Citi*	$546	UBS	$81	*JPMorgan*	$570
2	*JPMorgan*	$476	*JPMorgan*	$77	*Citi*	$522
3	Deutsche Bank AG	$429	*Citi*	$71	*Banc of America Securities LLC*	$337
4	Merrill Lynch	$370	Goldman Sachs & Co	$70	Royal Bank of Scotland	$210
5	Lehman Brothers	$365	Morgan Stanley	$64	Deutsche Bank AG	$180
6	Morgan Stanley	$361	Merrill Lynch	$60	Barclays Capital	$173
7	Barclays Capital	$349	Credit Suisse	$54	BNP Paribas SA	$172
8	Goldman Sachs & Co	$286	Deutsche Bank AG	$52	Goldman Sachs & Co	$134
9	Royal Bank of Scotland	$282	Lehman Brothers	$29	Calyon	$118
10	*Banc of America Securities LLC*	$276	China International Capital Co	$20	Credit Suisse	$116

Note: US companies are in italics.
Source: American Banker.

Industry consolidation and geographic expansion have also altered the nature of bank delivery systems. For example, while the number of commercial banks has declined by half since 1980, the number of commercial bank branches has nearly doubled from about 38,000 to more than 75,000 today (see Figure 31.1). Although some of this reflects bank purchases of thrift institutions with branching networks, the explosion in bank branches has been largely strategic in nature. By 'packing the map' with branches, a bank positions itself closer to its current customers, closer to its potential (i.e., its rivals') customers, establishes a market presence at relatively low cost, and if it does so successfully may be able to limit entry by its rivals. This strategy can be especially important for large, transactions-based banks; while it is difficult for these banks to offer personalized banking services, they can offer higher levels of customer convenience by deploying a combination of local branches, automated teller machines (ATMs), and Internet banking. A series of Federal Reserve surveys in the late 1990s and early 2000s documented that retail customers pay higher deposit-related fees at large banks, a likely indication that consumers are willing to pay for this higher level of convenience (Hannan, 2002). Physical branches located in prominent places can also serve as important advertising vehicles, especially in markets into which banks have recently expanded.

The implications of increased non-interest income

As documented in Table 31.2, large commercial banks generate a greater portion of their incomes from non-interest activities than to small banks. Thus, as the US banking industry consolidated into fewer and larger banks during the 1980s and 1990s, one would expect a greater portion of industry income to have flowed from fees rather than interest. Indeed, as shown in Figure 31.5, non-interest income roughly doubled as a percentage of total commercial bank income between 1980 and 2000, when bank merger activity was peaking in the US.

But, the increasing size of US banks, and with it the movement of the largest banks toward fee-based, transactions-based banking strategies, is not the only reason for the burgeoning importance of non-interest income. First, deregulation allowed commercial banking companies to expand into non-traditional lines of businesses—such as securities underwriting, securities brokerage, and insurance sales—that generate non-interest income. This came in stages, with the Federal Reserve's gradual relaxation of income limits at Section 20 securities subsidiaries during the 1990s, expanded insurance powers granted to national banks by the OCC during the late 1990s, and the passage of the Gramm-Leach-Bliley Act in 1999.

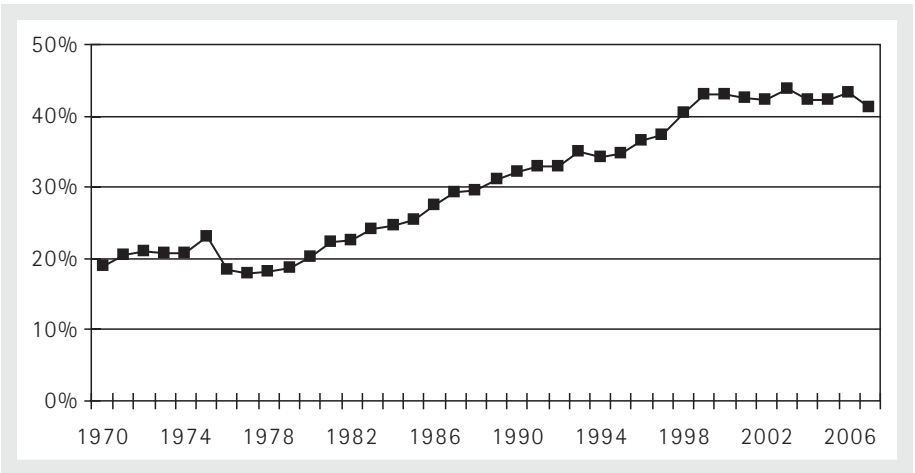

Fig. 31.5. Aggregate non-interest income as a percentage of aggregate operating income of US commercial banks, 1970–2007

Note: * Operating income is equal to net interest income plus non-interest income.
Source: Federal Deposit Insurance Corporation.

Second, while in the past banks would earn interest income by providing credit to their business clientele, today banks increasingly earn fee income by selling backup lines of credit that enable their business clients to issue their own debt securities. Similarly, the massive shift from portfolio lending to securitized lending has transformed consumer lending from an interest-generating to a fee-generating line of business for many banks. Third, in the past, banks were constrained by Regulation Q from paying market interest rates on transactions deposits, and banks made up for this by providing depositor services (e.g., certified checks, safe deposit boxes, overdraft protection) either free of charge or at prices that were well below costs. The relaxation of Regulation Q has resulted in market pricing for both deposit interest rates and depositor services, and thus increased fee income for banks.

DeYoung and Roland (2001) have argued that the increase in non-interest income at US banks has altered fundamentally their risk–return profiles. For example, compare (a) the fee income that a bank receives by originating and then securitizing a mortgage loan to (b) the interest income that a bank receives by making a small business loan and holding it in its loan portfolio. The former is a non-repeat transaction with the borrower, and the fees generated by this line of business are sensitive to the volatility of both the housing market and mortgage interest rates; the latter is a long-term relationship that both sides have an interest in preserving, which will continue to generate interest income (and perhaps fee income as well) into the future. Similarly, the fees associated with securities brokerage are typically based on the value of assets sold or assets under

management, so that the stream of fee income generated by these activities contains systematic (undiversifiable) risk from market fluctuations. Moreover, the production functions for fee-based activities are typically dominated by fixed costs (e.g., personnel expenses), while lending activities are based primarily on variable costs (e.g., interest expenses); thus, non-interest income requires high operating leverage (i.e., a high fixed-to-variable cost ratio) which amplifies any revenue volatility into even greater earnings volatility. The authors go on to show that (non-deposit-related) fee income is associated with higher revenue volatility, higher operating leverage, and higher earnings volatility at US commercial banks.

Several other empirical studies have investigated the riskiness of non-interest income. DeYoung and Rice (2004b) found that marginal increases in non-interest income are associated with a worsening of banks' risk–return trade-off. Stiroh (2004a; 2004b) found no evidence of diversification gains at banks that combine interest and non-interest income. Clark, et al. (2007) emphasize how the increasingly retail-focused strategies of large US banking companies expose these banks to economic and business cycle volatility.

The subprime mortgage crisis—which started in 2007 and lasted for several years afterwards—provides an illustration of the income volatility associated with fee-driven transactions banking. While the headlines in the financial press dwelled almost exclusively on the large capital losses suffered by banks and other investors in subprime mortgage-backed securities, banks that originated, serviced, and/or securitized mortgages experienced material, and in some cases crippling, reductions in fee income as investor demand for new MBS dried up and household demand for both new and existing houses declined. Total industry non-interest income fell from 43 percent of operating income in 2006 to just 38 percent during the first three quarters of 2008 (not shown in Figure 31.5), the largest two-year decline since the mid-1970s. Many of the largest financial institutions with non-diversified, 'mono-line' mortgage banking strategies failed (e.g., American Home Mortgage, New Century Financial, Countrywide Financial, Washington Mutual) owing to the combined impact of plummeting fee income and large losses in their portfolios of subprime mortgages and mortgage-backed securities.

Risk and return

Despite the fact that growing reliance on non-interest income was increasing the volatility of income and earnings at US commercial banks, by the mid-2000s the industry appeared to be very well positioned to handle this risk. Figure 31.6 shows the annual aggregate (book value) equity-to-assets and return-on-assets ratios for

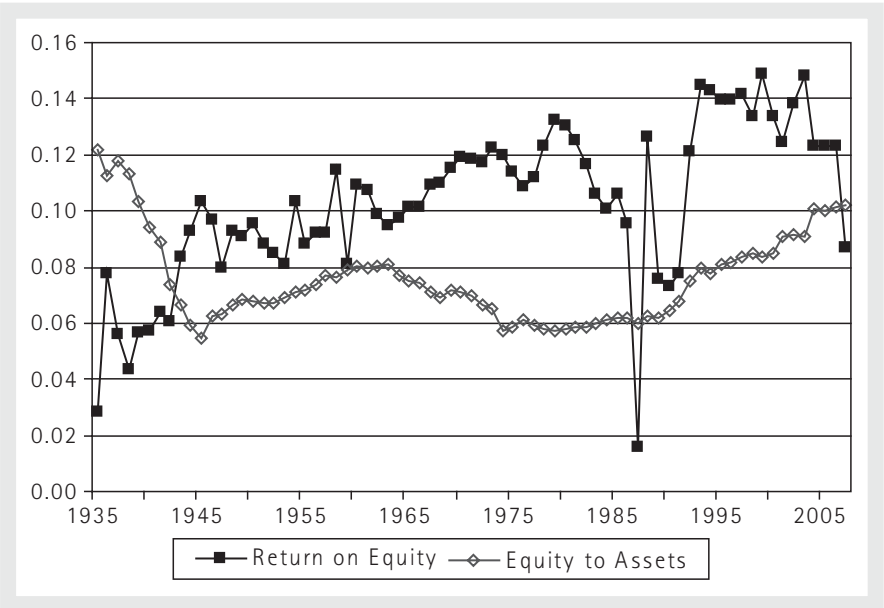

Fig. 31.6. Aggregate return-on-equity and equity-to-assets ratios for the US commercial banking industry, 1935–2007

Source: Federal Deposit Insurance Corporation.

US commercial banks going back seven decades. Capitalization in the banking industry has been steadily increasing since the early 1990s; by 2004, commercial banks were holding a dollar of equity capital against every ten dollars of assets, nearly double the capital levels of the early to mid-1980s. This large reservoir of capital provided an increased margin of safety and soundness against the increased opportunities for risk taking in today's deregulated and intensely competitive commercial banking industry.

This large capital cushion was the result of several independent developments. First and foremost was the stricter supervisory and regulatory framework mandated by the Federal Deposit Insurance Corporation Improvement Act (FDICIA) of 1991, the centerpiece of which is the practice of 'prompt corrective action' by bank supervisors that imposes costly restrictions on banks with low and diminishing levels of capital (Eisenbeis and Wall, 2002). Equally important were the historically high levels of bank earnings during the late 1990s and early 2000s (see ROE trend in Figure 31.6), which were caused by a combination of phenomena: two decades of strong macroeconomic growth, the elimination of costly regulatory constraints on banks, and increasing competitive pressures for banks to operate more efficiently. Berger, et al. (2008) show that banks retained a large percentage of these record earnings rather than distributing them to shareholders, a financial strategy that helps account for the simultaneous increase in bank capital levels.

The Berger, et al. (2008) study also estimates the desired or 'target' capital ratios of large US banking companies, and finds that during the 1990s and 2000s banks desired to hold equity capital well in excess of minimum capital levels set by their regulators. These regulatory capital requirements continue to evolve. The international Basel Capital Accord (Basel II), initially published in 2004, but not yet fully implemented in the US, allows the largest banking companies to determine how much capital they need to hold, based on their own internal credit risk estimates and value-at-risk modeling techniques (Gordy and Heitfield, 2008). It remains to be seen how Basel II will affect total capital levels in the US and the allocation of capital to risk taking at US banks. Because of these uncertainties, US bank regulators will include a simple equity-to-assets (or 'leverage') ratio as a backstop against modeling outcomes that recommend only low levels of capital. Consistent with the spirit of Basel II, the Federal Deposit Insurance Reform Act of 2006 allowed the FDIC much greater latitude to charge risk-based deposit insurance premiums that better reflected banks' estimated insolvency risk.

While holding a larger cushion of equity capital is the most fundamental hedge against risk for banking companies, US commercial banks are increasingly using other risk mitigation techniques as well. Banks of all sizes employ on-balance sheet techniques to mitigate interest rate risk, writing adjustable rate rather than fixed rate loans, and using increasingly sophisticated duration-based asset liability management programs. The advent of credit scores to analyze consumer and small businesses loan applications has reduced credit risk by improving lenders' estimates of loan default probabilities. For example, DeYoung, Glennon, and Nigro (2008) find that scoring small business loans reduces the probability of loan default. Larger and more sophisticated banks take off-balance sheet positions in interest rate swaps, credit default swaps, foreign exchange options, and other derivatives contracts to offset their exposures to movements in interest rates, foreign currencies, and loan defaults. The increased geographic dispersion of banks—either by operating banking offices in multiple economic regions, or by purchasing asset-backed securities backed by loans from other economic regions—has diversified banks' investment portfolios.

These improvements in risk mitigation at US commercial banks occurred at the same time that industry earnings had ascended to record levels; without the benefit of hindsight, these data suggested that the risk-adjusted returns of US banks had improved substantially, and that the likelihood of severe downturns in earnings and large capital losses were greatly reduced. However, one would have been wrong to conclude from this handsome financial performance that the US banking industry had become invulnerable to a banking crisis—unfortunately, history likes to repeat itself.

In mid-2007, US home prices began a steep decline and unusually large numbers of home owners began defaulting on their mortgage loans. Most of these bad mortgages were subprime loans—that is, loans to households with poor credit

histories, little collateral, and questionable long-run abilities to service the loans—and were underwritten using the automated lending and asset securitization processes central to the growth of large US retail and investment banking franchises during the 1990s and 2000s. Trillions of dollars of securities backed by these subprime loans declined in value, imposing large losses on the portfolios of commercial banks, investment banks, and other institutional investors in MBS. These losses, combined with the complex nature of many of these subprime MBS contracts, resulted in a 'credit crunch' as uncertain investors pulled their money out of mortgage investments.

As these funding markets seized up and a systemic collapse of the financial sector became a distinct possibility, US financial regulators took a series of unprecedented policy actions. The Federal Reserve (among other actions too numerous to detail here) opened wide its discount window in order to make liquidity freely available to all types of financial institutions; temporarily guaranteed all investments in money market mutual funds against losses; pledged to purchase up to $600 billion in agency (GSE) MBS from financial institutions; and, for the first time since the Great Depression, began lending directly to non-financial firms by purchasing their newly issued commercial paper. The Fed also set into motion a chain of events that marked the end of the large independent investment banking model in the US. The Fed subsidized JPMorgan Chase's acquisition of the insolvent Bear Stearns; a few weeks later denied assistance to Lehman Brothers, which required the firm to file for bankruptcy; and a week after that converted Goldman Sachs and Morgan Stanley to bank holding company charters—in essence, these two firms accepted stricter regulations and supervisory scrutiny in exchange for access to the Federal Reserve's discount window and the ability to issue inexpensive insured deposits. The US Treasury nationalized Fannie Mae and Freddie Mac (as discussed above); provided $150 billion in loans and other support to prop up and eventually nationalize AIG, the largest insurance company in the US; and gained authority under the Troubled Assets Relief Program (TARP) to inject up to $700 billion into commercial banks via equity injections and purchases of mortgage-backed securities, which it began doing in October 2008. The FDIC increased deposit insurance coverage from $100,000 to $250,000 per account; provided billions of dollars of open-bank assistance to prevent the insolvency of Citigroup, at the time the largest banking company in the world; and used a variety of innovative structures and techniques to resolve insolvent banks and thrifts, including the largest depository failure in US history (to date) in Washington Mutual.

As of the publication of this volume, the long-run ramifications of the breakdown in the transactions banking business model are unknown. Moving forward, we are likely to witness some retrenchment toward more traditional banking models: a larger percentage of retail loans will be funded by bank deposits rather than asset-backed securities, and there will be less credit made available for subprime borrowers. But technology does not disappear, and the second application of

a new technology learns from the mistakes of the first application. Although subprime mortgage securitization had essentially stopped by year end 2008, the loan-to-securitize credit channel generates far too much financial and informational efficiency to be abandoned; it will continue to be employed in the future, albeit with more prudence by lenders, more diligence by investors, and greater regulatory oversight. Similarly, it is far too early to gauge the efficacy of the unprecedented policy actions taken by US financial regulators in 2008 and 2009 to stabilize the financial sector. Having approved over $1 trillion of taxpayer funds to 'bail out' financial institutions in the short run, the US Congress faces a climate in which stricter regulation of financial institutions in the long run is a political necessity. If the new regulations stabilize financial markets and eliminate the poor incentives facing financial institution managers—and at the same time manage to avoid disrupting or destroying the true underlying efficiencies generated by technological and financial innovation—then the US banking industry should weather this storm and emerge more safe, productive, and profitable than would have been possible twenty-five years ago.

References

Berger, A. N. (2003). 'The Economic Effects of Technological Progress: Evidence from the Banking Industry', *Journal of Money, Credit, and Banking*, 35: 141–76.

—— Bonime, S. D., Goldberg, L. G., and White, L. J. (2004). 'The Dynamics of Market Entry: The Effects of Mergers and Acquisitions on De Novo Entry and Small Business Lending in the Banking Industry', *Journal of Business*, 77: 797–834.

—— and DeYoung, R. (2001). 'The Effects of Geographic Expansion on Bank Efficiency', *Journal of Financial Services Research*, 19: 163–84.

—— —— (2006). 'Technological Progress and the Geographic Expansion of the Banking Industry', *Journal of Money, Credit, and Banking*, 38: 1483–513.

—— —— Flannery, M. J., Lee, D., and Öztekin, Ö. (2008). 'Why Do Large Banking Organizations Hold So Much Capital?', *Journal of Financial Services Research*, 34: 123–49.

—— and Dick, A. (2007). 'Entry into Banking Markets and the First-Mover Advantage', *Journal of Money, Credit, and Banking*, 39: 775–808.

—— —— Goldberg, L. G., and White, L. J. (2007). 'Competition from Large, Multimarket Firms and the Performance of Small, Single-market Firms: Evidence from the Banking Industry', *Journal of Financial Economics*, 39: 331–68.

—— Frame, W. S., and Miller, N. H. (2005). 'Credit Scoring and the Availability, Price, and Risk of Small Business Credit', *Journal of Money, Credit, and Banking*, 37: 191–222.

—— and Mester, L. J. (1997). 'Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?', *Journal of Banking & Finance*, 21: 895–947.

—— Miller, N. H., Petersen, M. A., Rajan, R. G., and Stein, J. C. (2005). 'Does Function Follow Organizational Form: Evidence from the Lending Practices of Large and Small Banks', *Journal of Financial Economics*, 76: 237–69.

Berlin, M. (2007). 'Dancing with Wolves: Syndicated Loans and the Economics of Multiple Lenders, *Federal Reserve Bank of Philadelphia Business Review*, Q3: 1–8.

Carey, M., Prowse, S., Rea, J., and Udell, G. F. (1993). 'The Economics of Private Placements: A New Look', *Financial Markets, Institutions & Instruments*, 2: 1–66.

Clark, J. A. (1988). 'Economies of Scale and Scope at Depository Financial Institutions: A review of the Literature', *Federal Reserve Bank of Kansas City Economic Review*, September: 16–33.

DeYoung, R. (2005). 'The Performance of Internet-based Business Models: Evidence from the Banking Industry', *Journal of Business*, 78: 893–947.

—— Glennon, D., and Nigro, P. (2008). 'Borrower-Lender Distance, Credit Scoring, and the Performance of Small Business Loans', *Journal of Financial Intermediation*, 17: 113–43.

—— and Hasan, I. (1998). 'The Performance of De Novo Commercial Banks: A Profit Efficiency Approach', *Journal of Banking & Finance*, 22: 565–87.

—— —— and Kirchhoff, B. (1998). 'The Impact of Out-of-State Entry on the Efficiency of Local Commercial Banks', *Journal of Economics & Business*, 50: 191–203.

—— Hunter, W. C., and Udell, G. F. (2004). 'The Past, Present, and Probable Future for Community Banks', *Journal of Financial Services Research*, 25: 85–133.

—— Lang, W. W., and Nolle, D. L. (1997). 'How the Internet Affects Output and Performance at Community Banks', *Journal of Banking & Finance*, 31: 1033–60.

—— and Ors, E. (2004). 'Advertising and Pricing at Multiple-Output Firms: Evidence from U.S. Thrift Institutions', Federal Reserve Bank of Chicago Working Paper 2004-25.

—— and Rice, T. (2004a). 'How Do Banks Make Money? A Variety of Business Strategies', *Federal Reserve Bank of Chicago Economic Perspectives*, 28: 52–67.

—— —— (2004b). 'Non-Interest Income and Financial Performance at U.S. Commercial Banks', *Financial Review*, 39: 101–27.

—— and Roland, K. P. (2001). 'Product Mix and Earnings Volatility at Commercial Banks: Evidence from a Degree of Total Leverage Model', *Journal of Financial Intermediation*, 10: 54–84.

Eisenbeis, R. A. and Wall, L. D. (2002). 'The Major Supervisory Incentives Post-FDICIA: Are they Based on the Goals of PCA? Should They Be?', Federal Reserve Bank of Atlanta Working Paper 2002-31.

Evanoff, D. and Israilevich, P. (1991). 'Productive Efficiency in Banking', *Federal Reserve Bank of Chicago Economic Perspectives*, July: 11–32.

—— and Ors, E. (forthcoming). 'Local Market Consolidation and Bank Productive Efficiency', *Journal of Money, Credit and Banking*.

Frame, W. S., Srinivasan, A., and Woosley, L. (2001). 'The Effect of Credit Scoring on Small Business Lending', *Journal of Money, Credit, and Banking*, 33: 813–25.

—— and Wall, L. D. (2002). 'Financing Housing through Government-Sponsored Enterprises', *Federal Reserve Bank of Atlanta Economic Review*, Q1: 29–43.

—— and White, L. J. (2004). 'Empirical Studies of Financial Innovation: Lots of Talk, Little Action?', *Journal of Economic Literature*, 42: 116–44.

Gerdes, G. R. and Walton, J. K. (2002). 'The Use of Checks and Other Retail Noncash Payments in the United States', *Federal Reserve Bulletin*, August: 360–74.

Gordy, M. B. and Heitfield, E. (2008). 'Basel II Regulatory Capitol Accord', in *Oxford Handbook of Banking*, Berger, A. N., Molyneux, P., and Wilson, J. O. S. (eds.). Oxford: Oxford University Press.

HUGHES, J. P., LANG, W. W., MESTER, L. J., and MOON, G. (1996). 'Efficient Banking under Interstate Branching', *Journal of Money, Credit, and Banking*, 28: 1045–71.

HUMPHREY, D. B. (2002). 'U.S. Cash and Card Payments over 25 years', Florida State University, Unpublished Manuscript.

JONES, K. D. and CRITCHFIELD, T. (2005). 'Consolidation in the U.S. Banking Industry: Is the "Long, Strange Trip" About to End?', *Federal Deposit Insurance Corporation Banking Review*, 17: 31–61.

MESTER, L. J. (1987). 'Efficient Production of Financial Services: Scale and Scope Economies', *Federal Reserve Bank of Philadelphia Business Review*, January: 15–25.

—— (1997). 'What's the Point of Credit Scoring?', *Federal Reserve Bank of Philadelphia Business Review*, September/October: 3–16.

PETERSEN, M. A. and RAJAN, R. G. (2002). 'Does Distance Still Matter? The Information Revolution and Small Business Lending', *Journal of Finance*, 57: 2533–70.

ROSSI, C. V. (1998). 'Mortgage Banking Cost Structure: Resolving an Enigma', *Journal of Economics & Business*, 50: 219–34.

SCOTT, J. A. (2004). 'Small Business and the Value of Community Financial Institutions', *Journal of Financial Services Research*, 25: 207–30.

STEIN, J. C. (2002). 'Information Production and Capital Allocation: Decentralized versus Hierarchical Firms', *Journal of Finance*, 57: 1891–921.

STIROH, K. J. (2004a). 'Diversification in Banking: Is Non-Interest Income the Answer?', *Journal of Money, Credit, and Banking*, 36: 853–82.

—— (2004b). 'Do Community Banks Benefit from Diversification?', *Journal of Financial Services Research*, 25: 135–60.

CHAPTER 32

BANKING IN THE EUROPEAN UNION

JOHN GODDARD
PHILIP MOLYNEUX
JOHN O. S. WILSON[1]

Introduction

This chapter surveys the banking industry within the group of countries known as the European Union-15. This group comprises the fifteen countries that were members of the European Union prior to the accession of twelve new member countries in 2004 and 2007, which increased the total membership to twenty-seven. The European Union-15 are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Luxembourg, Italy, the Netherlands, Portugal, Spain, Sweden, and the UK.

Since the passing of the First Banking Directive in 1977, European Union legislation has been directed consistently toward the reduction of barriers to cross-border banking activity. Deregulation of financial markets at the national level has reduced or eliminated many of the lines of demarcation between banks and other financial service providers, and has helped facilitate cross-border

[1] The authors would like to thank Barbara Casu, Claudia Girardone, and David Marques for useful comments and suggestions. We would also like to thank Yalman Onaran of Bloomberg for supplying data on bank write-downs. The usual disclaimer applies.

competition. Despite these developments, however, by the end of the 2000s there were still significant barriers to the creation of a fully integrated European single market in banking and financial services. Retail banking and small business lending remained nationally oriented, with relatively little cross-border activity. In contrast, wholesale banking had become highly integrated.

During the 1990s and 2000s, many European banks expanded the scale of their operations, in some cases through merger and acquisition (M&A). Consolidation was motivated by the objectives of realizing scale and scope economies, reducing labor and other variable costs, cutting operational inefficiencies, and spreading risk through product or geographic diversification. Rapid growth in the loan portfolios of some banks was financed via the securitization of prospective cash flows from sources such as mortgages and credit card debt, which commonly took place off-balance sheet. Fast growth in European banks' non-interest income reflected the growing use of securities-based financing by private sector companies, and the increase in demand for protection products (insurance and personal pensions), and investment in mutual funds, by the household sector.

The geographic aspect to diversification entails increased cross-border banking activity. Expansion by banks into other European Union countries has taken place mainly through the establishment of subsidiaries (European Central Bank, 2007a). Changes embodied in the European Company Statute allowed banks to form single legal entities that can operate freely across European Union national borders, enabling the conversion of subsidiaries to branches. However, subsidiaries have been predominantly the preferred cross-border organizational form, suggesting that the benefits associated with risk spreading between different legal entities within a banking group are of strategic importance.

The remainder of this chapter is structured as follows. The second section outlines key characteristics of the banking sectors of the European Union-15 at the country level. The third section adopts a broader comparative approach in examining the structure and performance of the European banking industry. The fourth section examines the integration of European banking across national borders, and assesses the extent to which the European Union's objective of creating a European single market in financial services had been realized. The fifth section discusses the challenges for supervision and financial stability that are raised by the development of a European single market, and the impact of the credit crisis since 2007. The sixth section summarizes and concludes.

Banking in European Union countries

This section surveys the key characteristics of national banking sectors within the European Union-15. Sub-sections 1–5 provide country profiles for the five largest

banking systems—France, Germany, Italy, Spain, and the UK. Sub-section 6 provides a brief commentary on key features of the banking sectors of the other eleven European Union-15 countries. To provide a snapshot of the key operators in European banking, Table 32.1 identifies the thirty largest European banks in 2006, according to data published annually by the *Banker*. Based on the same source, Tables 32.2–32.6 identify the ten largest banks by country for France, Germany, Italy, Spain, and the UK in 1986, 1996, and 2006.

France

Until the early 1980s, French law distinguished between ordinary deposit banks and various other types of bank, including investment banks, medium- and long-term credit banks, and other financial institutions, each of which had distinct legal status (see Table 32.2). The French banking sector was highly compartmentalized, with banks being subject to a bewildering array of legal restrictions on their activities, and a variety of supervisory arrangements. For example, several of the leading cooperative and mutual banks, including Crédit Agricole, Banques Populaires, and Crédit Mutuel, had their own individual governance structures. Several of the largest deposit banks had been nationalized immediately after World War II; and in 1982 a number of other French banks that were not already nationalized were taken into state ownership (Commission Bancaire, 2002).

The 1984 Banking Act created a single legal framework and harmonized regulatory and supervisory arrangements for all banks. During the remainder of the 1980s and the 1990s, the trend was toward further financial liberalization and deregulation. The 1984 Act and other legislation recognized the principle of universal banking, eliminating many restrictions on bank lending and on the lines of business different types of bank were permitted to transact. French banks responded vigorously by increasing their mutual fund and insurance activities. Meanwhile, the French state progressively reduced its involvement in the financial sector through several waves of bank privatizations, in 1986–7, 1993 and again in the late 1990s. In 2006, the former financial division of the French post office, Le Banque Postal, was established as a common law bank. Several bank privatizations were structured with a view to promoting national champions through the formation of large shareholder-controlled universal banking groups (Commission Bancaire, 2002).

During the 1990s, a wave of consolidation reduced the number of French banks by more than 50 percent, from just over 2,000 in 1990 to below 1,000 by 2002. This decline in bank numbers reflected: reduced compartmentalization and increased universality within the banking sector; consolidation of mutual and cooperative banks; and the emergence of several transnational banking groups operating across national borders both within and beyond the European Union. Mutual and

Table 32.1. Largest banks (ranked by total assets) in European Union-15 countries

EU-15 rank	World rank	Bank	Country	Total assets ($million)	Tier 1 capital ($million)	Return on assets (%)	Cost–income ratio (%)	Risk-adjusted capital assets ratio
1	2	Barclays Bank	UK	1,956,786	45,161	0.72	58.69	11.70
2	3	Banque Nationale de Paris	France	1,896,935	45,305	0.73	61.07	10.50
3	5	HSBC Holdings	UK	1,860,758	87,842	1.19	51.33	13.54
4	6	Crédit Agricole	France	1,818,341	84,937	0.77	61.10	10.00
5	7	Royal Bank of Scotland	UK	1,710,703	58,973	1.05	53.01	11.73
6	9	Deutsche Bank	Germany	1,483,248	32,264	0.72	70.16	12.80
7	12	ABN AMRO	Netherlands	1,299,966	31,239	0.51	74.94	11.14
8	13	Société Générale	France	1,260,162	29,405	0.84	77.07	11.11
9	15	ING Bank	Netherlands	1,178,697	33,958	0.56	63.76	11.02
10	16	HBOS	UK	1,160,245	44,030	0.97	47.56	12.00
11	17	Banco Santander	Spain	1,098,213	46,805	1.05	49.57	12.49
12	18	Unicredit	Italy	1,084,267	38,700	1.65	56.50	10.50
13	21	Fortis Bank	Belgium	888,750	22,255	0.81	61.2	11.1
14	23	Commerzbank	Germany	801,184	20,410	0.39	59.75	11.00
15	24	Dexia	Belgium	746,402	17,158	0.60	49.6	10.3
16	25	Rabobank Group	Netherlands	732,708	34,757	0.49	68.53	11.00
17	26	Groupe Caisse d'Epargne	France	710,801	24,159	0.97	76.82	10.40
18	31	Lloyds TSB Bank	UK	674,515	25,183	1.24	47.74	10.70
20	32	Dresdner Bank	Germany	654,928	16,422	0.24	79.79	15.60
21	33	Crédit Mutuel	France	635,685	29,792	0.89	58.50	12.00

22	34	DZ Bank	Germany	574,750	11,721	0.32	72.87	12.60
23	35	Landesbank Baden-Wurttemberg	Germany	564,010	14,181	0.26	49.30	11.00
24	36	Banco Bilbao Vizcaya Argentaria	Spain	542,494	25,779	1.71	43.89	12.00
25	38	Danske	Denmark	483,901	16,988	0.68	53.17	11.44
26	41	Nordea Group	Sweden	456,855	17,315	1.10	51.81	9.80
27	42	Bayerische Landesbank	Germany	455,389	14,085	0.28	55.70	10.70
28	43	KBC Group	Belgium	428,553	14,820	1.41	62.01	11.74
29	44	Groupes Banques Populaires	France	402,090	22,257	0.85	74.27	12.70
30	45	Banca Intesa	Italy	384,276	16,736	1.43	58.90	9.50

Notes: The data in Table 32.1 are based on consolidated accounts for banking groups whose home country is an EU-15 member. The consolidated accounts data may include subsidiaries whose host country is outside the EU-15. Likewise, the data in Tables 32.2–32.5 are based on consolidated accounts for banking groups whose home country is France, Germany, Italy, and the UK, respectively. However, subsidiaries whose host country is France, Germany, Italy, or the UK are also eligible for inclusion in Tables 32.2–32.6, respectively, regardless of the home country of the banking group concerned. Therefore, there is an element of double counting in Tables 32.2–32.5. For example, in 2006, HVB (Hypovereinsbank), a German subsidiary of the Italian banking group Unicredit, is included in the consolidated accounts of Unicredit reported in Table 32.4, and appears separately in Table 32.3 in its own right.

Source: *Banker*.

Table 32.2. Largest banks (ranked by total assets) in France, 1986, 1996, and 2006

Country rank	1986		1996		2006	
	Bank	Assets ($million)	Bank	Assets ($million)	Bank	Assets ($million)
1	Crédit Agricole	154,407	Crédit Agricole	477,336	Banque Nationale de Paris	1,896,935
2	Banque Nationale de Paris	141,871	Banque Nationale de Paris	355,366	Crédit Agricole	1,818,341
3	Crédit Lyonnais	132,076	Société Générale	339,996	Société Générale	1,260,162
4	Société Générale	116,013	Crédit Lyonnais	310,040	Groupe Caisse d'Epargne	710,801
5	Paribas	93,240	Compagnie Financière de Paribas	290,720	Crédit Mutuel	635,685
6	Indosuez	40,372	Groupe Caisse d'Epargne	224,301	Natixis	604,020
7	Banques Populaires	38,801	Groupes Banques Populaires	115,524	Groupe Caisse d'Epargne	490,618
8	Crédit Commercial de France	32,038	Union Européenne de CIC	115,340	Groupes Banques Populaires	402,090
9	BFCE	26,144	Crédit Mutuel	111,432	Dexia	400,196
10	Compagnie Bancaire	21,564	Crédit National	55,089	HSBC France	151,115

Note: See notes to Table 32.1.

Table 32.3. Largest banks (ranked by total assets) in Germany, 1986, 1996, and 2006

Country rank	1986		1996		2006	
	Bank	Assets ($million)	Bank	Assets ($million)	Bank	Assets ($million)
1	Deutsche Bank	131,808	Deutsche Bank	569,906	Deutsche Bank	1,483,248
2	Dresdner Bank	101,185	Dresdner Bank	355,605	Commerzbank	801,184
3	Westdeutsche Landesbank	76,243	Westdeutsche Landesbank	295,774	Hypovereinsbank	669,081
4	Commerzbank	75,430	Commerzbank	287,692	Dresdner Bank	654,928
5	Bayerische Vereinsbank	72,125	Bayerische Vereinsbank	258,505	DZ Bank	574,750
6	Bayerische Hypotheken & Wechsel	62,019	Bayerische Landesbank	221,488	Landesbank Baden-Wurttemberg	564,010
7	DG Bank	57,442	Bayerische Hypotheken & Wechsel	218,294	Bayerische Landesbank	455,389
8	Bayerische Landesbank	55,870	Bankgesellschaft	216,265	West LB	312,013
9	Norddeutsche Landesbank	48,607	DG Bank	210,156	Norddeutsche Landesbank Girozentrale	267,474
10	Kreditanst. Fur Wiederaufbau	44,142	Kreditanst. Fur Wiederaufbau	153,207	HSH Nordbank	244,587

Note: See notes to Table 32.1.

Table 32.4. Largest banks (ranked by total assets) in Italy, 1986, 1996, and 2006

Country rank	1986		1996		2006	
	Bank	Assets ($million)	Bank	Assets ($million)	Bank	Assets ($million)
1	Banca Nazionale del Lavoro	72,695	Cariplo	125,910	Unicredit	1,084,267
2	Istituto Bancario San Paolo	58,589	Istituto Bancario San Paolo di Torino	171,317	Banca Intesa	384,276
3	Banca Commerciale Italiana	54,800	Banca di Roma	141,077	San Paolo IMI	380,022
4	Monte dei Paschi di Siena	52,869	Banca Commerciale Italiana	115,448	Banca Monte Dei Paschi di Siena	208,818
5	Cariplo	52,782	Banca Nazionale del Lavoro	112,857	Capitalia Gruppo Bancario	180,602
6	Banco di Roma	51,998	Instituto Mobiliare Italiano	52,228	Unione di Banche Italiane	153,968
7	Credito Italiano	45,380	Monte dei Paschi di Siena	93,177	Gruppo Banco Popolare di Verona e Novara	90,471
8	Banco di Napoli	44,496	Credito Italiano	114,378	Banca Popolare dell'Emilia Romagne	59,604
9	Banco di Sicilia	24,432	Cariverona Banca	23,017	Banco Popolare Italiana	55,825
10	Banca Naz. Dell'Agricoltura	23,253	Rola Banca 1473	40,618	Banca Popolare di Milano	52,919

Note: See notes to Table 32.1.

Table 32.5. Largest banks (ranked by total assets) in Spain, 1986, 1996, and 2006

Country rank	1986		1996		2006	
	Bank	Assets ($million)	Bank	Assets ($million)	Bank	Assets ($million)
1	Banco Central	28,195	Banco Santander	149,881	Banco Santander	1,098,213
2	Banco Hispano Americano	22,512	Banco Bilbao Vizcaya	131,069	Banco Bilbao Vizcaya Argentaria	542,494
3	Banco de Bilbao	21,904	Banco Hispano Americano	89,436	Caja de Ahorros y pen. De Barcelona	275,416
4	Banco Español de Crédito	21,798	Argentaria	83,832	Caja de Ahorros y Monte de Piedad de Madrid	180,367
5	Banco de Santander	20,946	Caja de Ahorros y pen. De Barcelona	75,773	Banco Popular Español	120,704
6	Banco de Vizcaya	20,793	Caja de Ahorros de Madrid	42,108	Banco Sabadell	95,851
7	Banco Exterior de España	19,141	Banco Popular Español	26,514	Caixa Catalunya	88,965
8	Caja de Pensiones La Caixa	14,563	Banco Sabadell	15,131	Caja Mediterráneo	81,975
9	Banco Popular Español	12,592	Bank Inter	14,087	Gruppo Bancaja	73,451
10	Caja de Ahorros de Madrid	10,515	Gruppo Bancaja	15,830	Unicaja	37,229

Note: See notes to Table 32.1.

Table 32.6. Largest banks (ranked by total assets) in UK, 1986, 1996, and 2006

Country rank	1986		1996		2006	
	Bank	Assets ($million)	Bank	Assets ($million)	Bank	Assets ($million)
1	National Westminster Bank	122,862	HSBC Holdings	401,686	Barclays Bank	1,956,786
2	Barclays Bank	116,380	Barclays Bank	315,846	HSBC Holdings	1,860,758
3	Midland	78,397	National Westminster Bank	314,716	Royal Bank of Scotland	1,710,703
4	Lloyds	705,23	Lloyds TSB	250,241	HBOS	1,160,245
5	Standard Chartered	47,514	Abbey National	210,581	Lloyds TSB	674,515
6	Royal Bank of Scotland	24,472	Royal Bank of Scotland	954,79	Abbey National	376,541
7	TSB Group	18,960	Bank of Scotland	770,44	Standard Chartered	266,047
8	Bank of Scotland	14,429	Standard Chartered	715,54	Northern Rock	198,293
9	Kleinwort	13,024	Ford Credit Europe	19,210	Alliance and Leicester	134,584
10	Morgan Grenfell	8,141	Schroders	17,610	Bradford and Bingley	89,035

Note: See notes to Table 32.1.

cooperative banks have been among the leading participants in the M&A trend—for example, Crédit Agricole acquired Indosuez in 1996 and the ailing Crédit Lyonnais in 2003, and Crédit Mutuel acquired CIC, previously in state ownership, in 1999. Other significant mergers in recent years include the acquisition by Banque Nationale de Paris (BNP) of Paribas in 1999, the acquisition by Banque Populaires of Natexis in 1999, the acquisition by HSBC of Crédit Commercial de France (CCF) in 2000, and the acquisitions by Caisse d'Epargne of the French subsidiary of the Italian group San Paolo IMI in 2003, and Compaigne Financière Eulia and CDC IMI in 2004.

In the late 2000s, the French banking sector was dominated by six domestic groups: BNP-Paribas, Crédit Agricole, Société Générale, Caisse d'Epargne, Crédit Mutuel and Banques Populaires; and one foreign group, HSBC France. Four of the top six domestic groups (the exceptions being BNP-Paribas and Société Générale) were mutuals or cooperatives. All four of these groups have adopted the universal banking model; all four can reasonably be described as substantially profit-oriented; and all four are owners of several non-cooperative subsidiaries. In 2006 BNP-Paribas was the largest bank based inside the euro zone. Having made a number of large-scale acquisitions abroad, BNP-Paribas was operating in more than twenty countries. Société Générale has also expanded significantly its operations internationally, but through a different model: by making large numbers of smaller-scale acquisitions, predominantly in developing countries (Fédération Bancaire Française, 2005).

Competition in the French banking sector appears to have intensified as financial market liberalization has proceeded. Interest margins have declined, while lending volumes have increased. The proportion of French households with a bank account is one of the highest in the European Union; and the French banking sector is highly active in life insurance and other specialized financial services such as asset management. Despite sweeping changes in bank ownership and the disappearance of many of the smaller banks, the market shares of the leading groups have remained relatively stable.

The French banking sector entered the credit crisis commencing from mid-2007 in a relatively healthy condition. As in many other countries, however, the situation has deteriorated subsequently. Société Générale has been hit by exposure to subprime lending losses, and an alleged fraud on the part of one of its traders that was uncovered at the start of 2008. Both Crédit Argricole and Caisse d'Espargne have sustained substantial losses, although BNP Paribas has fared rather better and was able to acquire the remaining assets of the Belgian-Dutch Fortis bank after it was nationalized by the Dutch government in October 2008. In October 2008, against a background of growing uncertainty concerning the financial health of the banking sector, the French government set aside up to €40 billion to recapitalize banks, and provided loan guarantees to the value of €320 billion.

Germany

Germany has the second largest banking sector by asset size in Europe. The German banking sector accounts for just under one-third of all banks in the euro area, has the largest number of branches, and is the largest employer of any European banking sector. Germany had more than 5,000 banks in 1980, and consequently has traditionally been considered as over-banked. Since the 1980s there has been less consolidation between major banks in Germany than elsewhere, although there has been a significant decline in the number of small mutual banks. The total number of banks had fallen to 2,050 by 2006 (see Table 32.3).

German banks subdivide into universal banks, and specialist financial institutions. Universal banks, which dominate the banking sector, combine commercial banking with investment banking, and the provision of insurance and other financial services, as well as non-financial business. Universal banks include commercial, public (land and savings), and cooperative banks.

Commercial banks are shareholder owned, and include several very large institutions, such as Deutsche Bank, Dresdner Bank, and Commerzbank. Some smaller, regional commercial banks have a traditional geographic focus in both lending and location, but during the 2000s several regional banks expanded throughout Germany, and beyond. A number of branches and subsidiaries of foreign banks are also classed as commercial banks.

Public sector banks include the Land banks (Landesbanken), which are owned by the state (Land) governments; and the savings banks (Sparkassen), predominantly owned by local government. These banks have the dual objectives of operating commercially, but also implementing the developmental objectives of state and local governments. The Land banks were originally established to operate the payments' system within each state. Financed by the issue of debt through wholesale markets, they provide wholesale banking services to smaller savings banks. Traditionally the Land banks obtained the highest credit ratings due to state government guarantees, prior to an European Union ruling in 2001 that these guarantees provided an unfair competitive advantage. The guarantees were withdrawn in July 2005 (although certain other guarantees will remain in force until 2015), triggering a wave of product and geographic diversification, as well as consolidation through M&A, among the Land banks.

German savings banks focus on retail deposit taking and lending within a local or regional geographic area. Savings banks dominate retail banking in Germany. Typically they have a close affiliation to their respective Land banks, which provide clearing and other forms of payment and settlement services. The profitability and capitalization of the savings banks has been similarly affected by the removal of the state government guarantees, and the sector has found it increasingly difficult to source cheap funding.

Cooperative banks (Kreditgenossenschaften) are member-owned and member-controlled not-for-profit organizations, historically closely affiliated to key trades such as agriculture. Like the savings banks, they focus on deposit taking activities within narrowly defined local geographic areas. They are also members of large and highly diversified regional cooperatives (Genossenschaftliche Zentralbanken), to which they lend funds for investment.

A number of German banks, including mortgage banks and building and loan associations, specialize in providing finance for specific purposes, such as construction, real estate, and home ownership. These specialist financial institutions can be classified by their ownership status: the Hypothekbanken are privately owned; while the Bausparkassen are owned by state or local government.

In common with many European Union-15 countries, the number and importance of foreign banks operating in Germany increased during the 2000s. Although the differences between German commercial, savings, and cooperative banks have lessened, the German banking sector remains relatively fragmented. Some recent consolidation has taken place with the acquisition of Dresdner Bank (previously owned by Allianz) by Commerzbank in August 2008. Continued public ownership of a large proportion of total banking sector assets, and the fragmented ownership structure of the cooperative banks, are significant barriers to further consolidation and rationalization.

Pressures to consolidate have been heightened by the credit crisis. By the end of 2008, the German government had provided up to € 400 billion in loan guarantees (similar to the 'blanket' guarantees provided by the governments of several smaller European Union countries, including Austria, Denmark, Hungary, Ireland, Slovakia, and Slovenia). The German government also allocated a maximum of €80 billion for capital injections. The first major bank to avail of the latter was Commerzbank, which sought an €8.2 billion injection in exchange for preference shares. At the time of writing, this arrangement was being examined by the European Union competition authorities, as it was widely reported that the state would receive a return below the 10 percent stipulated under the terms of the package.

Italy

The Italian banking sector comprises commercial banks, cooperative banks (banche popolari), mutual banks (banche di credito cooperativo), and foreign banks. At the end of the 1980s, the Italian banking sector was one of Europe's most heavily regulated. Subsequently, however, bank privatizations and the gradual withdrawal of government-owned holding companies from bank ownership and control has contributed toward a major transformation (see Table 32.4).

The 1990 Legge Amato Law, that was later incorporated into the 1993 Banking Act, created a legal framework that divided the public sector banks into public limited companies (banks) and public foundations. The Consolidated Law on Financial Intermediation (Ciampi Law) of 1998 allowed for the privatization of state-owned banks, as the public foundations which held bank shares were permitted to sell these shares to private investors. Notable bank privatizations since the mid-1990s include Credito Italiano, Banca Commerciale Italiana, Mediobanca, Banca Nazionale del Lavoro, Banco di Napoli, and Banco di Roma. Consolidation has followed privatization, and has led to the formation of large universal banking groups, which now dominate Italian domestic banking.

The domestic consolidation trend during the 2000s was encouraged by the Italian government, which seemed to consider the foreign acquisition of large Italian banks to be politically sensitive. Initially, the largest mergers were between domestic banks, but several cross-border mergers involving Italian banks have taken place. In 2005, UniCredit acquired the German bank Bayerische Hypo- und Vereinsbank (HVB). In 2006, the French bank BNP Paribas acquired Banca Nazionale del Lavoro; and Banca Intesa merged with Sanpaolo IMI. In 2007, Veneto Banca acquired Banca Popolare di Intra; Banche Popolari Unite, and Banca Lombarda e Piemontese S.p.A. merged to form the new cooperative group Unione di Banche Italiane; Banco Popolare di Verona e Novara merged with Banca Popolare Italiana; and UniCredit merged with Capitalia.

Despite this wave of consolidation, however, concentration is relatively low. Between 2002 and 2006 CR_5 (five-firm assets concentration ratio) fell from 31 percent to 26 percent; although the 2007 mergers listed above have increased this figure back above 30 percent. Consolidation involving large banks has been accompanied by an increase in the number of small retail banks operating at regional or local level. Moreover, the removal in 1993 of regulations that controlled branching has resulted in a proliferation of bank offices.

Historically, foreign bank involvement in the Italian banking sector has been limited, and restricted to niche areas such as investment banking and wealth management. During the 2000s, however, the presence of foreign banks has increased. For example, in addition to the large cross-border deals mentioned above, Crédit Agricole of France has acquired controlling interests in two mutual banks, Cassa di Risparmio di Parma e Piacenza and Banca Popolare Friuladria. In 2006, the branches and subsidiaries of foreign banks accounted for around 20 percent of Italian banking sector assets.

At the time of writing, the credit crisis appears to have had a less severe impact on the largest Italian banks than on their counterparts in several other major European countries. Italian banks had less involvement in holding securitized assets off-balance sheet, and in other types of financial innovation. An exception is Unicredit, which has encountered some difficulties arising from its acquisition of the German

bank HypoVereinsbank in 2005. In October 2008, the Italian government made up to €40 billion in treasury bills available to the banking sector for use in refinancing non-performing assets.

Spain

Spain has the European Union's fifth-largest banking sector, with total assets of €2.5 trillion in 2006 (see Table 32.5). There is a diverse range of ownership types, including commercial banks, savings banks, cooperative banks, and specialized credit institutions (which transact business such as hire purchase and leasing). Concentration is moderate, with $CR_5 = 40$ percent in 2006. The three largest banks by asset size in 2006 were Banco Santander, BBVA (Banco Bilbao Vizcaya Argentaria, and Caja de Ahorros y Pensiones de Barcelona (La Caixa). Commercial and savings banks accounted for 42 percent and 40 percent of banking sector assets, respectively. In 2006, Spain had the densest branch network of any European Union-15 country (one bank branch for every 1,009 inhabitants).

The universal banking model is well established in Spain, with many banks transacting a broad spectrum of financial services. For example, the leading commercial banks are dominant in cross-selling mutual funds to their retail clients. Foreign banks are large in number, but account for only about 5 percent of banking sector assets, playing a somewhat peripheral role in niche areas.

Several Spanish banks have made major forays into Latin America. The two largest Spanish commercial banks (Banco Santander and BBVA) were also the two largest in Latin America in 2006. In the 2000s, Banco Santander expanded its operations significantly in other European countries, by purchasing the UK banks Abbey National in 2004, Alliance and Leicester, and the deposit base of Bradford and Bingley in 2008, and by expanding its consumer lending business in France and Germany.

The regulatory approach adopted by the Bank of Spain requires off-balance sheet vehicles to adhere to traditional on-balance sheet capital ratios (as well as other restrictions). During 2007–8, the largest Spanish banks wrote off smaller proportions of their loans portfolios than many of their European competitors. However, a sharp downturn in the Spanish economy in 2008, driven partly by the collapse of a property market bubble, seemed certain to increase non-performing loans ratios and erode capital buffers. In 2008, Spanish banks enjoyed generally robust profitability, and appeared to be well placed to withstand the impact of lower business growth and higher bad debt provisioning as the economy slowed. Indeed, Banco Santander made several foreign acquisitions in 2008, including Alliance and Leicester and part of Bradford and Bingley in the UK. Despite this positive picture, however, growing concerns about the capitalization of Spain's largest banks prompted the Spanish government to offer guarantees of up to €100 billion of bank debt in October 2008.

United Kingdom

The UK is home to Europe's largest banking sector, with total assets of €9.7 trillion in 2006. The corresponding figure for the second largest, Germany, was €7.1 trillion. The number of banks and building societies (mutual institutions that focus on residential mortgages) operating in the UK fell from 451 to 401 between 2002 and 2006 (see Table 32.6). Structural indicators of banking sector characteristics are strongly influenced by London's role as a major financial centre, and by the presence of foreign banks. In 2006, there were only sixty-four UK banks and fifty-nine building societies responsible for the bulk of sterling-denominated banking activity in the UK. The rest of the sector included 108 banks from other European Union countries, thirty-one from the US, nine from Japan, and the remainder from other countries.

In 2006, the UK retail banking sector was dominated by nine groups: Abbey National (owned by Spain's Banco Santander), Alliance and Leicester, Barclays, Bradford and Bingley, HBOS, HSBC, Lloyds TSB, Northern Rock, and Royal Bank of Scotland (RBS). Four of these were originally mutual building societies that converted to bank status: Abbey National (converted in 1989), Alliance and Leicester (1997), Northern Rock (1997), and Bradford and Bingley (2000). In 2000, RBS acquired the National Westminster (NatWest) Bank, the largest UK retail bank during the 1980s. HBOS was formed through a merger in 2001 of Halifax (a mutual building society that converted in 1997) and Bank of Scotland.

Concentration of the UK banking sector is moderate, with $CR_5 = 36$ percent in 2006. However, standard concentration ratios understate the extent of domestic banking sector concentration because foreign bank assets, which account for more than half of all banking sector assets, are included in the denominators. Foreign banks typically operate in London, transact mainly foreign currency business, and (mostly) do not compete for business in the retail and small business sectors.

Strong profitability among UK banks during the 2000s has derived mainly from retail financial services, especially mortgages. UK banks have benefited from buoyant macroeconomic conditions, a lack of competition in key sectors such as small business lending and payments, aggressive pricing and cross-selling, and financial innovation: it is estimated that more than 3,000 different mortgage products were available in 2006. The major retail banks have also increased their involvement in non-bank financial services, including insurance and asset management. Typically, expansion into these areas has been through the establishment of subsidiaries, although there have been some major acquisitions such as Lloyds TSB's purchase of Scottish Widows in 1999.

While bank shareholders have benefited from aggressive profit maximization strategies, it has been suggested that poorer segments of UK society have fallen behind. The availability of banking services in rural communities has been reduced through branch closures. The issue of financial exclusion remains on the policy

agenda, and raises awkward questions for strongly profit-oriented banks. Allegedly excessive bank charges have been the subject of a high-profile media campaign, led by consumer groups and some angry bank customers; and concerns have been expressed about lack of competition in small business lending and retail banking (Cruickshank, 2000; and Office of Fair Trading, 2008). Prior to the credit crisis and the run on Northern Rock in September 2007 (see below), such concerns seemed to have ruled out the possibility of further M&A deals between the top UK retail banks. However, turbulence in financial markets during 2007–8, culminating in the acquisition by Banco Santander of Alliance and Leicester and the retail deposit business and branch network of Bradford and Bingley. Lloyds TSB's (government-backed) acquisition of HBOS in September 2008 reflected continued concerns over the solvency and stability of the UK banking sector. Mergers among smaller mutual institutions included the acquisition of the Cheshire and Derbyshire building societies by Nationwide, and the merger between the Barnsley and Yorkshire building societies.

In October 2008, the UK government announced a range of measures to support the banking sector (Bank of England, 2008). A total of £50 billion was earmarked for a Bank Recapitalization Fund to support, if necessary, a number of major banks and building societies. Banks that choose to participate in the scheme must meet certain conditions relating to dividend payouts and executive compensation. By mid-December 2008, HBOS, Lloyds TSB, and RBS had drawn on the Fund. Participating banks are also eligible for government-backed guarantees of any chosen senior unsecured debt instruments for a period of up to three years. The UK government also made available an additional £200 billion to the Bank of England's Special Liquidity Scheme, which allows banks to exchange high-quality illiquid securities for liquid government securities. The deteriorating condition of the UK banking sector was highlighted in late November 2008, when the recapitalization of RBS involved the government underwriting a £15 billion rights issue. This had translated into a publicly owned shareholding in RBS of more than 70 percent by January 2009.

Other European Union-15 countries

This section provides a brief snapshot of the key features of the banking sectors of the ten remaining European Union-15 countries.

The Nordea group occupies a prominent position in the banking sectors of the Scandinavian countries, of which Denmark, Sweden, and Finland are European Union members (while Norway is a non-member with close commercial links to the European Union). In 2006, Nordea was the largest bank in Finland and Sweden, the second largest in Norway, and the third largest in Denmark. For several years, Nordea has been pursuing the objective of being the first bank to convert into a

Societas Europaea (European Company), in accordance with the European Company Statute. At the time of writing, however, the issue of how to structure a deposit guarantee scheme for a bank that is incorporated across several national borders remained unresolved and this (among other factors) appears to have scuppered Nordea's objectives. Another interesting feature of Scandinavian banking is the high penetration of Internet banking. In 2006 around 60 percent of bank customers used Internet banking services; the average for the European Union-15 was 24 percent.

Perhaps benefiting from lessons learned during the Scandinavian banking crises of the early 1990s, at the time of writing, banks in Norway, Sweden, Finland, and Denmark had fared better than their competitors in several other countries during the credit crisis of the late 2000s. Nevertheless, in October 2008, the Swedish government announced credit guarantees to banks and mortgage lenders of up to kroner 1.5 trillion, and created a kroner 15 billion bank stabilization fund. Similar measures were implemented in Denmark and Finland.

The banking sector of the Netherlands is highly concentrated, with $CR_5 = 85$ percent in 2006. The three largest banks in 2006 were ABN AMRO, ING, and the cooperative Rabobank. In October 2007, a consortium of three European banks: RBS, Fortis, and Banco Santander acquired ABN AMRO after defeating a rival bid from the UK bank Barclays. The Belgian/Dutch banking group Fortis participated in the consortium with a view to using the ABN AMRO brand name for its retail banking operations in the Netherlands. This deal was the largest acquisition in banking history, but it has since become symbolic of a state of hubris in the higher echelons of the banking industry immediately prior to the credit crisis. Neighbouring Belgium also has a highly concentrated banking sector, with $CR_5 = 85$ percent in 2006. The three largest banks in 2006 were Fortis, Dexia, and the KBC Group. There is considerable overlap between the operations of banks in Belgium and the Netherlands. The latter hosts a number of foreign banks that transact business with international companies quoted on the Amsterdam stock market. In October 2008, the Dutch government injected €10 billion into ING. This followed the establishment of a €20 billion fund to protect the financial sector during the credit crisis. The Belgian government injected €4.7 billion into Fortis, thereby acquiring a 49 percent stake, and guaranteed all new financing by banks for one year.

The third Benelux country, Luxembourg, is a major financial centre, and a leading location for Europe's offshore mutual fund sector. Luxembourg also hosts a substantial private banking and wealth management sector, and most Luxembourg banks are foreign-owned. Unusually, the banking sector is Luxembourg's largest employer. The structure of the Luxembourg banking sector is quite dissimilar to those of Belgium and the Netherlands, with $CR_5 = 29$ percent in 2006. The three largest banks in 2006 were Fortis Banque Luxembourg, HVB Luxembourg, and Société Générale Bank et Trust. In September 2008, the Belgian, French, and Luxembourg governments invested €6.4 billion to keep Dexia solvent.

Although concentration in the banking sector as a whole in Ireland is relatively low, with $CR_5 = 45$ percent in 2006, retail banking is dominated by Ireland's two largest banks, Allied Irish Banks and Bank of Ireland. Their combined market share in personal current accounts was more than 70 percent in 2006. The competition authorities have expressed concern about a lack of competition in retail banking. A 2005 Competition Authority report identified, and introduced measures intended to remedy, several anti-competitive issues, in personal current accounts (consumer lock-in and switching costs, high entry barriers, and lack of new competition), and in small business lending (market power evidenced by price stickiness in response to changes in official interest rates). In October 2008, the Irish government introduced measures to guarantee €440 billion against the liabilities of the country's six largest banks. In December 2008, the government set aside €10 billion for the recapitalization of major banks.

The financial system in Austria is predominantly bank-based. The banking sector comprises several types of bank, including joint stock and private banks, savings banks, rural credit cooperatives, industrial credit cooperatives, state mortgage banks, building societies, and special purpose banks. However, the historical distinction between these categories has diminished over time, as Austria has evolved toward a universal banking model. Concentration is moderate, with $CR_5 = 44$ percent in 2006. The three largest banks in 2006 were Erste Bank, Bank Austria Creditanstalt, and Raiffeisen Zentralbank Osterreich. Despite some consolidation during the 2000s, by European Union standards the number of banks in Austria remains large relative to its population size. In October 2008, the Austrian government announced that it would provide support of up to €90 billion to the banking sector. The plan consists of a state guarantee for interbank loans capped at €75bn, and a further guarantee for liabilities and assets capped at €15bn.

Banks dominate the financial sector in Portugal. Many banks are highly diversified into insurance, securities, and other non-banking activities. Private commercial banks control approximately 60 percent of banking sector assets. Concentration in the Portuguese banking sector is relatively high, with $CR_5 = 68$ percent in 2006. The leading domestic banks are the state-owned Caixa Geral de Depósitos, Millennium bcp, and Banco Espírito Santo Group. Foreign-owned banks accounted for around 20 percent of banking sector assets in 2006. Conservative lending practices and absence of a real estate boom have limited the exposures of Portuguese banks during the credit crisis. Nevertheless, the Portuguese government set up a €20 billion loan-guarantee fund in October 2008, and in December 2008, announced the availability of up to €5 billion for bank recapitalization through the purchase of preference shares.

Finally, and in common with several other smaller European Union countries, the banking sector in Greece is highly concentrated, with $CR_5 = 66$ percent in 2006. The largest banks in 2006 were National Bank of Greece, EFG Eurobank Ergasius, and Alpha Bank. Two other important players in the Greek banking

sector are the special credit institutions, the Postal Savings Bank and the Deposit and Loans Fund. State-owned banks accounted for around 15 percent of banking sector assets in 2006, but the trend has been toward reduced state ownership and increased foreign ownership. The acquisition of the state-owned Emporiki Bank by the French bank Crédit Agricole in 2005 has contributed toward this trend: Emporiki was the fifth largest Greek bank by asset size in 2006. The Greek government announced a €28 billion banking sector stabilization program in October 2008, but by early December no banks had availed themselves of this support.

Structural change and bank performance

Table 32.7 reports structural indicators for the banking sectors of the European Union-15 countries. Between 1985 and 2006, the total number of banks operating in the European Union-15 fell from 12,473 to 6,296. In France, Germany, Italy, Spain, and the UK combined, the increase in nominal total assets over this period was in excess of 400 percent. Across the European Union-15 countries as a whole, rapid and sustained growth in banking sector activity reflects the effectiveness of deregulation and the single market programme in liberalizing formerly 'repressed' banking systems (Dermine, 2006).

In France, Germany, Italy, and Spain, the number of bank branches increased between 1985 and 2006, while the UK experienced a decline. Total European Union-15 banking sector employment increased by more than 10 percent over the same period, reaching a total of 2.7 million in 2006. Again, there were large differences between countries: the UK and (to a lesser extent) Germany experienced significant employment growth, while there was little change in France, Italy, and Spain.

There have been substantial variations in the average profitability of banks located in different European Union countries. Table 32.8 highlights the relatively low profitability of German banks during the early and mid-2000s, when banks in Belgium, Sweden, and UK enjoyed relatively high average profitability. Differences in average profitability between countries have been attributed to the following: variation in accounting and tax systems; structural factors such as the intensity of competition in specific product segments; the extent of product and geographic diversification; and business cycle effects (Goddard, Molyneux, and Wilson, 2004a; Goddard, Molyneux, and Wilson, 2004b; Llewellyn, 2005; Casu and Girardone, 2006; and Carbó and Rodriguez., 2007).

Table 32.7. Structural indicators for EU-15 banking sectors

Country	Number of banks			Assets (billion euros)			Number of branches			Employees (000s)			Concentration (Assets CR_5)		
	1985	1995	2006	1985	1995	2006	1985	1995	2006	1985	1995	2006	1985	1995	2006
Austria	1,406	1,041	809	–	–	789.8	–	–	4,258	–	–	76	–	39	43.8
Belgium	120	143	105	285.9	589.4	1,121.9	8,207	7,668	4,574	71	77	68	48	54	84.4
Denmark	259	202	191	96.3	125.5	822.4	3,411	2,215	2,144	52	47	46	61	72	64.7
Finland	498	381	361	–	–	255.1	–	1,612	1,598	–	31	24	–	70.6	82.3
France	1,952	1,469	829	1,348.8	2,513.7	5,728.1	25,782	26,606	40,013	449	408	435	46	41.3	52.3
Germany	4,739	3,785	2,050	1,495.1	3,584.1	7,122.8	39,925	44,012	40,282	591	724	693	–	16.7	22.0
Greece	41	53	62	69.2	94.0	315.1	1,815	2,417	3,699	27	54	62	80.6	75.7	66.3
Ireland	42	56	78	21.0	45.8	1,186.2	–	808	935	–	–	39	47.5	44.4	45.0
Italy	1,101	970	807	546.8	1,070.5	2,793.2	13,033	20,839	32,337	319	337	340	–	32.4	26.3
Luxembourg	177	220	154	169.8	445.5	839.6	120	224	234	10	19	25	26.8	21.2	29.1
Netherlands	178	102	345	226.7	650.0	1,873.1	6,868	6,729	3,456	92	111	117	72.9	76.1	85.1
Portugal	226	233	178	38.0	116.3	397.1	1,494	3401	5,618	59	60	58	61	74	67.9
Spain	364	506	352	311.3	696.3	2,515.5	32,503	36,405	43,691	244	249	262	35.1	47.3	40.4
Sweden	598	249	204	–	–	773.7	–	–	2,004	–	–	47	–	–	57.8
UK	772	564	401	1,293.6	1,999.5	9,651.5	22,224	17,522	12,880	350	383	483	–	28.3	35.9

Notes: The large increase in branches in France between 1995 and 2006 reflects the establishment of La Banque Postale in 2005, when approximately 12,000 former post office branches were converted into bank branches.

Source: Central Bank Reports, var.; ECB (2000; Goddard, et al. (2001: chap. 2); ECB (2006a), ECB (2007a).

Table 32.8. Average profitability (percentage return on equity) of EU-15 national banking sectors, 1990–2006

Country	1990–94	1995–99	2000–06	2000	2001	2002	2003	2004	2005	2006
Belgium	9.57	14.54	16.40	20.48	15.90	11.76	16.07	14.03	17.11	19.46
Germany	12.97	12.48	3.70	7.86	0.84	−1.71	−2.70	2.26	8.33	11.02
Greece	24.60	21.16	12.72	19.21	11.80	7.71	14.01	11.54	10.86	13.93
Spain	9.73	10.40	12.49	10.37	12.30	12.65	13.35	14.60	8.94	15.22
France	6.18	7.36	11.43	12.08	10.94	9.38	9.85	13.43	9.54	14.77
Ireland	n/a	19.80	14.81	17.88	10.77	11.90	14.50	18.30	12.46	17.84
Italy	11.14	9.29	10.02	17.58	8.42	6.44	7.59	11.45	8.17	10.50
Luxembourg	12.73	21.87	14.23	20.51	12.89	10.62	13.73	9.88	12.79	19.22
Netherlands	13.99	15.92	14.95	17.19	12.39	9.75	14.73	19.50	14.14	16.96
Austria	8.13	9.17	10.60	11.33	7.85	7.83	9.50	10.49	10.91	16.31
Portugal	10.07	7.78	11.55	8.84	13.43	12.30	13.44	11.40	8.04	13.40
Finland	−21.57	8.05	14.54	22.07	22.79	8.40	18.11	12.12	7.36	10.92
Sweden	17.09	18.42	16.04	19.50	18.85	13.39	15.34	18.45	11.07	15.70
UK	15.40	27.88	15.26	21.49	13.47	11.59	14.43	19.90	9.84	16.10
Denmark	−2.77	15.70	13.98	15.24	10.23	11.26	15.75	16.46	12.06	16.84
EU-15	9.05	14.66	12.09	–	–	–	–	–	–	–

Source: *Bankscope*.

As Table 32.7 shows, banking sector concentration ratios for markets defined by national boundaries have increased for some European Union countries and decreased for others, and there has been no consistent long-term trend. As far as the authors are aware, no regular published data are available on concentration at the level of the European Union-15 (or any alternative pan-European level). The non-availability of these data reflects the difficulties involved in either aggregating the unconsolidated company accounts data of dozens or hundreds of subsidiaries of the largest pan-European banking groups, or deconstructing the consolidated accounts of the same groups, which operate both within and beyond the European Union. It is clear that any concentration ratio calculated for the European Union-15 would be considerably smaller than the average of the same concentration ratio calculated for the fifteen countries individually. For example, Papademos (2005) estimates that the fourteen largest banking groups accounted for around one-third of total European Union bank assets in 2005. In view of the cross-border merger trend during the 2000s (see below), it seems probable that the long-term trend in European Union-level banking sector concentration has been upward.

Whatever the trends in concentration at either national or European Union levels, competition in European banking appears to have become more intense during the 2000s, as barriers to product and geographic diversification have been reduced or eliminated. Competition has prompted banks to make efficiency savings. Interest margins have come under pressure, encouraging many banks to develop fee-paying and commission-paying services. Many banks have diversified into non-traditional activities such as insurance and mutual funds, private banking, and asset management. Prior to the credit crisis of 2007–8, the securitization of European banks' loans portfolios had been proceeding rapidly. Securitization issues amounted to €496 billion in 2007, of which over 50 percent related to residential mortgages (European Securitization Forum, 2008). Meanwhile, insurance companies and investment and pension funds have encroached into territory previously occupied by banks, as household savings have been siphoned toward alternative savings and investment products. By the end of the 2000s, non-bank institutions such as supermarkets and telecommunications firms competed in financial services markets.

As elsewhere, advances in technology have impacted on the economics of European banking. Relevant aspects of technological change include innovations that reduce costs associated with the collection, storage, processing, and transmission of data, and those that transform the means whereby customers gain access to banking services and products. Significant front-office innovations are reflected in the growth in number and usage of automated teller machines (ATMs), electronic funds transfer at the point of sale (EFTPOS), Internet banking and e-money services. Meanwhile, back-office operations have been transformed by the adoption of new internal systems, such as customer relationship management and

business management technologies, core processing technologies, and various support and integration technologies.

As noted above, consolidation has been a prominent feature of European banking during the 2000s. M&A within national borders may present more straightforward opportunities for realizing cost savings or efficiency gains than cross-border M&A and, perhaps, fewer complications if the corporate cultures of the merger partners are homogeneous (Buch and DeLong, 2009). However, it is likely that increased emphasis will be placed on cross-border M&A in the future, as domestic banking markets become increasingly congested. Table 32.9 lists selected large bank cross-border mergers that took place between 1997 and 2007. The growth of the cross-border M&A movement suggests that there has been a reduction of several barriers that may have been insurmountable in the past, including difficulties in selling generic products across borders; differences in competition, employment, regulatory, and supervisory policy; political interference; and a lack of consumer trust in foreign banks.

An inclination for the competition authorities in some countries to disqualify mergers between banks that already dominate their domestic retail markets may have the unintended consequence of promoting cross-border bank M&A. For example, the UK's Competition Commission had indicated that mergers between the four largest retail banks (HSBC, Barclays, Lloyds TSB, and RBS) would be discouraged. Accordingly, these banks tended to look outside the UK for feasible acquisition targets. The bidding war between Barclays and a consortium including RBS to acquire the Dutch bank ABN AMRO in 2007 was a case in point. On the other hand, national chauvinism on the part of some European governments, unwilling to sanction the foreign acquisition of important domestic banks, may present an obstacle to cross-border consolidation in certain cases.

Integration of the European banking industry

Table 32.10 summarizes the key legislative changes at the European Union level that have contributed toward the integration of European banking and financial markets since the late 1970s. In 1989, the Second Banking Coordination Directive established European Union-wide recognition of single banking 'passports' issued in any member state, as well as the principle of home country supervision with minimum standards (including capital) at the European Union level. In addition, the Directive permitted banks to operate as universal banks—that is, to engage directly in business other than deposit taking and lending, including insurance,

Table 32.9. Selected cross-border mergers between EU-15 banks, 1997–2007

Year	Acquirer (country)	Target (country)	Value euro (millions)
1997	ING (Netherlands)	Banque Brussels Lambert (Belgium)	4,090
1997	Nordbanken (Sweden)	Merita (Finland)	4,135
1999	MeritaNordbanken (Finland)	Chrstiniana Bank og Kreditkasse (Norway)	3,302
2000	Nordic Baltic Holding Company (Sweden)	Unidanmark (Denmark)	4,779
2000	HSBC Holdings (UK)	Crédit Commercial de France (France)	11,229
2001	HVB (Germany)	Bank Austria CA (Austria)	7,169
2004	Banco Santander Central Hispano (Spain)	Abbey (UK)	13,853
2005	Unicredit (IT)	HVB (Germany)	15,371
2005	ABN Amro (Netherlands)	Banca Antoveneta (Italy)	5,865
2005	Unicredit (IT)	Bank Austria Creditanstalt (Austria)	2,695
2006	Crédit Agricole (France)	Emporiki (Greece)	3,100
2006	BNP Paribas (France)	Banca Nazionale del Lavoro (Italy)	9,000
2007	RBS (UK), Fortis (Belgium / Netherlands), Banco Santander (Spain)	ABN AMRO (Netherlands)	71,000

Source: PriceWaterhouseCoopers (2006) European banking consolidation (London: PWC) pp.-16, and authors' updates.

securities business, factoring, and leasing. Insurance companies and investment companies were likewise granted a single European Union 'passport' with mutual recognition through Directives enacted in the early 1990s.

Other major developments that have impacted significantly on the banking and financial services sectors include the 1985 White Paper on the Completion of the Internal Market; the 1986 Single European Act; the 1992 Maastricht Treaty (which consolidated the single market program); the introduction of the euro in 1999; and the accession of twelve new member countries in 2004 and 2007, increasing total European Union membership from fifteen to twenty-seven.

The Financial Services Action Plan (FSAP), launched in 1999, aimed to promote a more competitive and dynamic financial services industry with improved regulation. The FSAP specified measures to achieve three strategic objectives: (i) establishing a single market in wholesale financial services; (ii) making retail markets open and secure; and (iii) strengthening the rules on prudential supervision. A range of other regulatory actions focused on harmonizing the fiscal treatment of financial services. By the start of 2005, thirty-eight of the original forty-two measures outlined in the FSAP had been implemented and were incorporated into European Union law.

Table 32.10. Legislation affecting the EU banking industry

1977	First Banking Directive. Removed obstacles to the provision of services and establishment of branches across the borders of EU member states. Harmonized rules for bank licensing. Established EU-wide supervisory arrangements.
1988	Basel Capital Adequacy Regulation (Basel I). Minimum capital adequacy requirements for banks (8% ratio). Capital definitions: Tier 1 (equity); Tier 2 (near-equity). Risk weightings based on credit risk for bank business.
1988	Directive on Liberalization of Capital Flows. Free cross-border capital flows, with safeguards for countries with balance of payments problems.
1989	Second Banking Directive. Single EU banking license. Principles of home country control (home regulators have ultimate supervisory authority for the foreign activity of their banks) and mutual recognition (EU bank regulators recognize equivalence of their regulations). Passed in conjunction with the Own Funds and Solvency Directives, incorporating capital adequacy requirements similar to Basel I into EU law.
1992	Large Exposures Directive. Banks should not commit more than 25% of their own funds to a single investment. Total resources allocated to a single investment should not exceed 800% of own funds.
1993	Investment Services Directive. Legislative framework for investment firms and securities markets, providing for a single passport for investment services.
1994	Directive on Deposit Guarantee Schemes. Minimum guaranteed investor protection in the event of bank failure.
1999	Financial Services Action Plan (FSAP). Legislative framework for the Single Market in financial services.
2000	Consolidated Banking Directive. Consolidation of previous banking regulation.
2000	Directive on e-money. Access by non-credit institutions to the business of e-money issuance. Harmonized rules/standards relating to payments by mobile telephone, transport cards, and Basel payment facilities.
2001	Directive on the Reorganization and Winding-Up of Credit Institutions. Recognition throughout EU of reorganization measures/winding-up proceedings by the home state of an EU credit institution.
2001	Regulation on the European Company Statute. Standard rules for company formation throughout the EU.
2002	Financial Conglomerates Directive. Supervision framework for a group of financial entities engaged in cross-sectoral activities (banking, insurance, securities).
2004	New EU Takeover Directive. Common framework for cross-border takeover bids.
2005–10	White paper on Financial Services Policy. Plan to implement outstanding FSAP measures, consolidation/convergence of financial services regulation and supervision.
2006	Directive on the taking up and pursuit of the business of credit institutions. Recast version of the Consolidated Banking Directive.
2007	Payment Services Directive. Guarantees fair and open access to payments markets and to increase consumer protection.
2008	Capital Requirements Directive. Updates Basel I and incorporates the measures suggest in the *International Convergence of Capital Measurement and Capital Standards* (Basel II). Improved consistency of international capital regulations. Improved risk sensitivity of regulatory capital. Promotion of improved risk management practices among international banks.

Source: ECB (2005b: Table 2) and authors' updates.

Despite a sustained legislative drive at the European Union level toward the integration and harmonization of financial markets, by the end of 2008, significant barriers still remained to the integration of European retail banking markets (including banking services to small business). Retail banking is heavily segmented by national boundaries, and it is problematic to undertake cross-border activity without (physical) establishment. Regulation remains predominantly country-specific, reflecting differences in tax treatment, consumer protection legislation, marketing practice, product definitions, and investor protection. Issues of trust and confidence still incline many customers to prefer local or national banks to foreign banks.

Furthermore, local banks may have superior access to private information about a borrower's creditworthiness, creating a rent that is unavailable to a competing foreign bank. Bundling of financial services, enabling different prices to be charged for each component of the bundle in different markets, also can confer advantages on local banks. Barriers to cross-border activity have been less onerous for activities with an international dimension, such as investment banking. However, most of the world's largest investment banks have extensive physical market presence in many countries, suggesting that cross-border provision of financial services without establishment is not the preferred model, even in wholesale banking.

Quantification of the extent to which the integration of national European Union banking markets has been achieved is difficult (Baele, et al., 2004). Indicators that have been proposed include the number or asset value of foreign banks; product price convergence; quantity indicators highlighting foreign bank shares in specific product lines or payment systems; the presence of subsidiaries or branches in other countries; and cross-border M&A. Berger (2007) notes that, in the mid-2000s, foreign banks accounted for around 15 percent of banking sector assets in the European Union-15, but around 70 percent in the ten 2004 accession countries.

European Central Bank (2007a) summarizes the position as follows. Foreign bank subsidiaries and branches held only 17.9 percent of total assets in the euro area in 2006, and there were major differences between countries. For example, foreign banks accounted for 11 percent of bank assets in France, and 95 percent in Luxembourg. The share of subsidiaries in total banking assets in the European Union increased between 2002 and 2006, while those of branches was little changed. Prior to 2006 there had been an upward trend in the value of cross-border bank mergers, with large banks increasingly looking beyond their home countries for major expansion opportunities. Cross-border transactions accounted for a rising share of total wholesale banking transactions. The number of retail payments systems operated in European Union countries declined, from nineteen in 1998 to fourteen in 2006. Over the same period, the number of automated clearing houses was reduced by one, from seven to six. Even in retail payment systems the pace of integration has been somewhat modest.

Supervision and financial stability

The emergence of international banking groups within the European Union raises a number of important issues for bank regulation and supervision. Meanwhile, the turbulence which first impacted on international financial markets during the summer of 2007 represents the most serious challenge for bank supervision, and for the preservation of financial stability, since the establishment of the European Central Bank (European Central Bank) in 1998 and the launch of the euro at the start of the following year.

Supervision of pan-European banking groups

The integration of European banking markets has seen the emergence of several large cross-border banking groups. Among these groups, organizational structures, risk management practices, and strategic planning functions have evolved to deal with pan-European activity in traditional commercial banking as well as treasury and trading business. The emergence of pan-European banking groups gives rise to new forms of systemic risk, as banks become both too large and 'too complex to fail' (Herring and Carmassi, 2008; and Herring and Carmassi, 2009). The growth of pan-European banks places strains on a supervisory framework traditionally organized around national regulatory authorities. Some countries have centralized supervision across banking, securities and insurance, while, in others, different regulatory arrangements have evolved for various financial services (Garcia and Nieto, 2007). There is considerable variation in the duties and responsibilities of national central banks.

In sharp contrast to the nation-orientated architectures for safeguarding financial stability and the decentralized decision-making processes for allocating resources, cross-border European finance is continuing to grow and to become more complex and opaque. Moreover, a European financial system is gradually becoming a reality, with pan-European markets and the emergence of regional and European institutions. Driving the debate in Europe is a growing recognition by policymakers and politicians that, along with substantial benefits, the emergence of a European financial system will most probably be accompanied by a greater propensity for market turbulence, cross-border contagion, and regional and European systemic risk (Nieto and Schinasi, 2007: 3).

In response to these challenges, the Lamfalussy approach to the regulation of securities markets (named after the chair of a committee which reported in 2001) was extended to banking, and completed in 2004. The Lamfalussy approach split legislation into high-level framework provisions, and implementing measures which could be modified flexibly to keep pace with market and supervisory developments. Open consultation procedures and greater transparency were key

features, and there was regular contact between the Committee of Banking Supervisors (CEBS), Committee of European Securities Regulators (CESR), and the Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS). In 2006, the Financial Services Committee (FSC), comprising finance ministry representatives from each European Union member state and observers from CEBS, CESR, and CEIOPS, produced the Francq Report, which recommended improvements to supervisory arrangements within the European Union through convergence in supervisory practices, improved cost-efficiency of supervision, and improved cross-border supervision.

The introduction of Basel II in the European Union via the Capital Adequacy Directive (CAD3) in 2006, requiring greater transparency in the system of regulation and supervision, has increased the pressure for the development of a consistent European Union-wide regulatory and supervisory approach. By 2008, twenty-three European Union member states had committed to a common supervisory framework developed by CEBS, under which standardized information about national supervision practices was published. CEBS has also produced guidelines on standardized financial reporting for cross-border banking groups. In order to gauge market developments in bank lending, a quarterly survey of bank lending in the euro area addresses various issues, focusing particularly on credit standards for approving loans to firms and households.

There is no definitive answer to the question as to whether the supervision of international banks should be the responsibility of the home or the host country. Home country supervision may create conflicts of interest between countries if, for example, a decision is taken to close a bank that operates in several other countries. Small countries might find it difficult to bear the costs of bailing out a large international bank. Host country supervision creates difficulties for international banks in complying with standards in various countries, and encourages the creation of complex organizations with subsidiaries, making closure difficult in the event of bankruptcy proceedings under different jurisdictions.

The legal structure a financial institution adopts determines the division of responsibilities among national supervisory authorities. Since subsidiaries are independent legal entities owned by a parent bank, for regulatory and supervisory purposes they are treated separately from the parent. In contrast, branches form part of the parent bank. Accordingly, home authorities supervise the activities of foreign branches, while host authorities supervise subsidiaries. CEBS guidelines on voluntary supervisory cooperation between home and host authorities envisaged operational networking between the supervisors responsible for different parts of cross-border banking groups, and eventual convergence in supervisory practice.

Proposals to overhaul supervision include some that favor centralized supervision that tackles the risks associated with cross-border banking head-on, and others that favor localized supervision that accounts for local economic conditions. Given the differences in the size and organizational and ownership structures of

European Union banks, a hybrid approach seems likely to emerge. Čihák and Deceressin (2007) suggest the creation of a European Banking Charter to simplify the supervision of large pan-European banks. Mayes, Nieto, and Wall (2007) argue for the introduction of a Prompt Correction Action framework. Fonteyne and van der Vossen (2007) compile a list of proposals, including: (i) enhancing existing supervisory arrangements via colleges of national supervisors; (ii) designating a lead supervisor (e.g. the home country authority) to oversee both subsidiaries and branches throughout the European Union; (iii) establishing an European Union-level supervisory framework to oversee pan-European banking groups alongside existing national frameworks; and (iv) creating a European system of bank supervisors, operating with consolidated rules, but with some responsibilities devolved to the national level. The principle of subsidiarity, whereby action should only be taken at European Union level if national or local action is ineffective, is a key argument in favor of the retention of national supervisory arrangements. Another is the recognition that only a small number of European Union banks are truly international, while there remains considerable diversity at national, regional and local levels, especially in retail banking.

The credit crisis has demonstrated the dangers of cross-border contagion, although the policy response has mostly been through bailouts of banks by individual nation states. This perhaps is not surprising as no formal system is in place to allow the sharing of costs of recapitalizing large cross-border banks. Goodhart and Schoenmaker (2009) argue that *ex post* negotiations on sharing the costs of a cross-border bank bailout will result in an under provision of capital as participants outside the banks host country have an incentive to understate involvement in order to minimize costs. This leaves the home country supervisor with the burden. As an alternative, they propose two *ex ante* burden-sharing mechanisms. The first, termed a 'general mechanism', is financed collectively by the participating countries based on their GDP. The second is a specific mechanism funded by participants based upon the geographical spread of their banks' business. The latter might be more effective, because each country's benefits (e.g., financial stability) are more closely aligned to the costs of contributing to the scheme.

The credit crisis: turbulence in European financial markets in 2007–8[2]

The turbulence which first struck international financial markets in the summer of 2007, when sharp losses in the value of US subprime mortgages triggered a collapse

[2] Goddard et al. (2009) provide an account of the financial crisis in Western Europe, primarily from a country-level and banking sector perspective. They examine measures enacted by governments and central banks to deal with impaired bank assets, insufficient capital, and liquidity. Reform proposals aimed at creating a more secure and stable European financial system are examined.

of investor confidence in many types of asset-backed security, has posed fundamental challenges for bank supervision within the European Union (and beyond). Asset-backed securities are bonds backed by prospective cash flows from revenue sources such as mortgages and credit card debt. The securitization movement got off to a relatively slow start in Europe in comparison with the US. However, European markets for asset-backed securities grew rapidly following the launch of the euro in 1999, assisted by the general movement toward financial liberalization throughout the European Union. By unlocking future revenue streams through the issuance of asset-backed securities, banks were able to raise new funding and liquidity to support additional lending.

By issuing such securities through separately constituted off-balance sheet structured investment vehicles (SIVs), many banks were able to economize on costly regulatory capital requirements. The small German lender IKB Deutsche Industriebank, an early casualty of the credit crisis, was a case in point. Between 2002 and 2007, IKB accumulated a €12.7 billion portfolio of asset-backed securities, held off-balance sheet by its SIV, Rhineland Funding. Around 70 percent of the portfolio had a credit rating of AA or above; however, the same credit rating agencies that had provided these assessments had been heavily involved in structuring and valuing Rhineland Funding's portfolio (*Economist*, 9 August 2007).

Following the first downgrades of asset-backed securities backed by pools of subprime mortgages in the US, spreads on asset-backed securities backed by corporate bonds, bank loans, credit cards, and auto loans increased substantially, and the issuance of new asset-backed securities declined sharply. The repricing of credit and liquidity risk in secondary markets also impacted upon supply and demand conditions in primary markets, with sharp increases in spreads, and heightened nervousness on the part of both borrowers and lenders.

In August 2007, Rhineland Funding called on a €12 billion line of credit promised by IKB and several other banks. One of these, Deutsche Bank, exercised an option to cancel the commitment. This prompted a rescue in which the state-owned Kreditanstalt fur Wiederaufbau (KfW), a major shareholder in IKB, provided liquidity support and wrote off substantial losses on IKB's loans portfolio (*Economist*, 9 August 2007).

As liquidity within the banking system evaporated, and overnight interest rates increased dramatically on 9 August 2007, the European Central Bank (European Central Bank) moved rapidly to sanction an unlimited offer of one-day loans at the then-current policy rate of 4 percent. The European Central Bank was favorably positioned to take this step, because its rules enabled it to accept a wider range of assets as collateral than those available to the Federal Reserve or the Bank of England, including some asset-backed securities. European Central Bank rules permitted the acceptance of any tranche of securities that is most senior and rated 'A-' or above. In part, this liberal regime originated from the consolidation of a diverse set of national regulations that were in force at the time the European Central Bank was established. However, concerns were raised that the nature of the

European Central Bank's intervention may have inhibited the re-emergence of a publicly traded market in asset-backed securities, and may have left the European Central Bank, or ultimately the euro zone taxpayer, vulnerable to a future wave of defaults on asset-backed securities (*Economist*, 5 June 2008; 12 June 2008).

In the short term, this rapid injection of liquidity into the euro-area banking system appeared to have been effective in averting the danger of a major bank failure. However, the outlook deteriorated dramatically in September 2008 following the collapse of the US investment bank Lehman Brothers, and the agreed takeover of another US investment bank, Merrill Lynch, by Bank of America. The Federal Reserve then announced an $85 billion rescue package for the large insurer AIG, in return for an 80 percent stake. The US bank Washington Mutual was closed and its remaining assets sold to JPMorgan Chase. The US Federal Housing Finance Agency nationalized the government-sponsored mortgage finance enterprises, Federal National Mortgage Association (Fannie Mae) and Federal Home Loan Mortgage Corporation (Freddie Mac). These events created a crisis of confidence that appears to have brought the US and European banking systems perilously close to the brink of collapse in September and October 2008. In response, every European government has announced a combination of loan-guarantee schemes, bank-rescue plans, and fiscal-stimulus packages in an attempt to preserve their banking systems and avert the most damaging recessionary consequences of the banking crisis.

As one would expect, the profitability of European banks has been severely strained by the credit crisis. From the start of 2007 to October 2008, the largest European banks reported credit losses and write-offs amounting to $252 billion, and received capital injections totalling $290 billion, the bulk of the latter taking place during 2008 (see Table 32.11). Since October 2008, the capitalization of many of Europe's largest banks has been further bolstered through injections of public funding, predominantly through government purchase of preference shares. The UK experience is highlighted in Box 32.1.

The extreme difficulties experienced by European banks since 2007 have focused attention on those regulatory systems that may have coped better than others in the run-up to the crisis. Two distinctive features of the Bank of Spain's regulatory approach have attracted particular scrutiny. The first aspect concerns the 'dynamic provisioning' regime introduced in 2000, requiring banks to harmonize loan–loss provisioning with the lending cycle. The second is the requirement that assets channeled through SIVs are subject to the same capital requirements as on-balance sheet assets. With no opportunities for regulatory arbitrage, most Spanish banks simply abstained from creating SIVs, thereby avoiding the trap that snared banks elsewhere, of creating an overly optimistic impression of the extent to which credit risk was transferred through the creation of asset-backed securities issued through SIVs (*Economist*, 15 May 2008). As noted above, Spanish banks have not been immune from the credit crisis. However, the €100 billion support package

Table 32.11. European bank write-downs and capital raised until October 2008

Rank	Bank	Country	Write-down and losses (US$ billion)	Capital raised (US$ billion)
1	UBS	Switzerland	48.6	30.7
2	HSBC	UK	33.1	4.9
3	RBS	UK	15.4	49.3
4	Credit Suisse	Switzerland	13.2	11.4
5	IKB Deutsche Industriebank AG	Germany	13.0	10.8
6	Deutsche Bank	Germany	11.8	5.8
7	ING Group	Netherlands	9.5	16.9
8	HBOS	UK	9.4	23.3
9	Crédit Agricole	France	8.7	11.3
10	Fortis	Belgium	8.3	20.4
11	Société Générale	France	7.6	10.5
12	Barclays	UK	6.6	26.6
13	Bayerische Landesbank	Germany	6.2	8.1
14	BNP Paribas	France	5.4	3.2
15	Hypo Real Estate Holding AG	Germany	5.1	0.0
16	Dresdner Bank AG	Germany	4.6	0.0
17	Landesbank Baden-Wurttemberg	Germany	4.2	0.0
18	Unicredit SpA	Italy	4.0	9.2
19	Lloyds TSB Group Plc	UK	3.7	13.2
20	KBC Groep NV	Belgium	3.5	4.5
21	HSH Nordbank AG	Germany	3.1	1.6
22	Rabobank	Netherlands	3.1	0.0
23	West LB AG	Germany	3.1	6.4
24	Commerzbank AG	Germany	2.8	10.4
25	DZ Bank AG	Germany	2.3	0.0
26	Landesbank Saschen AG	Germany	2.2	0.0
27	ABN AMRO Holding NV	Netherlands	2.0	0.0
28	Dexia SA	Belgium	1.5	8.1
29	Alliance and Leicester Plc	UK	1.1	0.0
30	Groupe Caisse d'Epargne	France	1.0	0.0
31	Other European Banks (not listed above)		8.1	3.8

All the charges stem from the collapse of the US subprime mortgage market, and reflect credit losses or write-downs of mortgage assets that are not subprime, as well as charges taken on leveraged loan commitments since the beginning of 2007. They are net of financial hedges the firms used to mitigate losses and pre-tax figures unless the bank only provided after-tax numbers. Credit losses include the increase in the provisions for bad loans, impacted by the rising defaults in mortgage payments. Capital raised includes common stock, preferred shares, subordinated debt, and hybrid securities which count as Tier 1 or Tier 2 capital as well as equity stakes or subsidiaries sold for capital strengthening. Capital data begins with funds raised in July 2007. All numbers are in US $billion, converted at December 2008 exchange rates if reported in another currency.

Source: Bloomberg.

Box 32.1. The Credit Crisis in the UK

In the UK, the onset of the global credit crisis impacted most visibly upon Northern Rock, a former mutual building society that converted to become a shareholder bank in 1997. During the 2000s, Northern Rock had pursued an aggressive growth strategy, with rapid expansion of its mortgage loans portfolio financed heavily by securitization. Northern Rock's share price had been in decline throughout 2007 owing to concerns that its earnings were overly dependent on its ability to continue financing rapid lending growth. In September 2007, as finance from the securitization and covered-bond markets dried up, Northern Rock was forced to apply to the Bank of England for liquidity support, triggering a bank run as depositors queued to make withdrawals from high street branches.

The collapse of Northern Rock exposed several weaknesses in the UK's tripartite bank supervisory and regulatory framework, in which responsibilities are split between the Financial Services Authority (FSA), the Bank of England, and the Treasury. The FSA was widely criticized for failing to recognize the risk posed by the Northern Rock's rapid growth strategy until the crisis struck. During the immediate run-up to the crisis, the Bank of England was unable to provide liquidity support covertly, and the public nature of its intervention exacerbated the markets' loss of confidence. The UK's deposit guarantee scheme provided depositors with insufficient assurance to prevent large-scale withdrawals of deposits. Short of nationalization, the authorities' powers to intervene directly in the running of the bank were seen to be inadequate (House of Commons Treasury Committee, 2008). Full control of Northern Rock was belatedly achieved through nationalization in February 2008, but only after several months spent unsuccessfully attempting to find a purchaser, and following the investment of £30 billion of public funds to provide the guarantees required to restore confidence (*Financial Times*, 2 July 2008).

Subsequently, proposals were developed to address each of these issues. However, concerns for the stability the UK banking sector did not abate during 2008. In June, the share price of Bradford and Bingley, another former UK mutual turned shareholder bank, declined sharply following a mishandled attempt to raise new shareholder capital. In September 2008, concerns about the ability of HBOS to fund its long-term mortgage business precipitated a takeover by its rival Lloyds TSB.

The UK government has responded aggressively to support stability and restore confidence in the financial system (HM Treasury, 2008):

- The government guaranteed all new short- and medium-term debt issues. £100 billion was earmarked to allow banks to swap high-quality illiquid assets for treasury bills.
- The Bank of England reduced its base interest rate to 2 percent by December 2008.
- The coverage of the UK Deposit insurance scheme was extended to bank deposits up to £50,000.
- A Bank Recapitalization Fund was set up, making £25 billion available in permanent capital, and another £25 billion on stand-by. Northern Rock and the mortgage and loan business of Bradford and Bingley were nationalized. Bradford and Bingley's deposits business was sold to Banco Santander. Three of the UK's largest banks: RBS, HBOS, and Lloyds TSB received capital injections in return for public ownership stakes.
- Short-selling of financial stocks was banned until the end of January 2009.
- To assist small and medium-sized enterprises (SMEs) facing credit constraints, a Small Business Finance Scheme supported up to £1 billion of bank lending to small exporters; a £50 m fund was established to convert businesses' debt into equity; and a £25 million regional loan transition fund was created (HM Treasury, 2008).

announced by the Spanish government in October 2008 appears to have been necessitated primarily by the turmoil in interbank markets and the collapse of domestic property prices, rather than the holding of non-performing assets off-balance sheet.

Summing up and going forward

During the 1990s and 2000s, the regulatory drive to create a fully integrated European single market in financial services, together with factors such as technological advance, financial innovation, and deregulation, led to significant changes in the organization and delivery of banking and financial services. However, significant barriers to full European integration still remain, especially in retail banking. The latter remains nationally focused, with relatively little cross-border activity. In contrast, wholesale banking has become highly integrated. Banks have diversified into new product areas and services in order to pool risk, realize economies of scale and scope, and reduce x-inefficiencies. Cross-border M&A and the emergence of international banking groups within Europe raises a number of important issues for bank regulation and supervision. Meanwhile, the credit crisis of the late 2000s presents the most serious challenges for the preservation of financial stability since the creation of the European Central Bank and the launch of the euro.

So what does the future hold? A fundamental reappraisal of current regulatory structures is already under way. Basel I was found not to protect the solvency of banks, and even before implementation was complete Basel II already stood accused of similar failings. At the time of writing there is extensive discussion in European regulatory circles about a 'return to basics', with greater emphasis on: simple leverage and liquidity ratios; the curtailment of opaque business models (including SIVs); and greater scrutiny of financial innovations and capital charges. Consolidation will be a likely consequence of the crisis gripping European (and other) banking systems and this, of course, may have further repercussions in the future for competition, moral hazard ('too big to fail'), and stability. Discussions concerning the creation of a single European bank regulator continue, as national governments discuss methods for organizing coordinated liquidity injections in money markets. There is also debate about the regulation of consolidated groups, and fragility issues. Whatever the eventual outcome of these debates, it seems certain that the current crisis will have a transformative impact on all aspects of the future structure, conduct, and regulation of European banking.

References

Baele, L. M., Ferrando, A., Hordahl, P., Krylova, E., and Monnet, C. (2004). 'Measuring Financial Integration in the Euro Area', European Central Bank Occasional Paper 14.

Bank of England (2008). *Financial Stability Report*, 24.

Berger, A. N. (2007). 'Obstacles to a Global Banking System: "Old Europe" versus "New Europe" ', *Journal of Banking & Finance*, 31: 1955–76.

Buch, C. and DeLong, G. (2008). 'Banking Globalization: International Consolidation and Mergers in Banking', in Berger, A., Molyneux, P., and Wilson, J. O. S., *Oxford Handbook of Banking*. Oxford: Oxford University Press.

Carbó, S., Gardener, E. P. M., and Molyneux, P., (2005). *Financial Exclusion*. Basingstoke: Macmillan.

—— and Rodriguez, F. (2007). 'The Determinants of Bank Margins in European Banking', *Journal of Banking & Finance*, 31: 2043–63.

Casu, B. and Girardone, C. (2006). 'Bank Competition, Concentration and Efficiency in the Single European Market', *Manchester School*, 74: 441–68.

Čihák, M. and Decressin, J. (2007). 'The Case for a European Banking Charter', IMF Working Paper 07/173.

Commission bancaire (2002). *Annual Report*. Paris: Banque de France.

Correa, R. (2008). 'Cross-Border Bank Acquisitions: Is There a Performance Effect?', *International Finance Discussion Papers,* 922. Board of Governors of the Federal Reserve System.

Cruickshank, D. (2000). *Competition in UK Banking: A Report to the Chancellor of the Exchequer*. London: HMSO.

Dermine, J. (2006). 'European Banking Integration: Don't Put the Cart before the Horse', *Financial Institutions, Markets and Money*, 15: 57–106.

European Central Bank (2005). 'Consolidation and Diversification in the Euro Area Banking Sector', *European Central Bank Monthly Bulletin*, May: 79–87.

—— (2006). 'EU Banking Structures'. Frankfurt: ECB.

—— (2007a). 'Financial Integration in Europe'. Frankfurt: ECB.

—— (2007b). 'EU Banking Structures'. Frankfurt: ECB.

European Commission (2005). 'Cross-Border Consolidation in the EU Financial Sector', Commission of the European Communities SEC (2005) 1398.

—— (2007). 'Sector Inquiry under Article 17 of Regulation (EC) No 1/2003 on Retail Banking' (Final Report).

—— (2008). 'The Recapitalisation of Financial Institutions in the Current Financial Crisis: Limitation of Aid to the Minimum Necessary and Safeguards against Undue Distortions of Competition', Communication from the Commission. Brussels, 5.12.2008, C, 8259.

European Securitisation Forum (2008). 'ESF Securitisation Data Report, Winter 2008'. London: European Securitisation Forum.

Fédération Bancaire Française (2005). 'An Overview of Banking in France'. Brussels: Fédération Bancaire Française.

Fonteyne, W. and Van der Vossen, J.-W. (2007). 'Financial Integration and Stability', in Decressin, J., Faruqee, H., and Fonteyne, W. (eds.), *Integrating Europe's Financial Markets*. Washington, DC: International Monetary Fund.

Garcia, G. and Nieto, M. J. (2007). 'Preserving Financial Stability: A Dilemma for the European Union', *Contemporary Economic Policy*, 25: 444–58.

Goddard, J., Molyneux, P., and Wilson, J. O. S. (2001). *European Banking: Efficiency, Technology and Growth*. Chichester: John Wiley.

——————— (2004a). 'The Profitability of European Banks: A Cross-Sectional and Dynamic Panel Analysis', *Manchester School*, 72: 363–81.

——————— (2004b). 'Dynamics of Growth and Profitability in Banking', *Journal of Money, Credit and Banking*, 36: 1069–90.

——————— (2009). 'The Financial Crisis in Europe: Evolution, Policy Responses and Lessons for the Future', *Journal of Financial Regulation and Compliance*, forthcoming.

——————— and Tavakoli, M. (2007). 'European Banking: An Overview', *Journal of Banking & Finance*, 31: 1911–36.

Goodhart, C. E. and Schoenmaker, D. (2009). 'Fiscal Burden Sharing in Cross-Border Financial Crises', *International Journal of Central Banking*, 5: 141–65.

Herring, R. and Carmassi, J. (2008). 'The Structure of Cross-Sector Financial Supervision', *Financial Markets, Institutions & Instruments*, 17, February.

——————— (2009). 'The Corporate Structure of Financial Conglomerates: Complexity and its Implications for Safety and Soundness', in Berger, A., Molyneux, P., and Wilson, J. O. S., *Oxford Handbook of Banking*. Oxford: Oxford University Press.

HM Treasury (2008). 'Pre-Budget Report: Facing Global Challenges: Supporting People through Difficult Times'. London: HM Treasury.

House of Commons Treasury Committee (2008). *Run on the Rock*. London: HMSO.

Llewellyn, D. T. (2005). 'Competition and Profitability in European Banking: Why Are UK Banks So Profitable?', *Economic Notes*, 34: 279–311.

Mayes, D. G., Nieto, M. J., and Wall, L. (2007). 'Multiple Safety Net Regulators and Agency Problems in the EU: Is Prompt Corrective Action A Partial Solution?', Bank of Finland Discussion Paper, 7.

Nieto, M. J. and Schinasi, G. J. (2007). 'EU Framework for Safeguarding Financial Stability: towards An Analytical Benchmark for Assessing its Effectiveness', IMF Working Paper 07/260.

Office of Fair Trading (2008). *Personal Current Accounts in the UK: An OFT Market Study*. London: Office of Fair Trading.

Papademos, L. (2005). 'Banking Supervision and Financial Stability in Europe'. Speech Delivered at European Banking Federation Conference on Supervision of International Banks, Brussels, 28 October 2005.

PriceWaterhouseCoopers (2006). 'European Banking Consolidation'. London: PriceWaterhouseCoopers.

Tieman, A. and Čihák, M. (2007). 'Internationally Active Large Banking Groups', in Decressin, J., Faruqee, H., and Fonteyne, W. (eds.), *Integrating Europe's Financial Markets*. Washington, DC: International Monetary Fund.

CHAPTER 33

BANKING IN TRANSITION COUNTRIES

JOHN P. BONIN
IFTEKHAR HASAN
PAUL WACHTEL[1]

The distinctive character of banking in transition countries

Banking in the transition countries is particularly interesting because banks played no economic role in planned Soviet-style economies while financial sectors in most transition countries are now dominated by banks rather than equity markets. Hence, our first topic in this overview is the emergence of banking sectors from the planned economies. The birthing process was hardly smooth; it took place amidst massive macroeconomic collapse and considerable economic uncertainty. Not surprisingly, these nascent banking sectors experienced crises ranging from serious bad loan problems to total collapse. The next section deals with the responses to the bad loan problem, the process of bank privatization, and the development of the necessary regulatory framework. The following section

[1] The authors are grateful to Rainer Haselmann and Stephan Barisitz for helpful suggestions.

characterizes the structure of the more mature banking sectors in the leading transition countries, with particular attention paid to the dominant role of foreign banks. Banking sectors in many transition economies have developed remarkably quickly and now look little different from their counterparts in other emerging market economies except for the distinctive high percentage of foreign ownership. Nevertheless, these banking sectors are not immune to problems and do not always provide sufficient impetus for economic development, which is problematic because of the bank-dominated financial sectors of most transition economies. Our last section considers the problems of, and prospects, for banks fulfilling this role in the European transition countries. To illustrate the commonalities and differences in the transition experience, we have selected ten representative countries from three regions: Central Eastern Europe (CEE), South Eastern Europe (SEE), and the former Soviet Union. We divide the countries into two groups. The Czech Republic, Hungary, Poland, and Slovakia (CEE) along with Russia, as the largest and most important of the former Soviet republics, make up the first group; Bulgaria, Croatia, Romania, Serbia, and Slovenia (SEE) constitute the second group.

The emergence of banking institutions in the early stages of transition

Banking sectors in the European transition economies were relatively underdeveloped compared with the real economies in these countries due mainly to the legacies of the pre-transition centrally planned economy. As examples of real-sector development, Czechoslovakia had a relatively modern automobile industry, Hungary produced buses, and Bulgaria made computers and software for use within the Soviet bloc. However, in the planning framework, financial intermediation between savers and borrowers was internalized wholly within the state banking apparatus. Capital was allocated through a system of directed credits to state-owned enterprises (SOEs) for both investment needs and budget allocations for the working capital necessary to meet the output plan. Credit evaluation and risk management played no role in lending decisions. The national monobank served only as an accounting clearing house for inter-enterprise transactions. Cash issuances by enterprises were based on planned wage bills that were calibrated to the expected aggregate value of consumer goods sold to households at administered prices. Money was entirely passive in that it was used solely as a unit of account in enterprise transactions and as a medium of exchange between households and the state distribution sector. Household savings, oftentimes the result of forced accumulation of monetary balances owing to the unavailability of desirable consumer

goods to purchase, were collected by a state savings bank that operated an extensive branch network throughout the country.

Pre-transition banking sectors typically included a foreign trade bank that handled all foreign currency transactions to isolate these from the domestic financial system and often contained separate specialty banks to oversee the financing of the agricultural and construction sectors. In this environment, banking was segmented along functional lines and credit allocation was entirely subservient to the plan. Hence, structural segmentation, state control of banking activities, and high concentration ratios are the major legacies inherited from the planning period by the banking sectors in the European transition economies. Despite these commonalities, important differences among the experiences of countries both prior to and during the transition period yield unique characteristics. As an example, we begin with a brief discussion of banking in the SEE transition countries that were former republics of Yugoslavia because their sectors inherited somewhat special legacies. We continue with a consideration of the initial developments in banking during the first half decade of the transition followed by a more-detailed look at several European transition countries. This section concludes with a discussion of foreign bank participation in the early transition years.

In the 1950s, Yugoslavia established a two-tier banking system with a traditional central bank located in Belgrade, the National Bank of Yugoslavia (NBY), and republic-level commercial banks. Banks were owned collectively, as were all enterprises under the Yugoslavian system of self-management. Because Yugoslavia was a small, open economy, commercial banks made a significant number of loans denominated in foreign currency throughout the 1980s. However, these republic-level banks were required to remit most of their foreign exchange deposits to the NBY in exchange for credits in dinars. Hence, the balance sheets of republic-level banks exhibited a serious currency mismatch between assets and liabilities by the late 1980s. Upon the secession of Croatia and Slovenia in 1991, the NBY froze the forex deposits of the republic banks in these two countries creating large holes in their balance sheets. Although the legacies of segmentation and state ownership found in the banking sectors of CEE transition economies were not initially present in Croatia and Slovenia, high concentration ratios and a substantial accumulation of problem loans are important legacies from the Yugoslavian past. Government rehabilitation policies that were designed to deal with bank insolvency led to the nationalization of most banks; hence, state-owned banks were created at the beginning of the transition in Slovenia and Croatia.

The first step in banking sector reform for most transition economies involved the creation of a two-tier system with commercial activities carved out of the portfolio of the national monobank. The top tier consists of a traditional central bank that is charged with pursuing monetary policy, including exchange rate policy, and is given responsibility for supervising and monitoring the nascent banking sector. The second tier consists of the newly created state-owned commercial

banks (SOCBs), the state-owned specialty banks, which themselves morphed into SOCBs, any operating foreign and joint venture banks, and all private domestic banks, including those that entered after the political change. As a rule, lax entry requirements led to the creation of many new private banks, some of which were of dubious quality, or even fraudulent, and virtually all of which were severely undercapitalized. In the former republics of Yugoslavia, this entry occurred well prior to transition, in the late 1970s, when the establishment of many internal company banks led to excessive numbers of small, unhealthy and undercapitalized banks. Hence, the seeds for a banking crisis were planted at the beginning of the transition, or even before, in virtually all transition countries due partly to the adoption of lax entry requirements with the intent of fostering competition for state-owned banks in highly segmented banking sectors. Moreover, the nascent regulatory systems were overwhelmed by the mismatch between their capabilities, which were severely restricted by a lack of human capital, and their mandates provided by quickly adopted standard financial rules and regulations, especially given the inherited loan portfolios of the SOCBs.

Although each country's financial restructuring program involved hiving off the commercial bank portfolio of the national bank to establish the two-tier system, different approaches were taken toward the creation of SOCBs, all of which were established initially as wholly state-owned joint stock entities. In Hungary, the commercial portfolio was divided along sectoral lines—for example, industry, agriculture, and infrastructure, plus the nascent small business sector, to create three SOCBs. In Poland, the commercial portfolio was divided along regional lines to create nine SOCBs from regional offices of the national monobank. The commercial portfolio of the Czechoslovak national monobank was separated into two parts regionally to create two SOCBs—a Czech and a Slovak one. After the Velvet Divorce, each new country had a single, large SOCB. Similarly, in Romania, only one SOCB was created from the entire commercial portfolio of the national monobank. All CEE countries and Russia had specialty banks that obtained universal banking licenses and, thus, became SOCBs after the transition.

At the opposite extreme, full separation of all commercial activities from the Bulgarian national bank's balance sheet occurred in 1990 when each of its 145 branch offices was granted a universal banking license that allowed it to pursue commercial business either as an individual entity or in combination with other branches. Again, the intent of this policy was to foster competition. As a result, fifty-nine SOCBs were formed and, in 1992, the Bank Consolidation Company was established to oversee and orchestrate the eventual consolidation of the Bulgarian banking sector by the government. By 1995, forty-one banks were operating in Bulgaria and the two largest SOCBs were the former state foreign trade bank and the former state savings bank.

In Russia, then the Soviet Union, the two-tier banking system was established in 1987 with the separation of all commercial bank functions from the national

monobank and the creation of sectoral banks by enterprises or former branch ministries. As in Bulgaria, branches of the national bank became independent entities and then regrouped into larger banks. In addition, new entry into Russian banking was dramatic. By 1995, about 2,300 banks were licensed and operating in Russia. Most of the newly created banks were small and poorly capitalized. Some of them were merely internal or house banks owned by industrial enterprises. However, by 1996, six of the 'de novo' domestic private banks had grown sufficiently to be among the ten largest banks in Russia, a group that included the former state foreign trade bank and the former state savings bank as the two largest SOCBs (Abarbanell and Meyendorff, 1997).

Policies toward foreign bank participation, both in establishing subsidiaries and in purchasing equity stakes in SOCBs, differed considerably across the transition countries. In some countries, policies that invited entry—for example, providing tax holidays—encouraged Greenfield foreign operations. In others, licensing was restrictive and foreign banks were limited to taking minority stakes in SOCBs or to participating in the resuscitation of ailing smaller domestic banks. Foreign participation in the banking sector was viewed initially by most governments as a vehicle for importing banking expertise and training to augment the scarce domestic human capital in the sector. Even before the political change, the Hungarian government pursued a liberal licensing policy toward foreign financial institutions. The Central-European International Bank Ltd. was founded as an off-shore joint venture bank by six foreign banks and the Hungarian National Bank in 1979; in 1986, Citibank Budapest Ltd. began operations as a foreign-majority-owned, joint venture bank. By 1995, foreign financial institutions held almost 42 percent of banking assets in Hungary due in large part to the privatization of two SOCBs to foreign owners. The next highest percentage of foreign ownership by the middle of the first decade of transition is found in Slovakia, a country that opened up to foreign bank penetration rapidly after the Velvet Divorce. In contrast, the Czech Republic and Poland restricted new licenses for foreign Greenfield operations and invited foreign owners to take only minority equity positions in existing Czech and Polish banks. These governments followed a more protectionist strategy, taking an infant industry approach according to which domestic banks are nurtured to become strong enough to fend off foreign competition when it arrives. By 1995, only about 16 percent and 4 percent of the banking assets in the Czech Republic and Poland, respectively, were owned by foreign financial institutions. As Table 33.1 indicates, no other major transition country had more than 4 percent of its banking assets held by foreign institutions by the end of 1995.

For the most part, governments in transition countries succeeded in establishing the foundations for building commercial banking sectors early in the transition period. However, developing efficient banking sectors required the completion of three interrelated tasks—namely, the resolution of non-performing loans, the privatization of the SOCBs, and the establishment of effective regulatory

Table 33.1. Banking sector in early transition, 1995

	Number of banks (foreign-owned banks)[1]	Asset share of foreign banks (%)[2]	Credit/Gross domestic product (%)[3]	Nonperforming loans (% of total)[4]	European bank for reconstruction and development index[5]
Central Eastern Europe and Russia					
Czech Republic	55 (23)	15.9	46.7	26.6	3.0
Hungary	43 (21)	41.8	22.7	12.1	3.0
Poland	87 (18)	4.2	12.7	23.9	3.0
Slovakia	33 (18)	32.7	26.3	41.3	2.7
Russia	2,297 (21)	<1	8.7	12.3	2.0
South Eastern Europe					
Bulgaria	41 (3)	<1	21.1	12.5	2.0
Croatia	54 (1)	<1	22.9	12.9	2.7
Romania	24 (1)	<1	7.8	37.9	3.0
Serbia	112 (3)	<1	9.2*	12.0	1.0
Slovenia	39 (6)	3.8	27.3	9.3	3.0

Notes: * indicates data for 1996.

1 The number of banks, with the number of foreign-owned banks in parentheses taken from the country tables in European Bank for Reconstruction and Development, *Transition Report*, var. issues.

2 The asset share of foreign-owned banks in total banking assets from Barisitz (2007), except for Slovenia. Slovenian data are from Bonin (2004).

3 Credit/Gross domestic product is domestic credit to the private sector at year end from Barisitz, except for Slovenia. Slovenian data are from European Bank for Reconstruction and Development, *Transition Report* (2003).

4 Non-performing loans as a percentage of total loans at year end from Barisitz (2007), except for Slovenia. Slovenian data are from European Bank for Reconstruction and Development, *Transition Report* (2003).

5 European Bank for Reconstruction and Development index of banking sector reform from European Bank for Reconstruction and Development, *Transition Report* (2003); it takes values between 1.0 and 4.0+.

institutions. We discuss the progress made on these fronts during the first decade of transition in the next section.

The development of modern banking sectors during the first decade of transition

As described in the previous section, the typical banking sector in a transition economy consisted initially of state-owned banks that were carved out of the planned economy structure along with newly established small private domestic banks. Some countries began to privatize the large SOCBs quickly and also opened up to foreign bank entry early in the transition. However, the creation of market-based legislation and institutions did not lead automatically to good banking practices. To the contrary, the SOCBs and the newly created banks often did not behave like proper commercial banks due to distorted incentives.

First, the SOCBs continued to maintain banking relationships with their large clients—that is, state-owned enterprises (SOEs). Such lending was either politically mandated or simply the result of long-standing relationships between clients having little experience in choosing viable projects and banks unable to evaluate the risk of loans. Second, in many countries, 'de novo' banks were created without adequate regulatory oversight. As a result, some 'de novo' banks were used to channel loans improperly to their owners, many of which were enterprises so that these banks acted as pocket banks for their owners. Entry requirements for 'de novo' domestic banks were initially very lenient because policy was based on the mistaken notion that competition would be enhanced by easy entry. The proliferation of new, often undercapitalized, banks placed an added burden on an underdeveloped regulatory structure. Although most countries adopted modern banking and regulatory legislation immediately, effective supervision did not follow automatically due partially to the scarcity of knowledgeable staff.

Not surprisingly, bad loans were a serious problem for all transition economies due partly to the inherited legacies but also to continuing lending practices. The ratio of non-performing loans to total loans in 1995 averaged 25.9 percent in the four CEE countries (Poland, Hungary, Czech Republic, and Slovenia/Slovakia—see Table 33.1). For the four major SEE countries (Slovenia, Croatia, Romania, and Bulgaria) the average in 1995 was 18.8 percent of the total loans on the books. However, information about the performance of borrowers in a rapidly changing environment is revealed only slowly under the best of circumstances so that these

measures are only illustrative of the serious overall problem of bad loans. Most governments responded to failing banks with efforts to save them from closure by recapitalization and the removal of bad loans from their balance sheets. For small insolvent banks, mergers with state-owned banks were used commonly. Repeated problems were inevitable because recapitalizations addressed only the stock of existing bad loans.

In the absence of independent market-oriented banking institutions, the flow of new bad loans continued to accumulate. Regulators did not have proper incentives, the requisite expertise, or sufficient independence to cope with this problem. To some extent, the bad loan problem was unavoidable because transition recessions and the dissolution of trading relationships within the Soviet bloc generated severe real sector shocks that were mirrored on the balance sheets of the banks. Nonetheless, even though the roots of this problem were difficult to resolve, the average ratios of non-performing loans to total loans had fallen to 6.4 percent for the four CEE countries and 5.9 percent for the four major SEE countries by 2005. To examine the resolution of the bad loans problem in more detail, we consider several countries' experiences.

The Hungarian government began to clean up the portfolios of its banks in the early 1990s when it enacted strong bankruptcy laws, new accounting regulations, and a new banking law. At the time, the Hungarian government provided guarantees to cover a portion of the debts of SOEs. However, firms continued to accumulate debts in arrears so that a second policy to address bad loans was introduced in 1992. The government replaced non-performing loans on bank balance sheets with government securities and transferred these assets to a government collection agency. Further recapitalizations introduced an element of moral hazard into the banking relationship. The situation changed when the authorities began to pursue an aggressive strategy of selling controlling stakes of the large SOCBs to foreign investors, signaling a credible commitment to no further bailouts. However, such a privatization strategy was not without difficulties as exemplified by an early transaction. The sale of a controlling stake in Budapest Bank, the third largest SOCB in Hungary, to GE Capital in 1995 was controversial because the buyer was given the right to off-load bad loans that were uncovered after the sale. Nevertheless, the banking expertise and discipline imposed by foreign owners of the three major SOCBs in Hungary led to rapid improvements in the banking environment. By the end of the 1990s, the Hungarian banking sector was well capitalized, loan quality had improved, claims on the state were a declining share of bank assets, bank staffing declined, bank margins narrowed, and, incidentally, bank regulation improved markedly (Organization for Economic Cooperation and Development, 1999; Hasan and Marton, 2003).

The government in the Czech Republic developed an explicit and detailed plan for privatization of most state-owned institutions, including SOCBs, using vouchers rather than direct sales. Initially, in 1991, bad loans were removed from

bank balance sheets and replaced with government bonds while the bad assets were taken over by a newly established hospital bank, Konsolidacni Bank. These resulting recapitalized large SOCBs were privatized by placing a minority stake of bank stock in the voucher program. However, non-state ownership of these partially privatized banks was dispersed with the largest stakes held by bank-related investment funds. Furthermore, the bank-related funds held ownership interests in their unrestructured industrial clients so that the large banks continued to lend to SOEs, which resulted in more bad loans. Hence, the key problems in the Czech Republic were interconnectedness between banks and their clients resulting from voucher privatization and the lack of independence of bank governance from a state holding controlling stakes in the banks. As a result, the resolution of bad loans required several rounds of recapitalization by the government, which increased the state's stake further and necessitated a second round of privatization. In this final round, foreign investors were allowed to take majority stakes in the large Czech banks, and bank behavior changed accordingly. The continuing efforts to restructure the Czech banks over the first decade of transition were expensive, with total costs amounting to more than 25 percent of 1998 GDP (Bonin and Wachtel, 2004).

In other countries, banking crises reached systemic proportions and severely impeded the overall transition to a market economy. In Bulgaria, weak bank governance and poor regulation of the many small SOCBs created from the commercial portfolio of the original monobank resulted in considerable asset stripping and insider lending. In addition, the macroeconomic shock of transition in Bulgaria was severe; in 1996, real GDP declined by 10 percent. Repeated rounds of recapitalization of banks resulted in a total cost to the government of 42 percent of 1998 GDP, which made the Bulgarian banking crisis one of the most costly of all transition countries. A currency board introduced in 1997 restored macroeconomic stability in Bulgaria and the banking system was rationalized quickly thereafter. In Romania, the dominant SOCBs accumulated large portfolios of bad loans and also required massive capital injections from the government. Non-performing loans peaked at 58 percent in 1998. In both of these SEE countries, severe macroeconomic shocks led to serious banking crises and sustainable economic growth resumed only after these crises were resolved.

After a decade and a half of transition, privatization of SOCBs is largely completed in CEE and SEE, although the situation is different in many countries emerging from the former Soviet Union. As the Czech and Hungarian experiences indicate, the privatization process differed considerably across the European transition countries. In Poland, the first bank privatizations utilized a combination of domestic initial public offerings (IPOs) and tenders to sell non-majority stakes to a strategic foreign investor. The Polish stock market was not very large; trading was not very extensive and bank stocks were the largest issues traded. Thus, bank IPOs were difficult to price and accusations of market manipulation lead to the political defeat of one of the early governments. The new government developed a

bank consolidation program as an alternative approach to privatization and attempted to force mergers and acquisitions of banks, but not without controversy. In one case, the attempt to include an already partially privatized bank (BPH) in the program caused a public uproar. Delays in privatization followed; almost a quarter of Polish bank assets remained in state hands as late as 2005. The two large banks that were still state-owned in that year, PKO (zloty savings bank) and BGZ (agricultural bank), had not participated in either consolidation or the privatization program.

Most of the later bank privatization programs in Romania, Bulgaria, Croatia, and the Czech Republic involved negotiated deals between the government and a single foreign bank, sometimes after a tender. In most transition countries, state ownership basically disappeared over a five-year period around the turn of the century. For the four CEE countries, average assets in state-owned banks were 27.1 percent of the total in 2000 and 5.9 percent five years later. For the four SEE countries, the average was 45.6 percent in 2000 and 8.0 percent in 2005. However, both the method (e.g., attracting a strategic foreign investor) and the timing of privatization matter to bank performance. Even after considering selection effects, Bonin, Hasan, and Wachtel (2005b) conclude that voucher-privatized and late-privatized banks lagged in performance and efficiency relative to non-voucher and early privatized banks.

The surprising aspect of banking in the transition countries is not the depth of the crises after the end of communism but the speed with which financial restructuring took place subsequently. The rapid changes in the last decade can be attributed to two related phenomena. First, the desire of European transition countries to qualify for European Union membership was a strong force for reform, not only in the eight original transition accession countries but also in the later joiners and in countries still hoping to join. Thus, improvements in bank regulation and investments in the banking sector took place rapidly. Second, the prospect of European Union membership (and ultimately the adoption of the euro) made these under-serviced banking markets attractive to European banks once macroeconomic stability was attained and reasonable regulations were in place. However, the governments in many transition countries were reluctant to allow foreign ownership for all the common arguments that attempt to show that foreign direct investment (FDI) in banking, unlike all other FDI, is dangerous. The usual claims that foreign-owned banks would facilitate capital flight and fail to provide credit for local economic development were made. As noted earlier, Hungary was the exception in that foreign banks were allowed to operate even before the transition and SOCBs were sold to foreign investors early in the transition. However, other transition governments took longer to realize that privatization to foreign buyers is not only a source of revenue but also a means of improving bank performance to support intermediation in the new market economy.

The proportion of assets in foreign-owned banks rose from virtually zero in the early 1990s to more than half in most countries a decade later. By 2005, the average share of assets in foreign-owned banks was 84.5 percent in the four CEE countries and 61.9 percent in the four SEE counties. In most cases, privatization by itself was not sufficient to improve bank performance; rather, joint ownership with foreign strategic investors was the crucial determinant in behavioral change (Bonin, Hasan, and Wachtel, 2005a). The FSU countries are an exception; foreign banks are not a major factor in Russia or in any other former Soviet republic, except for the Baltic countries. To some extent, this outcome follows from banking regulations that inhibit foreign entry and from reluctance on the part of many governments to accept foreign dominance of the banking sector. For example, although Russia has relaxed its limits on the overall size of the foreign banking sector, it sets minimums for the number of Russian employees and board members in foreign banks. In addition, unstable supervisory environments and weak legal protection have deterred foreign interest in such investments.

The characteristics of banking in Russia differ considerably from patterns found in CEE and SEE. In addition to three dominant SOCBs, Russia has a large number of mostly very small private commercial banks and many pocket banks having industrial owners. Some of these banks were involved in speculative activity and many were insolvent when the Russian government defaulted on its debt in 1998. At the time, weak bankruptcy laws and poor regulation made it difficult to close institutions so that the managers or owners were able to strip banks of any remaining good assets. The severe crisis in the banking sector did not have too large an impact on the real economy because the credit-to-GDP ratio was considerably lower in Russia than such ratios in the CEE transition countries, and cash was used widely for transactions throughout the FSU. Exacerbating the economic crisis in 1998 was uncertainty about the economic and legal environment.

Since the crisis, the Russian banking sector has shown some signs of improvement. Although more than one thousand banks still operate, this number is roughly half of the total in 1995 (Table 33.1) due to consolidations and closures. In addition, the influence of foreign banks is increasing as three foreign-controlled banks (including Citibank) are among the fifteen largest banks in Russia. Moreover, financial intermediation has increased as the bank asset to GDP ratio is double its level before 1998, though still lower than in the European transition countries. Nonetheless, some of the private banks still operate as private financial services institutions for their energy sector owners and provide little overall intermediation. The banking system is still fragmented with many small and poorly capitalized institutions characterized by poor governance, inadequate risk management, and high operating costs. Although deposits have increased, household savings are still largely held in the state savings bank, Sberbank, or in cash (Steinherr, 2006). Sberbank and Vneshtorgbank, the former foreign trade bank, have begun to provide credit to the private sector even though the government has no current

plans to privatize either of these SOCBs. Sberbank is the dominant bank in Russia holding more than 25 percent of all banking assets at the end of 2005. The next two largest banks in Russia were also SOCBs; Vneshtorgbank had about 7 percent market share and Gazprombank had 4.5 percent market share. At that time, no other Russian bank had a market share above 2.4 percent (Barisitz, 2007: Table 5.20).

In all countries, successful restructuring and privatization in the financial sector depend on the establishment of an effective institutional and legislative infrastructure to support proper regulation. In addition to developing an arm's-length legislative framework for banking regulation and supervision, bankruptcy laws, and international accounting standards are required to change the behavior of economic agents who are accustomed to operating in a non-market environment. Moreover, training of bank supervisors and other types of professional human capital development are needed to promote effective implementation of the legislation. Although the basic legal framework for modern banking was established early in the transition, additional related elements that are crucial for its effective functioning took more time to develop. In particular, a modern banking sector needs a functioning credit information system, which includes a credit registry and ratings agencies, and a reliably functioning court system to mediate contract disputes.

Hungary took the lead among the transition countries in promoting such institutional development with a legislative shock therapy program in 1992. In January, the government promulgated new, modern banking legislation, instituted international accounting standards, and revised its bankruptcy law to include a draconian trigger that resulted in a large number of company insolvencies. In addition, Poland developed a computer-supported system of bank oversight at the beginning of the transition and had in place rather stringent bankruptcy legislation for private firms even before the political change. Other countries took considerably longer to address these problems and, as a consequence, bank restructuring and privatization took longer to complete.

The maturation of transition banking sectors

The distinctive characteristic of the rationalization of banking sectors observed in virtually all transition countries is the rapid emergence of foreign-dominated ownership. As Table 33.2 reports, the asset share of foreign-owned banks was less than 50 percent in 1999 in all ten of the countries listed, except for Hungary and

Table 33.2. Banking market characteristics in later transition, 1999

	Ownership—asset shares[1]		Intermediation[2]		
	State-owned	Foreign-owned	Deposits/Gross domestic product	Loans/ Gross domestic product	European Bank for Reconstruction and Development index[3]
Central Eastern Europe and Russia					
Czech Republic	41.2	27.1	66.1	44.6	3.3 (+)
Hungary	7.8	65.3	36.4	24.6	4.0 (+)
Poland	23.9	69.5	35.4	27.1	3.3 (+)
Slovakia	50.7	24.6	57.1	48.4	2.7 (0)
Russia	41.9 (1998)	10.6*	10.2 (2000)	13.1 (2000)	1.7 (−)
South Eastern Europe					
Bulgaria	50.5	44.7	21.3	10.7	2.7 (+)
Croatia	39.8	39.9	34.0	35.7	3.0 (+)
Romania	50.3	47.8	20.5	10.6	2.7 (−)
Serbia	90.9#	0.5#	9.7	29.6	1.0 (0)
Slovenia	42.2	4.9	32.0##	35.8	3.3 (+)

Notes: *indicates share in registered statutory capital. # indicates data for 2000 and state-owned includes social ownership. ## indicates data for 2000 and for primary deposits only.

1 Ownership is the asset share of state-owned and foreign-owned banks, respectively, in total banking assets from Barisitz (2007) except for Slovenia. Slovenian data are from European Bank for Reconstruction and Development, *Transition Report* (2005).

2 Deposits of and loans to the private sector at year end are in percentages of gross domestic product from Barisitz (2007), except for Slovenia where data for deposits are from Bonin (2004), and data for credits are from European Bank for Reconstruction and Development, *Transition Report* (2005).

3 European Bank for Reconstruction and Development index of banking sector reform from European Bank for Reconstruction and Development, *Transition Report* (2005); it takes values between 1.0 and 4.0+. Also, (+) indicates an increase in the index from 1995, (0) indicates no change, and (−) indicates a decrease in the index from 1995.

Poland. Table 33.3 indicates that by 2005 only the Russian and Slovenian banking sectors exhibited such a small level of foreign participation. The asset-share of foreign-owned banks in CEE and SEE countries is now among the highest of any banking sector in the world, with Croatia, the Czech Republic, and Slovakia recording percentages above 90 percent and Hungary not far behind at 84 percent, in 2005. In addition, Serbian banking experienced a remarkable transformation over a five-year period; foreign ownership increased from a negligible amount in 2000 (0.5 percent) to 66 percent in 2005. Russia and Slovenia remain outliers on this measure, with foreign participation at only about 11 percent and 23 percent, respectively, in 2005. However, the asset share of state-owned banks was lower in Slovenia at 12 percent than in Russia (38 percent), Serbia (24 percent), and Poland (21.5 percent). Hence, Slovenia appears to be an anomaly among European transition economies with respect to the ownership structure of its banking sector.

Regarding the pace of restructuring, the results from 1995 to 1999 are mixed. The EBRD index of banking reform increased for six of the countries but it actually decreased for Russia and Romania, with no change in the index recorded for Serbia and Slovakia (Table 33.2). By 1999, only Hungary had a rating of 4 on a scale from 1 to 4+, where the highest score reflects full convergence to performance norms and regulation standards of advanced industrial economies. By 2005, the Czech Republic and Croatia joined Hungary with scores of 4 while six of the seven other countries recorded an increase in the index from 1999 (Table 33.3). Hence, banking sectors in most transition countries have reached, or are rapidly approaching, the performance of their counterparts in developed market economies, with one major difference—namely, an extremely high foreign bank presence. Russia and Slovenia are the outliers on both counts.

Based on the legacy of segmented sectors and exacerbated by consolidation programs, banking concentration is high in most transition countries. In 2005, the three-firm concentration ratio ranged from a high of over 65 percent in the Czech Republic to about 33 percent in Bulgaria, with six of the ten countries listed in Table 33.3 having a ratio above 40 percent. Moreover, the five-firm concentration ratio in all SEE countries was 50 percent or above. Only Poland and Russia had five-firm ratios below this threshold. However, high concentration ratios have not prevented competition from developing in many of these banking sectors. As Table 33.4 indicates, interest rate spreads declined considerably since the beginning of the transition, which may be attributable more to reduced risk in the macroeconomic environment than to increased banking competition.

Considerable differences exist among countries with respect to interest rate spreads. In 2005, Hungary had the lowest spread, while Romania and Serbia still had spreads above 10 percent. Of these ten countries, only Hungary, Slovakia, and Slovenia had average spreads from 2001 to 2005 under 5 percent, which we take to indicate a reasonably competitive banking sector. Interestingly, the Czech Republic, Croatia, and Poland had lower average inflation rates during this period but higher

Table 33.3. Banking market characteristics in 2005

	Ownership[1]		Intermediation—ratio to gross domestic product[2]				Concentration[3]		European Bank for Reconstruction and Development index[4]
	State	Foreign	Deposits	Loans	Household loans	Mortgages	3-firm	5-firm	
Central Eastern Europe and Russia									
Czech Republic	2.5	94.5	55.9*	27.3*	13.9	7.9	65.4	75.6	4.0 (+)
Hungary	7.0	84.5	40.1	44.8	15.7	11.6	41.9	56.6	4.0 (0)
Poland	21.5	74.2	34.6*	27.4*	12.6	5.1	33.4	45.3	3.7 (+)
Slovakia	1.1	97.3	47.7*	32.5*	13.0	3.6	47.7	66.3	3.7 (+)
Russia	38.1	11.2**	17.7	25.7	4.9	0.2	37.0	41.7	2.7 (+)
South Eastern Europe (SEE)									
Bulgaria	1.7	72.8	36.0*	34.9*	14.7	4.8	32.8	49.9	3.7 (+)
Croatia	3.4	91.2	59.8*	55.8	34.3	12.0	55.3	75.3	4.0 (0)
Romania	6.5	59.2	26.1	20.9	7.3	0.6	48.3	59.8	3.0 (+)
Serbia	23.9	66.0	25.1	25.0	7.6	1.4	37.3	53.2	2.7 (+)
Slovenia	12.0	22.6	n/a	56.4	14.5	3.6	50.4	73.3	3.3 (+)

Notes: *indicates data for 2004. **indicates share in registered statutory capital.

1 Ownership is the asset share of state-owned and foreign-owned banks, respectively, in total banking assets from Barisitz (2007), except for Slovenia. Slovenian data are from European Bank for Reconstruction and Development, *Transition Report* (2006).

2 For deposits and loans, see Table 33.2, note 2. Household loans is the ratio of total outstanding bank credit to households at year end to gross domestic product and Mortgages is the ratio of mortgage lending to households at year end to gross domestic product, both from European Bank for Reconstruction and Development, *Transition Report* (2006).

3 Concentration measures are computed as market shares of the top number of banks indicated, both from Barisitz, except for Slovenia. Slovenian data are for 2000 from Bonin (2004).

4 European Bank for Reconstruction and Development index of banking sector reform from European Bank for Reconstruction and Development, *Transition Report* (2005); it takes values between 1.0 and 4.0+. Also, (+) indicates an increase in the index from 1999, (0) indicates no change, and (−) indicates a decrease in the index from 1999.

Table 33.4. Interest rate spreads and inflation, 1991–2005

	Interest rate spread			Consumer price index inflation rate		
	1991–5 average	2001–5 average	2005	1991–5 average	2001–5 average	2005
Central Eastern Europe and Russia						
Czech Republic	6.5 (1992–5)	6.5	6.0	12.1 (1992–5)	2.2	2.2
Hungary	8.0	2.5	2.2	24.9	5.2	3.3
Poland	10.5	7.1	5.9	38.7	2.2	0.7
Slovakia	6.1 (1993–6)	4.8	4.5	12.4 (1993–6)	5.8	3.7
Russia	43.9 (1996–8)	8.7	7.5	39.1 (1996–8)	13.7	10.9
South Eastern Europe						
Bulgaria	29.4	6.1	4.9	127.4	5.0	6.5
Croatia	13.6 (1994–6)	8.7	8.3	1.4 (1994–6)	2.4	3.6
Romania	21.5 (1993–6)	15.8	11.5	110.6 (1993–6)	16.0	8.6
Serbia	86.4 (1996–8)	15.8	11.7	37.5 (1996–8)	18.3	17.5
Slovenia	6.9 (1995–7)	4.3	4.5	8.9 (1995–7)	4.9	2.3

Notes:

1 Spreads are computed as the difference between lending rates and deposit rates from the country tables in European Bank for Reconstruction and Development, *Transition Report*, var. issues. Maturities are always less than one year but they differ across countries.

2 Inflation rates are year end changes in the Consumer Price Index. Data are taken from the country tables in European Bank for Reconstruction and Development, *Transition Report*, var. issues.

interest rate spreads. By 2005, Bulgaria joined the countries having interest rate spreads below 5 percent. Of these four countries, Croatia and Slovenia have relatively high three-firm concentration ratios at over 50 percent. Regarding foreign participation in the banking sector, Slovenia is the outlier with less than 23 percent of assets in foreign-owned banks in 1995. Moreover, the Czech Republic and Poland have high percentages of banking assets in foreign banks and low inflation rates but relatively high interest rate spreads. Thus, the experiences of the European transition countries indicate that neither high foreign participation in the banking sector nor low inflation is a sufficient condition for competitive interest rate spreads.

The ratio of bank deposits to GDP is a measure of both banking sector development and public confidence in the banking system. By comparing the entries in Tables 33.2 and 33.3, we find considerable differences across countries in this ratio and in its changes from 1999 to 2005. In the Czech Republic and Slovakia, the ratio of bank deposits to GDP was fairly high in 2005 although it had decreased considerably since 1999, which may suggest some decline in public confidence. In Croatia, the ratio of bank deposits to GDP increased dramatically to the highest of any of the ten countries by 2004, which reflects both a credit boom and increased confidence in banks. The 2005 ratios for Hungary, Romania, and Russia show modest growth of around five percentage points from 1999 while Poland experienced virtually no change in this ratio. Both Bulgaria and Serbia experienced considerable increases in deposits to GDP from 1999 to 2005. Public confidence in banks is important to a well-functioning banking system in any transition economy. Based on the ratio of deposits to GDP, the evidence is mixed but the laggards are improving rapidly.

According to the EBRD Transition Report of 2006, the banking sectors of transition economies have exhibited considerable growth and diversification since 2000, although further progress in financial deepening is considered to be both feasible and desirable. On the lending side, four of the ten transition countries listed in Tables 33.2 and 33.3 experienced increases in the ratio of loans to GDP of more than 20 percent from 1999 to 2005. Ratios in 2005 (or 2004 when indicated) in Slovenia and Croatia were around 56 percent, which equaled the worldwide average of domestic credit to the private sector as a percentage of GDP. Hungary, at about 45 percent, and Bulgaria, at about 35 percent, have the next highest ratios. Of the other six transition countries, only Slovakia had a ratio of loans to GDP above 30 percent by 2005, with the other five between 21 percent (Romania) and 28 percent (Czech Republic and Poland). As a further basis for comparison, the European Union average for this measure of financial depth was 86 percent in 2005. Hence, even the four leading transition countries are well below the European Union average in providing credit to the private sector.

In the same document, the EBRD reports that the share of loans to households increased sharply in CEE and SEE countries, with much of the increase due to

mortgage lending. By 2005, domestic credit to the household sector as a percentage of GDP ranged from a high of over 34 percent in Croatia to less than 10 percent in Romania, Russia, and Serbia (Table 33.3). Retail credit accounted for well over half of all loans in Croatia and around half of the total in the Czech Republic and Poland. Mortgage lending as a percentage of GDP in 2005 ranged from highs of around 12 percent in Croatia and Hungary to moderate levels of about 8 percent in the Czech Republic and around 5 percent in Bulgaria and Poland to virtually nothing in Romania and Russia (Table 33.3). Non-mortgage household credit is particularly large in Croatia (22.3 percent of GDP). To what extent the recent explosion of retail credit in some transition countries will lead to instability in the banking sector is yet to be determined, but it will be influenced considerably by the use to which credit has been put and the possibility of real estate bubbles occurring.

Household credit, in particular mortgage lending, depends on well-defined property rights over collateral and an effective legal infrastructure to facilitate the collection of collateral in case of default. Hence, the dramatic growth of both types of lending in many transition countries reflects significant improvements in supportive institutions. Nonetheless, differences in retail lending ratios across these ten countries are large. Consistent with the other measures of financial intermediation, retail credit data indicate considerable progress in banking in Bulgaria, Croatia, and Hungary. More sluggish development in the Czech Republic, Slovakia, and Poland may be inferred from the intermediation data. Romania, Russia, and Serbia appear to be either laggards or late starters in all areas of banking sector reform. Finally, Slovenia is an anomaly in that its ratio of loans to GDP is near the top of all ten countries in 2005, but retail credit, and especially mortgage lending, lag well behind these activities in many other countries.

Credit growth throughout the region slowed in 2007 and 2008 as the international financial crisis affected economies, particularly those that were closely integrated with the euro area (Hungary and the Baltics) or vulnerable to swings in energy prices (Russia and Kazakhstan). Countries with macroeconomic imbalances were particularly vulnerable to the worldwide 'credit crunch' that reduced volume in international bond and syndicated loan markets. However, the banks in the transition countries were relatively unaffected in the initial stages of the crisis. They did not experience large write-offs and short-term funding from parent banks seemed to hold up through 2008. However, the resiliency of transition banking does not mean that the sector will be immune to the upheaval in world financial markets (Gardó, Hildebrandt, and Walko, 2008).

Hungary was among the first emerging market countries to suffer the fall out of the global 'credit crunch'. It was vulnerable because of a large fiscal deficit, its reliance on external financing, and the extent of domestic, particularly household, borrowing in foreign currency. The 'credit crunch' led to pressure on the florint and an increase in the country risk premium. In October 2008, the IMF, the World Bank, and the European Union joined forces to provide a $25 billion support

program. Importantly, the program included provisions for pre-emptive additions to bank capital and guarantees for the interbank market. That is, the macroeconomic issues and financial sector stability are inseparable problems.

The Russian banking system encountered serious liquidity problems late in 2008. The problems in the banking system stem from the fall in oil prices and the depreciation of the Ruble while many institutions borrowed abroad in foreign currencies. The central bank eased its refinancing terms and extended deposit insurance coverage and the government offered support to enterprises in trouble. The Russian banking system is much stronger than it was before the 1998 crisis but it is still vulnerable to large macroeconomic shocks.

Evaluation of transition banking and prospects for the future

Although banks in the transition countries have made rapid strides in improving performance and services since the early 1990s, the banking sectors in the European transition economies still do not posses the financial depth of their European Union counterparts nor are banking services as well developed in these countries. Nonetheless, with few exceptions (primarily in the FSU), the transition in banking is complete. State monobanking structures have been replaced by privately owned, market-oriented, well-capitalized banking institutions that are independent from the government and from state-owned clients. The legal environment has improved with respect to bankruptcy laws, collateral laws, and confidence in the application of the law. Furthermore, banking regulatory and supervisory capabilities have developed considerably. Thus, any evaluation of the structure of banking in transition countries must be positive. However, banking conduct is a somewhat different matter; any evaluation of what banks are doing and how they are contributing to economic performance in the transition economies must be more nuanced.

The ratio of bank credit to GDP depends on the financial structure of a country; it will be larger in bank-centered financial systems than in countries having more-developed capital markets. For the transition countries, the financial depth ratio is well below industrial country levels, although the numbers are not unusual for countries with similar GDP levels. In some CEE countries, this ratio has fallen as bad loans have been removed from balance sheets while GDP has grown. Deepening has occurred in the major FSU countries with the achievement of financial stability and the resulting return of public confidence in banks. Financial deepening or increasing intermediation has been shown to be associated with more rapid economic growth in cross-country studies (Wachtel, 2001). Thus, the increased

credit ratios in the SEE should be viewed as a positive development even though they have been met with concern in some countries—that is, in Croatia where the ratio went from 35.7 to 55.8 in five years as Tables 33.2 and 33.3 indicate and in Bulgaria where it increased from 10.7 to 34.9 over the same period. The main concern is that credit deepening has come in the form of rapid growth in mortgage lending and other forms of consumer credit.

Lending to households has grown rapidly in many countries. In 2005, it was more than one-half of total bank lending in Croatia and in the Czech Republic. Despite rapid increases in household credit, ratios of household credit to GDP are still not large by developed-country standards. However, the ratio of household credit to the financial wealth of the consumer sector is high in Croatia and elsewhere, suggesting some vulnerability of consumers to economic shocks (European Bank for Reconstruction and Development Transition Report, 2006). Although rapid credit increase might have long-term growth benefits in general, it could also be a sign of excessive risk taking and financial vulnerability.

The expansion of household lending in transition countries may be related to the dominance of foreign-owned banks. Once the legal environment is in place, lending to households is a commodity business that can be entered easily through the application of banking technology from abroad. In contrast, lending to enterprises requires developing client relationships and having the ability to evaluate unique situations, both of which require expertise that is generally lacking in foreign banks. Using a recent EBRD survey, Haselmann and Wachtel (2009) show that banks in many transition economies have shifted their asset portfolios out of government securities toward mortgages and consumer credit. Foreign banks in particular have increased consumer lending and only maintained the existing level of lending to enterprises. The EBRD/World Bank surveys of enterprises in transition countries indicate that many firms are financially constrained in the sense that they are unable to obtain bank lending. Based on these surveys, the EBRD concludes that 'despite some regional variation, bank loans still play a limited role in enterprise financing' (European Bank for Reconstruction and Development Transition Report, 2006: 47). Since lending to enterprises is important to support economic growth, this finding has important implications for any evaluation of the conduct of banking in transition countries.

Foreign banks have had a positive influence on the banking environment by introducing technology, operational efficiencies, and new products and services. However, Haselmann and Wachtel observe that foreign banks have focused on lending to households and large firms. In addition, the EBRD surveys provide little evidence of increased lending to small and medium-sized enterprises (SMEs). To some extent, the lack of SME lending in a foreign bank-dominated country is understandable because such lending requires local knowledge. However, the large foreign banks were created by mergers and acquisitions of local entities so that this knowledge should not be prohibitively difficult to acquire. Moreover, the surveys

suggest that improvements in the legal environment for banking have been associated with greater risk taking and more credit extended to SMEs (European Bank for Reconstruction and Development Transition Report, 2006; and Haselmann and Wachtel, 2007). Frequently, the survey respondents indicate that a lack of creditworthy borrowers and difficulty in evaluating risks were the main reasons for slow loan growth. In their lending activity, banks in transition countries tend to favor large firms and foreign affiliates currently. However, improvements in the legal and regulatory institutions are expected to induce more SME lending. Hence, environmental improvements such as good bankruptcy laws, efficient ownership structures, reliable court systems for their application, credit registries, and defined legal rights to collateral should lead to more lending to SMEs and more support of local entrepreneurs in the future (de Haas and Lelyveld, 2006).

Moreover foreign bank ownership makes banking systems more vulnerable to the worldwide 'credit crunch'. Although, there are no reports of transition country banks suffering large losses on US mortgage securities, their European parent banks may have. In this case, the parent banks may be less wiling to provide funding to their transition subsidiaries and credit standards may tighten as the parent banks reduce risk exposures across the border. Further it is not clear that every transition country central bank would be able to maintain liquidity in the banking sector and confidence in domestic institutions if the foreign parent banks withdrew support.

Overall the growth in banking in transition countries has increased considerably the availability of financial services, many of which were simply not obtainable before. Whether banks can become formidable engines of sustainable economic growth in transition economies is an open question. Many large enterprises, particularly in the European Union new member states, are able to take advantage of recent increases in European capital market integration and obtain financing from abroad. However, these sources of funding fell with the global 'credit crunch' starting in 2007. Furthermore, non-bank financial institutions are emerging in the transition economies. Nonetheless, the rapid expansion of credit in some countries has become a source of concern because of the accompanying potential increases in risk to the banking sectors. In addition, much of the lending by banks in some transition countries, particularly the SEE countries that experienced hyperinflation in the 1990s is denominated in foreign currencies and many deposits are denominated in non-national currencies as well. Thus, the balance sheets of banks in these countries are exposed to foreign exchange risk. In Croatia, 70 percent of mortgages are denominated in euros. Even though the deposit base of these banks is also in euros, foreign exchange risk is not eliminated by this matching because a domestic slowdown or exchange rate shock would affect the ability of domestic borrowers to repay in euros.

These risks and indeed many of the problems, faced by banks in transition countries are familiar to banks in small, open, emerging-market economies around the world. Moreover, the trade-off between bank consolidation and bank

concentration is relevant to other small banking sectors. Although consolidation eliminates inefficient and undersized institutions, it also increases concentration, which may limit competition and create systemic risks. To some extent, free entry and foreign bank participation can mitigate this anti-competitive tendency. Although foreign bank penetration is a worldwide phenomenon in emerging-market economies, it is more prevalent and more concentrated in a subset of home countries in transition economies. European banks, mainly from the Netherlands, Italy, and Austria, are most active due to particularly strong trading relationships or to a desire to enter expanding new markets close to their own countries. Overall, foreign-owned banks have maintained their lending activities in the presence of local shocks, although their aggressive growth targets may be a source of instability in the future.

The relationship between parent banks and their local partners is a mixed blessing. In some cases, the parent bank provides assistance for a troubled local institution—for example, KBC from Belgium supported its troubled Polish subsidiary, Kredytbank. However, parent bank support cannot be taken for granted—for example, Bayerische Landesbank walked away from its Croatian subsidiary, Rijecka Banka, when fraud was uncovered. In addition, ownership changes in the parent bank can affect the structure of banking in the host country. When HVB joined the Unicredito banking group, several Polish subsidiaries were merged to create the largest bank in Poland with a market share in excess of 25 percent despite objections from the Polish authorities. These close connections with specific foreign banking sectors combined with high concentration in local banking may leave some transition countries vulnerable to economic shocks in other countries—for example, Netherlands, Italy, or Austria.

Banking regulation in the European Union follows the home country principle in that the home country regulators supervise the consolidated balance sheet of multinational banks. At the same time, the host country regulators have responsibility over the local subsidiaries. Hence, a potential for conflict arises if a home country regulator does not have sufficient interest in a foreign subsidiary that is a small part of a multinational bank but an important player in the financial sector of the host country. Unfortunately, the lack of explicit coordination of bank regulation across borders is a problem that is overdue for attention. For example, the British authorities were not prepared to deal with the failure of the Icelandic banks that had large UK subsidiaries. It is unclear how authorities in both home and host countries would respond to the failure of any parent bank with subsidiary operations in the transition countries. Since foreign-owned banks dominate some of the transition banking systems, the potential for systemic crisis is clear—something the European Union will probably avoid facing until the first cross-border banking crisis hits. Most recently, an exchange rate crisis in Latvia in 2009 placed pressure on the Swedish banks that own the local banks and have already received support form the Swedish authorities.

In summary, considerable strides have been made in developing mature banking sectors in virtually all European transition countries. However, this positive evaluation must be tempered by some concerns about future stability owing to the dominance of foreign banks from a handful of countries. The less advanced transition countries, largely the smaller republics of the FSU, are just beginning to create modern banking sectors. These countries now have models to emulate; hence, their progress toward achieving mature and effective banking institutions warrants careful watching to see if the relevant lessons have been learned.

Banks in the transition economies have become part of the competitive global financial industry. As such, they are exposed to the shocks of the world financial crisis and the macroeconomic shocks affecting many transition countries. It remains to be seen how resilient the banks will be to these challenges. The experience of transition banks and banking authorities in this era will increase our understanding of the role of openness, performance, and ownership in banking.

References

Abarbanell, J. and Meyendorf, A. (1997). Bank Privatization in Post-Communist Russia: The Case of Zhilsotsbank, *Journal of Comparative Economics*, 25: 62–96.

Barisitz, S. (2007). *Banking in Central and Eastern Europe 1980–2006: A Comprehensive Analysis of Banking Sector Transformation in the Former Soviet Union, Czechoslovakia, East Germany, Yugoslavia, Belarus, Bulgaria, Croatia, the Czech Republic, Hungary, Kazakhstan, Poland, Romania, the Russian Federation, Serbia and Montenegro, Slovakia, Ukraine, and Uzbekistan*. New York and London: Routledge.

Bonin, J. P. (2004). 'Banking in the Balkans: The Structure of Banking Sectors in Southeast Europe', *Economic Systems,* 28: 141–53.

—— and Wachtel, P. (2003). 'Financial Sector Development in Transition Economies: Lessons from the First Decade', *Financial Markets, Institutions & Instruments*, 12: 1–66.

—— —— (2005). 'Dealing with Financial Fragility in Transition Economies', in D. Evanoff and G. Kaufman (eds.), *Systemic Financial Crises: Resolving Large Bank Insolvencies*. Singapore: World Scientific Publishing, 141–59.

—— Hasan, I., and Wachtel, P. (2005a). 'Bank Performance, Efficiency and Ownership in Transition Countries', *Journal of Banking & Finance,* 29: 31–53.

—— —— —— (2005b). 'Privatization Matters: Bank Efficiency in Transition Countries', *Journal of Banking & Finance,* 29: 2155–78.

De Haas, R. and Lelyveld, I. V. (2006). 'Foreign Banks and Credit Stability in Central and Eastern Europe: A Panel Data Analysis', *Journal of Banking & Finance,* 30: 1927–52.

European Bank for Reconstruction and Development (various issues), *EBRD Transition Report*.

Gardó, S., Hildebrandt, A., and Walko, Z. (2008), 'Walking the Tightrope: A First Glance on the Impact of the Recent Global Financial Market Turbulence on Central,

Eastern and Southeastern Europe', Oesterreichische Nationalbank, *Financial Stability Report,* 15.

Hasan, I. and Marton, K. (2003). 'Banking in Transition Economy: Hungarian Evidence', *Journal of Banking & Finance,* 27: 2249–71.

Haselmann, R. (2006). 'Strategies of Foreign Banks in Transition Economies', *Emerging Markets Review,* 7: 283–99.

—— and Wachtel, P. (2006). 'Institutions and Bank Behavior', New York University Stern School of Business Working Paper EC06-16.

—— —— (2007). 'Risk Taking by Banks in the Transition Countries', *Comparative Economic Studies,* 49: 411–29.

—— —— (2009). 'Bankers' Perceptions of the Legal Environment and the Composition of Bank Lending: Evidence from the Transition Countries', New York University Stern School of Business Working Paper EC06-16, Revised.

Organization for Economic Cooperation and Development (1999). *OECD Economic Surveys* 5 (Hungary).

Steinherr, A. (2006). 'Russian Banking Since the Crisis of 1998', *Economic Change and Restructuring,* 39: 235–59.

Wachtel, P. (2001). 'Growth and Finance: What Do We Know and How Do We Know it?', *International Finance,* 4: 335–62.

CHAPTER 34

BANKING IN LATIN AMERICA

FERNANDO J. CARDIM DE CARVALHO
LUIZ FERNANDO DE PAULA
JONATHAN WILLIAMS

Introduction

During the 1980s and, even more notably, the 1990s, banking systems in Latin America were deeply transformed. Liberal reforms were widely adopted in the region. Common features of these reforms were the liberalization of interest rates, the attenuation of barriers to entry in the provision of banking services, large-scale privatization of state-owned banks and the facilitation of entry for foreign banks (see Singh, et al., 2005; and Stallings and Studart, 2006). In parallel, but in a largely independent process, liberalization of the capital account of the balance of payments also influenced the evolution of domestic financial systems, since it opened new opportunities of investment for resident wealth holders at the same time in which it made possible for non-residents to buy assets and offer financial services to residents. The downside of such a process, of course, is the increasing exposure of these economies to the volatility of international financial markets.

The joint impact of all these changes was to transform deeply the ways financial systems work in Latin America. In fact, the transformation process is still unfolding, although nowadays in the shape of a sharp expansion of securities markets and

also of a fast increasing supply of bank credit to private borrowers, even though it has generally started from very low levels in the region. Among the most visible changes already achieved is the strong process of bank consolidation that has taken place in the period. Of particular interest are the effects of consolidation on competition and efficiency in banking sectors.

This chapter is structured in the following way. After this Introduction, the second section provides an overview of the recent consolidation of banking sectors in Latin America. In the third section, we consider the evolution of financial policy and how it has contributed toward the recent consolidation process. The fourth section investigates the effects consolidation has had on banking sectors with the fifth section offering some concluding remarks. As the subprime mortgage crisis exploded after this chapter was completed, and since so far its impact on the region has been relatively moderate, we have added an epilogue discussing perspectives for the three major economies in the region: Argentina, Brazil, and Mexico.

Banking consolidation in Latin America: A quick overview

Banking crises, financial deregulation, and the globalization of financial services have led to a significant increase in foreign bank penetration of emerging market banking sectors over the latter half of the 1990s. The effects of these developments have been summarized as follows:

> global market and technology developments, macroeconomic pressures and banking crises in the 1990s have forced the banking industry and the regulators to change the old way of doing business, and to deregulate the banking industry at the national level and open up financial markets to foreign competition. . . . These changes have significantly increased competitive pressures on banks in the emerging economies and have led to deep changes in the structure of the banking industry. (Hawkins and Mihaljek, 2001: 3)

Whilst the process of bank consolidation in industrialized and emerging markets has been shaped by the above forces, some specific features have characterized consolidation in emerging markets (International Monetary Fund, 2001; and Gelos and Roldós, 2004). First, cross-border mergers and acquisitions (M&A) have been an important source of consolidation in emerging markets, yet the exception in industrialized markets. Second, consolidation was used to restructure emerging-market banking sectors after financial crises rather than to eliminate excess capacity or improve bank efficiency as in industrialized markets. Finally, emerging

market governments actively participated in the process of consolidation, whereas consolidation tended to be 'market-driven' in industrialized markets since it represented financial institutions' response to policies of financial deregulation that were implemented in the 1970s and 1980s.

Bank consolidation has been more advanced in Latin America compared to other emerging markets. National governments actively participated in bank restructuring and implemented substantial bank privatization programs, though in countries such as Argentina and Brazil some large banks remain under state ownership. Since the 1990s ended, the consolidation process—especially in Brazil and Mexico—has become increasingly market-driven (as in industrialized markets). Generally, the desire to enhance competition and efficiency and, in some cases, to restructure public finances formed the background to almost all privatization programs in the region. The role played by foreign banks in the restructuring and consolidation of domestic banking sectors should not be underestimated. The mid-1990s banking sector crises offered foreign banks

> a one time set of opportunities to invest in financial institutions and to expand business . . . A standard response to crises by EME (emerging market economies) government, encouraged by the international financial institutions, was to accelerate financial liberalization and to recapitalize banks with the help of foreign investors.
>
> (Committee on the Global Financial System, 2004: 6)

This has happened in Argentina, Brazil, and Mexico. Foreign banks' shares of banking sector assets have increased substantially in Latin America, and although foreign bank penetration is not as extensive as in Central and Eastern Europe it is higher than in Asia (see Table 34.1).

In the 1990s, Latin America received record levels of foreign direct investment (FDI). In 1998 alone, the region received an inflow of $76.7 billion which was equivalent to 41 percent of total FDI to developing countries (Economic Commission for Latin America and the Caribbean, 2000: 35–6). The majority of investments were made in banking sectors. Between 1991 and 2005, a total of $121 billion was expended on cross-border M&A involving the acquisition of banks in emerging markets (Domanski, 2005). Of the total, 48 percent was spent in Latin America, with Asia and Central and Eastern Europe receiving 36 percent and 17 percent, respectively. The main source of investment in Latin America came from Spanish banks (46.6 percent of the value of acquisitions made by foreign banks of domestic banks in the region) followed by US (26.5 percent), UK (10.0 percent), Dutch (6.4 percent), and Canadian banks (3.6 percent).

Bank restructuring has increased the level of concentration in regional banking sectors. Whilst banks numbers have fallen—considerably in some countries—the accompanying increases in concentration were not as sharp. Measured by the three-firm deposit concentration ratio between 1994 and 2000, the largest banks in Brazil and Mexico grew their share of deposits to more than 55 percent, whereas

Table 34.1. Share of bank assets held by foreign banks'

Countries	1990	2004[2]	Gross Domestic Producy (%)	US$ billion
Central and Eastern Europe				
Bulgaria	0	80	49	13
Czech Republic	10	96	92	99
Estonia	—	97	89	11
Hungary	10	83	67	68
Poland	3	68	43	105
Emerging Asia				
China	0	2	4	71
Hong Kong	89	72	344	570
India	5	8	6	36
Korea	4	8	10	65
Malaysia	—	18	27	32
Singapore	89	76	148	159
Thailand	5	18	20	32
Latin America				
Argentina	10	48	20	31
Brazil	6	27	18	107
Chile	19	42	37	35
México	2	82	51	342
Peru	4	46	14	11
Venezuela	1	34	9	9

Notes:
1 Percentage shares of total banking sector assets.
2 Or latest available year.
Source: Domanski, 2005: 72, based on data from European Central Bank and national central banks.

the comparative shares of the largest banks in Argentina and Chile remained stable (see Table 34.2). Across the region, concentration of the ten largest banks increased (except Venezuela). Whilst Latin American banking sectors were more highly concentrated (in 2000) than Asian sectors, they were slightly less concentrated than sectors in Central Europe. However, Table 34.2 clearly shows that concentration increased in Latin America whereas it decreased in Asia and Central Europe (albeit with limited exceptions).

In countries such as Mexico and Argentina, the rise in the level of consolidation was closely tied to foreign bank penetration. In Mexico, foreign banks had unrestricted access to all sectors of the banking market and became market leaders. Whilst foreign banks came to dominate domestic banks in Argentina—as they increased their market share from 16.1 percent of total bank deposits in November 1994 to 51.8 percent in December 2001 (Fanelli, 2003: 52)—their presence partially wavered after the 2001–2 financial crisis as their market share declined whilst the

Table 34.2. Banking concentration in some selected emerging countries, 1994–2000

Countries	1994 Market share total of deposits (%)				2000 Market share total of deposits (%)			
	Banks quantity (1994)	Three major banks	Ten major banks	Herfindahl index (1994)	Banks quantity (2000)	Three major banks	Ten major banks	Herfindahl index (2000)
Asia								
Korea	30	52.8	86.9	1,263.6	13	43.5	77.7	899.7
Malaysia	25	44.7	78.3	918.9	10	43.4	82.2	1,005.1
Philippines	41	39.0	80.3	819.7	27	39.6	73.3	789.9
Thailand	15	47.5	83.5	1,031.7	13	41.7	79.4	854.4
Latin America								
Argentina	206	39.1	73.1	756.9	113	39.8	80.7	865.7
Brazil	245	49.9	78.8	1,220.9	193	55.2	85.6	1,278.6
Chile	37	39.5	79.1	830.4	29	39.5	82.0	857.9
Mexico	36	48.3	80.8	1,005.4	23	56.3	94.5	1,360.5
Venezuela	43	43.9	78.6	979.2	42	46.7	75.7	923.1
Central Europe								
Czech Republic	55	72.0	97.0	2,101.5	42	69.7	90.3	1,757.8
Hungary	40	57.9	84.7	1,578.8	39	51.5	80.7	1,241.2
Poland	82	52.8	86.9	1,263.6	77	43.5	77.7	899.7

Source: International Monetary Fund, 2001: 127.

market shares of private and mainly public-owned banks increased. Domestic private and public banks are market leaders in Brazil. Indeed, privately owned banks responded proactively to foreign bank penetration and became active in domestic M&A (Paula and Alves, 2007). The consolidation process in Chile proceeded more gradually: it has increased because of M&A in Spain (the home country of the parent banks of the two largest banks in Chile); technically, the enlarged Spanish parent has operated its Chilean subsidiaries as individual entities (Ahumada and Marshall, 2001).

Bank restructuring and privatization ushered in a new wave of cross-border (and domestic) M&A activity. Cross-border bank M&A partially reflects country-specific factors: positively related to shared language (Spanish bank entry: Sebastián and Hernansanz, 2000) and geographical proximity (North American bank entry: Buch and DeLong, 2001); and the availability of access to large, relatively poor countries with widely spaced populations and underdeveloped financial sectors (Buch and DeLong, 2001; and Focarelli and Pozzolo, 2001). M&A can be analyzed in terms of the financial condition of buyers and targets. A recent application to Brazil differentiates between M&A involving domestic-owned and foreign-owned banks. The results suggest that domestic and foreign buyers

acquired target banks that had alternative profiles: domestic buyers have tended to buy underperforming banks whilst foreign buyers tended to acquire large, slow-growing institutions; the implication is that foreign banks have used M&A as the vehicle to increase bank size and market share (Cardias Williams and Williams, 2008).

The evolution of financial policy in Latin America

Although the process of transformation of Latin American banking systems has exhibited basically the same features, and took place in roughly the same period, its causes diverged from country to country. Post-1945, Latin American financial systems were typically repressed, and governments across the region attempted from the late 1940s on, with varying success, to accelerate economic growth and transform national social and economic structures. The key to become a developed country was thought to be becoming industrialized as quickly as possible. Inspired by the experiences of Central European countries (cf. Gerschenkron, 1962), Latin American governments, particularly in the largest countries (Brazil, Mexico, Argentina, and Chile) saw in the banking system a powerful instrument to centralize and direct the necessary resources to finance the growth of manufacturing production. Unwilling to rely on the eventual ability of freely operating financial markets to support an accelerated growth process, governments in those countries imposed financial repression (Fry, 1995), which consisted in this case mostly of creating, or enlarging the functions of, existing, state-owned banks, setting maximum interest rates to be charged on loans by private banks (frequently adopted in the context of usury laws), and directing the credit supplied by these banks to sectors considered strategic to enhance economic growth.

This is not the place to assess how successful these initiatives were in promoting growth.[1] The region suffered heavily with the oil shocks of the 1970s. The attempts to deal with the effects of those shocks by increasing short-term foreign debt led to the debt crisis of the early 1980s that brought the most important economies of the region to a standstill that lasted so long it became known as the 'lost decade of economic growth'. As part of the negotiated resolution package for that crisis, practically all countries in Latin America accepted to promote liberalizing reforms,

[1] Higher growth rates were in fact achieved, although at the cost of the emergence of some important disequilibria.

including in the financial sector and thereby ending the financial repression experiment.

Chile was the pioneer in this process (see Stallings and Studart, 2006; and Foxley, 1983). Liberal reforms in the banking markets, including privatization of state-owned financial institutions began right after the 1973 military coup that ousted then-President Salvador Allende. The root cause of financial liberalization in the case of Chile was the radically conservative nature of the military regime led by General Pinochet, which aimed at erasing all and any trace of the policies adopted before. As it has happened in similar experiences, strong liberalization policies created new profitable opportunities for banks which raised their competitiveness. However, financial regulation and bank supervision were deficient either because regulators lacked experience with open markets or because the state was assumed to be an inefficient player in the economic game so no investment in upgrading the skills of regulators and supervisors was made. Inevitably, as has been the general experience, this first wave of liberalization ended up generating a profound banking crisis in the early 1980s. To resolve the crisis, the government intervened heavily in the banking system. On the one hand, banks were allowed to sell to the government their non-performing assets under the obligation of buying them back over time, when the crisis was expected to be over. In addition, tougher bank regulation was adopted to prevent the disorderly expansion of the past from repeating itself.

In the case of Mexico, banking reforms were inspired by less dramatic events (see Avalos and Trillo, 2006; Singh, et al., 2005; and Stallings and Studart, 2006). Mexico had also followed the general pattern set by the largest economies of Latin America in the post-war period of creating strong state-owned banks to stimulate economic development. Room for private banks was very limited and foreign banks were all but banned from operating in the domestic markets. As late as in the early 1980s, foreign banks were still prevented from controlling more than 7 percent of the net worth of the largest banks. The 1982 debt crisis, the ensuing period of economic stagnation, and the conditionality clauses included in the rescue packages negotiated by the Mexican government with creditor banks and multilateral institutions led the Mexican authorities to a change of heart. The government endeavored to promote liberal reforms in the economy, of which banking reform was an important element (see de Vries, 1987). Later, this drive was strengthened by Mexico's adhesion to NAFTA which led to a drastic reduction of barriers to entry to American and Canadian banks. The defining act of Mexico's reforms, however, was the bungled privatization process of 1991, which took place when there were still strong restrictions against foreign participation in the domestic banking sector. Banks were acquired by businessmen inexperienced in the banking business, at prices widely considered to be excessive. The rush to recover their investments and to obtain profits led to a credit boom unrestrained by any kind of proper regulation. Credit was expanded without any attention being given

to credit risks. The fast expansion ultimately caused the 1994 crisis, when bank assets were virtually re-nationalized. In fact, the Mexican government, first in 1995, and again in 1996, bought the huge amount of non-performing assets in banks' balance sheets through a crisis resolution entity created to manage the problem (Fobaproa).[2] Contrary to what was done in Chile, however, those assets were not to be reabsorbed by the banking system; rather, taxpayers' money paid for the losses of banks, since Fobaproa's liabilities were transformed into public debt. The weakness of the banking system led the Mexican government to change the law to allow an increasing participation of foreign banks in domestic markets, including the acquisition of local problem banks. Consequently, the market share of foreign banks in Mexico was over 80 percent in 2000 (Hernandez-Murillo, 2007: 416).

In Brazil and Argentina, the causes of the liberalization process were somewhat more complex, owing to persistently high inflation. In both cases, most (but not all) reforms were adopted as elements of price-stabilization strategies. Until the 1970s, the Brazilian banking system was highly repressed (Carvalho, 1998). Although the presence of private banks was strong, the system was dominated by state-owned institutions. Foreign banks were confined to attending mostly foreign companies, and, as in other countries, prevented from reaching domestic clients (Carvalho, 2000). In the mid-1960s, the structure of the Brazilian financial system had been changed, and a segmented market model, similar to the one set by the Glass-Steagall Act in the US, was imposed. Commercial banks would provide short-term credit and payment services, investment banks should help develop an incipient securities market, specialized institutions would finance the acquisition of durable consumption goods, and public institutions would give financial support to productive investments in manufacturing, agriculture, and construction.

Thanks to loopholes in the legislation, financial conglomerates, with interests in practically all segments of the financial system, and in non-financial sectors as well, emerged in the 1970s and early 1980s. In parallel, the acceleration of inflation after the oil shocks of the 1970s steadily reduced the access of private borrowers to credit markets. Banks were increasingly devoting the resources they controlled to buying public debt issued by the Federal government, unable as the latter was to control its fiscal deficits. Market segments other than deposit taking and public debt buying, and the institutions supposed to operate them, gradually faded and disappeared. Under these circumstances, in 1988, the Central Bank of Brazil passed a resolution adopting the German-type universal banking model in the place of the aforementioned segmented model.[3] In the same resolution, interest rate controls were lifted. Financial liberalization in Brazil, therefore, began as the result of the

[2] The ratio of non-performing loans to total loans is estimated to have reached 52.6% by December 1996 (Hernandez-Murillo, 2007: 421).

[3] Universal banks are called *multiple* banks in Brazil.

acknowledgment that past regulations had become obsolete rather than being the first step of a well-defined strategy.

In Argentina, similar, up to a point, developments took place in the same period.[4] Accelerating inflation, as in Brazil, was the most important problem faced by policymakers at the time. In the late 1980s, the arsenal of instruments to control inflation was fast being depleted, after many failed attempts at price stabilization. Moreover, foreign creditors were demanding implementation of financial liberalization policies as a conditionality clause in the resolution package for the debt crisis of 1982. The Argentine government had little choice but to begin a liberalization process, by freeing interest rates and moving toward a universal bank model, leaving to each financial institution the choice of sectors where to operate.

After 1991, with the adoption of the Convertibility Plan (also known as the Cavallo Plan, named after then Finance Minister Domingo Cavallo), in contrast to the more pragmatic Brazilian experience, a radical liberalization strategy was put in place. A central element of this strategy was the opening of the domestic banking market to foreign banks. As a result, foreign penetration of the Argentinean banking system increased dramatically as it was deliberately promoted by a restructuring and concentration policy, which had been implemented after the contagion of Mexico's Tequila crisis that severely tested both the Convertibility system and the financial sector. Among the ten largest banks in Argentina in December 2000, seven banks were foreign owned, two were publicly owned—the market leaders, Banco de la Nación (Federal) and Banco de la Provincia de Buenos Ayres (provincial)—and only one bank was domestic, privately owned (Paula and Alves, 2007: 97).

The process of privatization of state-owned banks in Argentina illustrated an important change of views that had already taken place in countries, such as Chile and Uruguay. In these cases, privatization was seen not only as a temporary convenience or an unavoidable evil. Liberalization was adopted as a *strategy*, rather than as an expedient. Bank privatization was conducted as an element, no matter how important, of an overall liberalization process that was expected to help the region to overcome its long-term inefficiencies. The deep crisis of the early 2000s led Argentina partially to repudiate this view. It is still dominant in Chile and Uruguay, even after center-left administrations were elected in the latter countries at the beginning of the new century.

In Brazil, in contrast, this path was explored with caution. In fact, the end of inflation in 1994 caused severe stress in a large number of banks that earned their profits mostly from securing deposits to finance the purchase of public debt, the yield of which was indexed to the rate of inflation. When inflation fell precipitously, after the implementation of the Real Plan in 1994, many banks were revealed to be

[4] Decisions concerning financial liberalization in Argentina since the late 1980s are listed (in Portuguese) in Studart and Hermann (n.d.) and are reproduced and discussed in Carvalho (2008). For an overview of the process, see O'Connell (2005).

practically bankrupt. The Brazilian government, to avoid panic, took measures to allow splitting problem banks in two parts: a 'sane' one, with healthy assets and its corresponding share of liabilities; and the failed one, with the non-recoverable assets. The sane part was to be sold to other banks; the failed part would be liquidated by the Central Bank.

The same scheme, actually inspired by rules used in the US to deal with the Continental Illinois Bank in the mid-1980s, was adopted in Argentina (De la Torre, 2000). In Argentina and Brazil, panic was avoided, at the cost of pushing bank consolidation forward. In Brazil, the Central Bank decided to invite foreign banks to buy domestic banks that were either being privatized (the banks owned by the States) or facing difficulties that would probably lead them to fail. The decision to allow foreign banks in was made to prevent excess concentration, which was expected to ensue should the leading domestic banks be allowed to buy problem banks. Although, the Brazilian government never lifted the legal restrictions banning the entry of new foreign banks in the domestic system, it allowed 'exceptions' to take place while they were needed. Once the economy was stabilized and the stock of problem banks was sold, practically no new foreign bank was authorized into the country. Mexico was the last large country in Latin America to open its market to foreign banks. However, it is also the country where foreign banks were granted the most unrestrained access to domestic markets, leading to an almost complete disappearance of domestic private banks, let alone state-owned banks.

The recent trend toward consolidation is not new to the region. Previously, waves of bank consolidation had taken place in some countries, mostly induced by domestic policies. In Brazil, for instance, in the early 1970s, a strong consolidation process was promoted by the Federal government under the expectation that taking advantage of supposedly strong economies of scale would allow the reduction of interest rates necessary to keep the economy growing as rapidly as it was. Financial repression was still in force, and no increase of foreign participation was envisaged. Increasing efficiency via scale economies should lighten the burden of interest rate control on banks, attenuating the incentives to evade these controls. In any case, Latin American economies are still relatively small. If to the small dimension of these economies one also adds the generally high degree of income concentration, markets for banking services would be even smaller. If scale economies exist in banking, one would expect to find a relatively high degree of concentration in the region anyway.

In the 1980s, and more so, in the 1990s, the push for consolidation came from many sources. Political and ideological factors were very important in the case of Chile in the mid-1970s, to allow banks to decide their own policies, including larger and stronger banks to absorb smaller ones. After the early 1980s crisis, the push for consolidation was strengthened by the assumption that larger banks, especially foreign ones, are capable of managing risk more efficiently, specially if prudential regulation was improved, thereby making the system more stable.

Concerns with systemic stability help to explain consolidation, in one way or another, in nearly all of the region's recent experiences. Many of the regulatory initiatives adopted to strengthen the stability of banking systems contributed to push consolidation forward. The introduction of modern payments systems, the increasing use of ATMs, Internet banking and so forth, also lead to increased consolidation if individual banks have to provide their own equipment and other facilities. Even privatization initiatives were frequently defended with systemic safety arguments, on the notion that state-owned financial institutions increase the risk of feeding dangerous forms of crony capitalism. Consequently, at the turn of the millennium, the most important banking systems in Latin America came to exhibit a relatively similar ownership structure.

Financial penetration in Latin America

Latin American financial systems are characterized by similar features: financial depth is limited; financial sectors are bank-based since stockmarkets are mostly small and illiquid and corporate debt markets even more so; intermediation margins are high by international standards; banking sector concentration has increased; and bank lending is low relative to overall economic activity. Indeed, the limited access to bank credit and uncertainty about financial stability are factors that have contributed to economic volatility in the region (Singh, et al., 2005: ch. 1).

Whereas Latin American financial systems are deeper than they were a decade ago (Rojas Suarez, 2007: 3), the level of financial depth is low compared to industrialized countries and some emerging market regions (see Table 34.3). There is considerable heterogeneity in financial depth in the region: financial penetration is deeper in Chile and Uruguay (which operates as an offshore financial center) and the largest markets—Argentina, Brazil and Mexico—exhibit only modest levels of financial depth (see Table 34.4). Chile is the only country to have achieved a level of deepening comparable with industrialized countries (Rojas Suarez, 2007). Chile's stockmarket depth (measured as the ratio of stock-market capitalization to GDP) is greater than Japan's and some European countries such as France, Germany, and Spain (Betancour, De Gregorio, and Jara, 2006). Furthermore, households' access to financial services in Chile is closest to levels observed in industrialized countries; over 90 percent of households have access to financial services in western industrialized countries compared with 60 percent to 80 percent in Chile, 40 percent to 60 percent in Brazil and Colombia, and 20 percent to 40 percent in Argentina and Mexico (Honohan, 2007).

If financial depth and access to financial services are to increase, the institutional environment which conditions the effective operation of financial intermediaries and financial markets must be developed further. The World Bank Governance Indicators show an improved level of governance in Brazil, Chile, and Mexico

Table 34.3. Financial depth by region, 1990s

Region	Number of countries	Credit to private sector (% of gross domestic product)	Credit and market capitalization (% of gross domestic product)	Gross domestic product per capita, 1995 (US$)
Developed countries	24	84	149	23,815
East Asia and the Pacific	10	72	150	2,867
Middle East and North Africa	12	43	80	4,416
Latin America and the Caribbean	20	28	48	2,632
Eastern Europe and Central Asia	18	26	38	2,430
Sub-Saharan Africa	13	21	44	791
South Asia	6	20	34	407

Note: Values are simple averages for the regions for the 1990s.
Source: Inter-American Development Bank (2005: 5) with data from International Monetary Fund and World Bank.

Table 34.4. Financial depth in some selected countries of Latin America, 2003 (percentage of GDP)

Country	Banking system*			Outstanding domestic debt securities
	Deposits	Loans	Assets	
Argentina	25.3	14.2	44.8	25.2
Brazil	30.6	21.5	74.6	47.2
Chile	38.1	68.5	79.8	52.9
Colombia	19.7	19.7	37.9	—
Ecuador**	16.8	15.1	22.0	24.8
Mexico	25.5	16.1	52.3	13.9
Paraguay	24.6	17.6	31.7	—
Peru	14.5	13.7	19.2	7.1
Uruguay	36.4	64.3	82.6	—
Venezuela**	20.0	8.1	23.9	1.3
Latin America***	25.2	25.9	46.9	24.6

Notes: In 2003, the supply of credit decreased sharply due to the effects of economic crises in Argentina and Brazil (with a greater decline in Argentina due to the crisis of the convertibility system). Indeed, the credit to gross domestic product ratios of Argentina and Brazil stood at 21.4% and 24.8% in 2000 (see Belaisch, 2003: 4), which was greater than in 2003, but still very low compared to industrialized countries and some other areas of emerging market countries.
* Only deposit taking commercial banks are considered.
** Domestic debt securities data are as of 2000.
*** Mean values.
Source: Singh, et al. (2005: 64).

between 1996 and 2004, but only Chile achieved a level comparable with industrialized countries (Rojas Suarez, 2007: 25).

The effectiveness of financial liberalization in Latin America may be gauged from the evolution of interest rate spreads. Weaknesses in the institutional environment are offered as a partial explanation for the relatively high, by international standards, spreads observed and the dispersion of spreads across the region (Gelos, 2006). Although spreads have narrowed recently, the continued presence of high spreads has limited the benefits of liberalization.

Across Latin America, credit is not only scarce but costly too. Comparatively speaking, the region has one of the highest interest margins in the world (8.5 percent), above East Asia and the Pacific (5.1 percent) and the developed countries (2.9 percent), yet slightly lower than Eastern Europe and Central Asia (8.8 percent). Table 34.5 points to the negative correlation between private sector credit and interest spread (Singh, et al., 2005: 5–7).

Figure 34.1 shows the evolution of interest rate spreads in Latin America from 1993 to 2006. Salient features include the narrowing of spreads over time—although remaining high by international standards—and considerable cross-country variation. Spreads are largest in Brazil (Brazil has some of the highest short-term interest rates in the world), Uruguay, and Peru. Chile has the narrowest spread, comparable with industrialized countries. After 2004, it appears that spreads began to converge (except Brazil, Uruguay, and, to a lesser extent, Peru). Spreads are correlated more with loan rates than deposit rates (especially in Argentina and Peru) meaning that a shock that causes spreads to widen will raise

Table 34.5. Interest spread and efficiency by region, 1995–2002

Region	Number of countries	Interest margins (%)	Overhead costs (% of assets)	Credit to private sector (% of gross domestic product)
Developed countries	30	2.9	1.8	89
East Asia and the Pacific	16	5.1	2.3	57
Middle East and North Africa	13	4.0	1.8	38
Latin America and the Caribbean	26	8.5	4.8	37
Eastern Europe and Central Asia	23	8.8	5	26
Sub-Saharan Africa	32	10.6	5.1	15
South Asia	5	4.6	2.7	23

Note: Values are simple averages for the regions for the 1990s.

Source: Inter-American Development Bank (2005: 7) with data from International Monetary Fund and World Bank. Authors' calculations with data from International Financial Statistics.

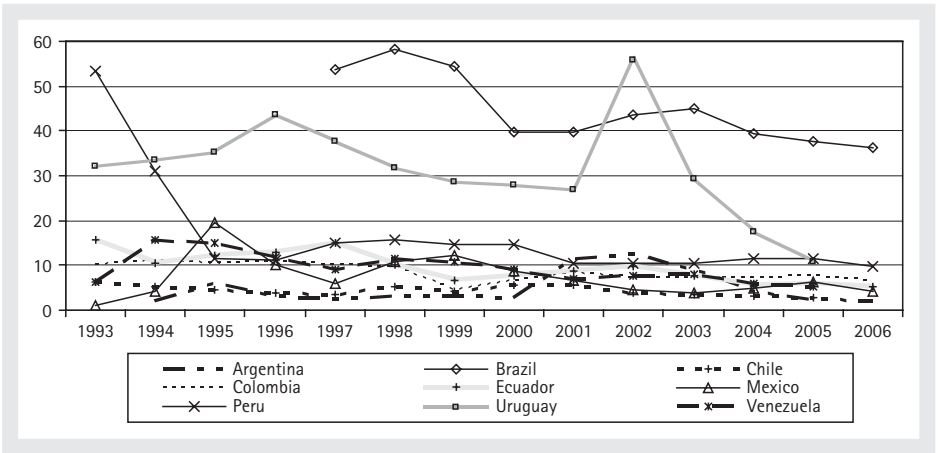

Fig. 34.1. Banking spread (percentage per annum)

Notes: In Figure 1, the banking spread (*ex ante* spread) is calculated as the difference between the average lending rate and the average deposit rate, that is, the measurement of the *ex ante* spread, while net interest (*ex post* spread) is a measurement of the net yield of bank financial intermediation, according to the revenues actually generated by credit operations and the actual cost of deposit taking, normally calculated from accounting data made available by the bank itself.

lending rates rather than decrease deposit rates. Finally, spreads are more dispersed across banks than over time (Brock and Rojas Suarez, 2000).

One should exercise caution when interpreting the narrowing of spreads. Under competitive conditions with weak bank regulation and supervision, explicit government guarantees, and the absence of political will to liquidate failing banks, poorly performing banks could have raised deposit rates but not passed on the higher funding costs to borrowers for fears that higher loan rates may raise default risk for risky borrowers. In addition, these banks may have tried to grow market share by expanding loans to risky borrowers (Brock and Rojas Suarez, 2000; and Rojas Suarez, 2001). This suggests there may be an inverse relationship between interest rate spreads and banks' portfolio risk in Latin America, which, if true, is contrary to the observed relationship in industrialized countries. It implies that stronger banks served the 'better-quality' customers meaning that poorly capitalized, weaker banks have had to operate with lower spreads in order to compete. Furthermore, weaknesses in provisioning caused spreads to narrow when the loan portfolio (and bank income) deteriorated (Brock and Rojas Suarez, 2000).

Factors influencing bank spreads include microeconomic factors (high operating costs, poor loan quality, high capitalization, and reserve requirements) and macroeconomic factors (interest rate volatility, GDP growth, and inflation). Empirical evidence suggests that microeconomic factors have been the main determinant of spreads in Bolivia; micro- and macroeconomic factors impacted on spreads in Chile and Colombia (Brock and Rojas Suarez, 2000; Barajas, Steiner,

and Salazar, 1998; and Barajas, Steiner, and Salazar, 2000); macroeconomic factors were more important in determining Brazilian spreads that were particularly high at large banks due to market power (Afanasieff, Lhacer, and Nakane, 2002); neither micro- nor macroeconomic factors adequately explained the evolution of spreads in Argentina and Peru (Brock and Rojas Suarez, 2000). A comparison of spreads in Latin America with the industrialized countries found that the main difference is the effect that non-performing loans have had on spreads, with deteriorating loan quality associated with narrower spreads in Latin America. It is suggested that this feature was indicative of inadequate loan-loss provisioning, or that banks with large amounts of bad loans lowered spreads in an attempt to grow out of difficulties (Brock and Rojas Suarez, 2000).

Besides the generally low level of credit identified in Table 34.3, the pattern of credit growth in Latin America has been marked by boom and bust cycles, particularly in economies with the lowest amounts of bank credit to GDP. Credit had expanded sharply across the region in the early 1990s, in part due to increased capital inflows, but it collapsed in many cases after the mid-1990s banking crises and remained subdued for many years. Only after 2004 has credit begun to recover, due to stronger economic growth, easier global monetary conditions, and progress in bank restructuring. Credit growth has been particularly strong in Argentina and Brazil (Jeanneau, 2007: 6–7). Yet, in most Latin American countries, the unstable macroeconomic environment has been a critical factor in holding back financial system development and generating a high volatility of credit growth. For example, high short-term interest rates used to combat inflation or defend the exchange rate has added to banks' funding costs and increased loan-default rates (Singh, et al., 2005: 70–1).

There are legitimate concerns that the composition of bank portfolios probably led to some crowding out of private sector credit. This was because of banks' tendencies to hold high proportions of government securities in their portfolios, which possibly reflected historical patterns of behavior associated with hyperinflation. In the late 1990s, banks replaced non-performing loans with sizable portfolios of government securities in Argentina, Mexico, and Venezuela. More recently, and due to fiscal consolidation (in Argentina, Mexico, and also Brazil), the amount of government securities in banks' portfolios has declined.

Banking sector heterogeneity and dollarization

The observed heterogeneity across Latin America's financial systems results from a variety of different historical and institutional features. Financial sector penetration (depth) of the economy is highly variable but unrelated to country size and per capita incomes. The relatively large economies of Argentina and Mexico have smaller banking sectors than implied by their levels of economic development,

which can be attributed to long-lasting effects of financial crises. As a result of the 1990s Tequila crisis, the ratio of Mexican bank assets to GDP fell from a historical high of almost 70 percent in 1994 to between 32 percent and 35 percent from 2000 to 2005 (Sidaoui, 2006: 287). Estimates imply that private sector credit should be around 50 percent of GDP given the economic size of Argentina, rather than the observed 1990s average of 20 percent (Inter-American Development Bank, 2005: 6). Uruguay has one of the most internationalized and open financial systems in the region with the banking sector performing the role of regional offshore financial center. Owing to its longer track record of greater macroeconomic stability, sustained economic growth, and earlier financial sector reform, Chile has achieved a more even pattern of credit growth.

Latin American financial systems are characterized by varying degrees of dollarization with the conspicuous exception of Brazil and Venezuela (see Singh, et al., 2005). In several countries, relatively large shares of bank deposits and loans have been denominated in US dollars: the average dollarization ratio across 1998 to 2004 shows that over 75 percent of total banking sector deposits were foreign currency denominated in Bolivia, Peru, and Uruguay whilst the ratio was approximately 60 percent for Paraguay (see Table 34.6). In these countries, informal dollarization has been developed partially as a response to the hyperinflation of the 1980s (in Bolivia and Peru), when confidence in the value of domestic currencies was severely undermined.

Table 34.6. Dollarization ratio* in selected countries in Latin America (percentage)

Countries	1998	2000	2001	2004
Argentina	58.4	66.6	2.9	11.0
Bolivia	93.1	93.8	92.1	90.5
Brazil	0.0	0.0	0.0	0.0
Chile	6.2	0.0	11.5	13.0
Colombia	0.0	0.0	0.0	0.0
Costa Rica	44.4	44.9	48.0	48.0
Ecuador	—	—	100.0	100.0
Mexico	8.0	5.6	4.7	3.4
Panama	100.0	100.0	100.0	100.0
Paraguay	47.5	61.6	68.5	61.9
Peru	76.5	76.9	73.2	68.9
Uruguay	90.6	91.6	93.6	90.0
Venezuela	0.0	0.1	0.2	0.1
South America	21.4	23.2	27.5	27.0

Note: * Total foreign currency deposits in the domestic banking system/total deposits in the domestic banking system.

Source: Bank for International Settlements (2007: 68).

In Ecuador, full dollarization was implemented for price stability purposes in 1999. Elsewhere, economic policies that stimulated the dollarization of the domestic economy were adopted. In 1991, Argentina implemented its currency board regime that guaranteed the full convertibility of dollars and pesos; financial intermediation increasingly became dollar-denominated until the regime collapsed in 2002. In contrast, other countries (Brazil, Chile, Colombia, Mexico, and Venezuela) have avoided dollarization, by either prohibiting most holdings of foreign currency deposits, or imposing prudential constraints on such holdings. Prohibition has had the adverse effect of shifting deposits and loans offshore; consequently, financial-system vulnerability increased because of greater liquidity and solvency risks (Jeanneau, 2007: 7–8).

The effects of banking consolidation

Market structure, privatization, foreign bank penetration, and bank performance

Bank privatization of state-owned banks dramatically altered the market structure of domestic banking sectors.[5] Privatization has transformed the governance structure of domestic banks as new, private owners (domestic and foreign) assumed control of banks. Generally speaking and across the region, state-owned banks had served political and social purposes and they shared certain characteristics: weak loan quality, underperformance, and poor cost control. Indeed, privatization was deemed to be a cheaper option than restructuring and recapitalization. The outcomes of bank privatization have varied across countries. For Argentina and Brazil, the evidence suggests that privatized bank performance improved

[5] In Argentina in the early 1990s, each province had its own bank which often dominated the local financial sector. Provincial state-owned banks held between 40% and 70% of banking sector assets in each province. In 1993, 25 provincial banks held an assets market share of 21.6% of the Argentinean banking sector with a further 23.3% of assets held by nine state-owned national and municipal banks. By 1999, the remaining 10 provincial banks and five national and municipal banks held 13% and 18.5% of total assets, respectively (Clarke, et al., 2005). In Brazil in 1994 there were 32 public sector banks holding 51.53% of total banking sector assets. In 2002, 14 public sector banks held market share of 35.01% (Nakane and Weintraub, 2005). In Colombia, state-owned banks held a 55% asset market share in June 1991; following liberalization this share stood at 10.3% in June 1998 (Barajas, et al., 2000). The Mexican authorities responded to the 1982 debt crisis by nationalizing the banking sector and implementing a restructuring program involving M and A. By 1991, at which time it had been decided to return the 18 state-owned banks to private ownership, they controlled around 70% of banking sector assets (Montes-Negret and Landa, 2001).

post-privatization (Berger, et al., 2005 for Argentina; Nakane and Weintraub, 2005 for Brazil). In stark contrast, the 1991 privatization program in Mexico failed in the mid-1990s with the onset of the Tequila crisis.[6] The crisis revealed deep-seated problems in the banking sector which had been masked by weak property rights and ineffective bank regulation that failed to prevent imprudent behavior by newly privatized banks. Bank privatization failed to the tune of a bailout costing an estimated $65 billion (Haber, 2005). Yet, unlike in Argentina and Brazil, the 1991 Mexican program disbarred foreign banks from entering the auctions. Beginning in February 1995, a post-Tequila second round of restructuring and privatization liberalized the treatment of foreign ownership of domestic banks and was completed in 1996 (with effect from 1997). This lead to a large-scale transfer of bank ownership from domestic to foreign hands: foreign banks held 5 percent of banking sector assets in 1995 that leapt to 82 percent by 2003 (Haber, 2005). In 2007, two of the three largest Mexican banks were owned by foreign banks.

One must be cautious when interpreting the apparent positive outcome of bank privatization. The observed post-privatization improvements in bank performance may reflect selection bias. In order to raise the viability of state-owned banks to prospective buyers, bank balance sheets were sanitized and healthy banks were privatized whilst bad banks were funded using public funds (for further details on the privatization of Argentina's provincial banks, see Clarke and Cull, 2000). Certainly, statistically significant differences in the balance sheet structures of privatized and non-privatized state banks are reported for Argentina (Berger, et al., 2005). Similar transfers were carried out in Brazil. The utilization of the bad bank model can be expected to have influenced post-privatization bank performance.

Bank privatization assisted foreign bank penetration in Latin America as foreign banks acquired large, domestic banks. For policymakers, foreign bank entry was expected to raise competition leading to efficiency gains and banking sector recapitalization. Foreign bank entry increased banking sector capitalization in Mexico between 1997 and 2004 by more than $8.8 billion—equivalent to 42 percent of total banking sector capital in 2004 (Schulz, 2006). Country-level evidence suggests bank efficiencies improved at the same time as foreign bank penetration increased. Arguably, this is too general a claim since there are caveats to consider. First, one should distinguish between the performance of existing foreign banks and domestic banks acquired by foreign banks—mainly large banks purchased via cross-border bank M&A. We refer to the latter as foreign bank acquisitions. Second, it is difficult to disentangle the effects of foreign bank entry from other

[6] For details of the new owners of the 18 privatized banks, see Hoshino (1996). The bank auction process is described by Haber (2005). It is suggested that some new bank owners lacked suitable experience and that this was a contributing factor to the problems which befell the banking sector in the mid-1990s (Hoshino (1996); Montes-Negret and Landa (2001); Haber (2005)).

liberalization effects that could have affected bank efficiency. Finally, many studies use proxy measures of efficiency like the ratio of overhead costs to assets; there is limited evidence where econometric estimates of bank efficiency were employed (Berger, 2007).

One exception reports there were inter-country differences in bank cost efficiencies with very small and very large banks more inefficient than large banks. Cost inefficient banks tended to be small, undercapitalized, relatively unprofitable, less risk averse, facing unstable deposit bases and intermediating less. Country-level factors also determined bank-level cost efficiencies: countries with higher rates of economic growth, denser demand for banking services, and lower levels of market power achieved better cost efficiency performance (Carvallo and Kasman, 2005).

It is very difficult to identify the separate effects of bank privatization and foreign bank entry on bank condition and performance. One study reported little difference in the performances of privately owned domestic banks and foreign-owned banks, though the former did outperform state-owned banks (Crystal, Dages, and Goldberg, 2002). Foreign banks achieved higher average loan growth than domestic banks (in Argentina, Chile, and Colombia) with loan growth stronger at existing foreign banks compared to acquired foreign banks. It is suggested that management at foreign bank acquisitions focused on restructuring the former domestic banks and integrating operations with the parent (foreign) bank. This implies that foreign bank acquisitions adopted a defensive strategy toward market share and growth until the integration process was completed. The cautious nature of foreign bank strategies explains why foreign banks, and foreign bank acquisitions in particular, had better loan quality than domestic-owned banks, although stronger provisioning and higher loan recovery rates translated into weaker profitability at foreign banks. Foreign banks were relatively more liquid, have relied less on deposit financing, and realized stronger loan growth during episodes of financial difficulty than domestic banks. The available evidence suggests that foreign banks have achieved a greater efficiency in intermediation because they were better able to evaluate credit risks and allocate resources at a faster pace than their domestic-owned competitors (Crystal, Dages, and Goldberg, 2002).

In Argentina, foreign banks typically entered the market via cross-border M&A rather than 'de novo' entry. The targets of foreign banks tended to be the larger and more profitable domestic banks. On average, foreign banks achieved better loan quality and were more highly capitalized and profitable than domestic banks (Clarke, et al., 2005). The effects of the governance changes on bank performance are summarized by Berger, et al (2005): state-owned banks underperformed against domestic and foreign banks due partly to poor loan quality associated with directed lending and subsidized credit. The privatization of provincial banks realized efficiency gains as the amount of non-performing loans fell and profit

efficiencies increased. However, the improvement in profit efficiency may simply reflect selection bias since cost efficiencies were consistent before and after privatization. M&A activity involving domestic banks and foreign bank entry were reported to have had little effect on bank performance (Berger, et al., 2005).

These findings do not generalize to Brazil. Foreign banks operating in Brazil have faced difficulties in adapting to the peculiarities of the Brazilian banking sector, which remains dominated by private domestic banks (Paula, 2002). Incidentally, the empirical record offers no support for the hypothesis that foreign banks are either more or less efficient than domestic banks (Guimarães, 2002; Paula, 2002; and Vasconcelos and Fucidji. 2002). This is unsurprising in the light of evidence that the operational characteristics and balance sheets of domestic and foreign banks are similar (Carvalho, 2002). Hence, the expected benefits of foreign bank entry have yet to materialize in Brazil, because foreign banks have witnessed and graduated toward similar operational characteristics of the large private domestic banks (Paula and Alves, 2007).

Market concentration and competition effects

The recent consolidation process has increased concentration in Latin American banking sectors. Whereas the expectation of policymakers has been that higher concentration would lead to more competition and efficiency improvements, there was the possibility that competitive gains would not materialize, and instead bank market power would increase. The latter implies that the evolution of highly concentrated market structures could limit the deepening of financial intermediation and the development of more efficient banking sectors (Rojas Suarez, 2007). Since a non-competitive market structure often produces oligopolistic behavior by banks, the suggestion is that further consolidation could incentivize banks to exploit market power rather than become more efficient.

It is an empirical matter to determine if bank consolidation (more concentration) has raised competition and banking sector efficiency, or instead realized market power gains for banks. Some degree of market power can check bank risk taking, and there are trade-offs between increased competition and financial stability. One difficulty when considering the relationship between consolidation and competitive conditions is the measurement of competition. The literature commonly employs the H statistic showing the sum of the elasticities of bank revenue with respect to input prices (Panzar and Rosse, 1987). Using this approach, banks in Latin America were found to operate under monopolistic conditions, consistent with results from industrialized countries and other emerging markets.

Importantly, the recent increase in consolidation has not weakened competitive conditions (Yeyati and Micco, 2007; Yildirim and Philippatos, 2007; and Gelos and

Roldós, 2004). Despite this general finding, there are country-level features of note and some inconsistencies between studies. For instance, there is agreement that banking sector competition increased in Argentina and remained constant in Mexico from the mid-1990s to the early 2000s. On the contrary, competitive conditions in Brazil and Chile are reported to have changed little (Gelos and Roldós, 2004) or weakened (Yildirim and Philippatos, 2007).

In general, the literature rejects the notion of collusion between banks, but evidence from Brazil suggests that banks possessed some degree of market power (Nakane, 2001; and Nakane, Alencar, and Kanczuk, 2006). Other Brazilian evidence has illustrated the complexities associated with identifying competition effects. Whereas the banking sector has operated under conditions of monopolistic competition, this finding cannot be generalized across bank ownership and size. Whilst small banks and state-owned banks operated under the above banking sector conditions, large banks and foreign banks behaved competitively. This implication is of markedly different competitive conditions in local markets and the national market (Belaisch, 2003). In local markets, privately owned banks were more pro-competitive than state-owned banks, although the latter entered markets which the former did not service (Coelho, de Mello, and Rezende, 2007). Small and large banks have also faced different competitive conditions in Argentina and Chile (Yildirim and Philippatos, 2007). Lastly, evidence from Colombia finds increased competition reduced banks' market power (Barajas, Steiner, and Salazar, 1998).

At first sight, greater competition brought about by the changes in the banking industry in the 1990s has not weakened bank safety. As one can see from Table 34.7, capital coefficients have been comfortably above the Basel 1988 minimum of 8 percent of risk weighted assets everywhere. Even if one considers the higher floor of 11 percent, as suggested in the twenty-five Core Principles for Effective Banking Supervision, all countries in the Table would be in compliance.

One should read these data with some care, though. It is true that the 1990s and the 2000s witnessed a widespread effort at modernization of regulatory and

Table 34.7. Bank regulatory capital-to-risk-weighted assets

Country	2002	2003	2004	2005	2006
Brazil	16.6	18.8	18.6	17.9	18.9
Chile	14.0	14.1	13.6	13.0	12.5
Colombia	12.6	13.1	13.8	13.2	12.2
Mexico	15.7	14.4	14.1	14.5	16.3
Peru	12.5	13.3	14.0	12.0	12.5
Venezuela	20.5	25.1	19.2	15.5	14.3

Note: No data for Argentina is provided.

Source: International Monetary Fund, *Global Financial Stability Report* (September 2007).

supervisory methods and institutions everywhere in the region. Nevertheless, in part the data may be hiding one important source of fragility which is the dependence of the banking industry, at least in some of the largest economies, on the supply of credit to the government. Public debt securities tend to benefit from zero risk weighting (and thus do not require any capital to cover credit risk) adding one more incentive to banks to accumulate them, instead of private credit. As a result, in countries like Argentina, Brazil, or Mexico, high capital coefficients may not necessarily translate into higher defences against insolvency, but, in fact, to higher dependency on Treasury policies.

In the discussion so far, no attempt has been made to disentangle the impact of foreign bank entry on competition. A priori greater foreign bank penetration was expected to increase competition and to offset the potential rise in domestic bank market power resulting from higher concentration. Consistent with expectations, there is cross-country evidence that suggests the increased foreign bank penetration raised the level of competition (Yildirim and Philippatos, 2007). An alternative view claims that increased concentration had little effects on competition and financial stability. Rather, foreign bank entry caused competitive conditions to weaken (Yeyati and Micco, 2007). The intuition for this claim is that foreign banks typically acquired domestic banks that were under duress and consequently operating with relatively high interest margins. For new foreign owners, the franchise value of high margins and the time needed to transform the fortunes of their acquisitions can explain why increased foreign bank penetration was associated with weaker, rather than stronger, competition. Whilst this apparent feature was inconsistent with policymakers' objectives, the franchise value of the high margins disciplined banks' risk taking because of fears that the increased bank profitability of the period could be dissipated away. In short, although foreign bank entry may have weakened competition it appears to have had a beneficial effect on banking sector stability (Yeyati and Micco, 2007).

Foreign bank entry raises the threat of increased competition which conditions the behavior of domestic banks and reduces their market power (Claessens, Demirgüç-Kunt, Huizinga, 2001). The evidence suggests this has happened in Latin America: greater foreign bank penetration has caused lower interest margins and profits at domestic banks (Yildirim and Philippatos, 2007). Individual country studies offer a richer interpretation of events. Evidence from Colombia suggests that foreign bank and domestic bank behavior began to evolve differently following the announcement (in 1990) that financial liberalization policies were to be implemented (Barajas, Steiner, and Salazar, 2000). This study was able to control for other liberalizing reforms that affected bank behavior—for instance, it differentiated between foreign bank entry and the entry of new domestic institutions, and it controlled for the opening of the capital account as well as improvements made to bank regulation and supervision. Whereas foreign bank entry did condition domestic bank behavior by reducing excess intermediation spreads over

marginal costs, the effect of new domestic entrants on bank behavior was greater, reducing non-financial costs and interest spreads. The Colombian evidence implies bank behavior reflected the degree of market power of banking groups; since foreign banks had relatively little market power they were more able to adapt to changes in competitive conditions (Barajas, Steiner, and Salazar, 2000).

In Mexico, the lower administrative costs of foreign banks released downward pressure on administrative costs across all banks, which improved bank efficiency (Haber and Musacchio, 2005). Others have suggested that the impact of foreign bank entry on bank efficiency was limited because the low level of competitive intensity in the banking sector abated pressures for banks to improve operational efficiency (Schulz, 2006). Evidence from Argentina and Brazil has reported there was no significant difference in the behavior of foreign and domestic banks; both types of bank reacted similarly to the macro-institutional environments (Paula and Alves, 2007).

Consolidation and the allocation of credit

The governance changes resulting from bank privatization and foreign bank penetration raised concerns in relation to the supply of bank credit. Three concerns were voiced: first, that increased foreign bank penetration would affect the stability of bank lending; second, foreign bank entry and/or new private ownership of banks might lead to a reallocation of credit toward certain geographic or product market segments; and third, given the governance changes, would bank credit be responsive to market signals.

Foreign bank penetration has raised foreign banks' share of total banking sector loans in Latin America. Foreign bank lending has tended to concentrate in specific market segments, mostly the commercial loans markets (including government and interbank sectors) in Argentina, Colombia, and Mexico (Dages, Goldberg, and Kinney, 2002; Paula and Alves, 2007; and Barajas, Steiner, and Salazar, 2000). In these countries foreign banks limited their exposure to the household and mortgage sectors. In Chile, household credit has dominated foreign banks' loan portfolio increasing from 18.4 percent to 27 percent of total foreign bank loans between 1990–9 and 2000–5 (Betancour, De Gregorio, and Jara, 2006). In Argentina and Brazil, foreign and domestic banks competed in loans markets and shared loan portfolio characteristics (Dages, Goldberg, and Kinney, 2002; and Paula and Alves, 2007). However, foreign banks in Argentina weighted the loan portfolio toward relatively less risky loans (Dages, Goldberg, and Kinney, 2002), which was not the case in Brazil where no distinction was found between interest rates charged by foreign banks and domestic banks. This gave rise to claims that variations in pricing occurred within the foreign bank and domestic bank sectors rather than between the two sectors (Carvalho, 2002).

Foreign banks have become an important source of finance for specific customer segments. Indeed, they have achieved higher loan growth (better quality and less volatile) than domestic banks (especially vis-à-vis state-owned banks) (Dages, Goldberg, and Kinney, 2002). Foreign banks—and also private domestic banks—are responsive to market signals: in particular, lending is pro-cyclical and sensitive to movements in GDP and interest rates, which is indicative of transactions-based activities. The finding of higher loan growth and lower volatility at foreign banks—even during crisis periods—implies they were important stabilizers of bank credit (Dages, Goldberg, and Kinney, 2002).

After being granted unrestricted access in 1997, foreign banks came to dominate the Mexican banking sector quicker than they had done in other countries: in 1997, foreign banks supplied 11 percent of bank credit, which grew to 83 percent in 2004 (Haber and Musacchio, 2005). During this time, a 'credit crunch' occurred and private sector lending fell by 23 percent in real terms between December 1997 and December 2003 (Haber, 2005). It appeared foreign bank penetration had altered bank lending strategies, but this was not the case because the acquired foreign banks had begun to reduce private lending before acquisition. Prior to the 1991 bank privatizations, the ratio of commercial bank loans to GDP was 24 percent and rose to 26 percent in 1996; subsequently, it declined to 14 percent in 2003 (Haber, 2005). Furthermore, the behavior of foreign bank acquisitions, pre- and post-M&A, differed little from domestic banks. In brief, the 'credit crunch' was driven by factors affecting all banks and unrelated to foreign bank entry.

In Argentina, bank privatization and foreign bank entry raised fears of a reallocation of bank lending. Initially, fears arose because the acquirers of the privatized provincial banks tended to be small, wholesale banks based in Buenos Aires who were expected to raise deposits in the provinces and allocate resources more in the centre (Clarke, Crivelli, and Cull, 2005). State-owned bank lending had been geographically diversified though concentrated more in the public sector with fewer manufacturing loans. Other concerns were that the volume of bank credit would decrease post-privatization because the transfer of non-performing loans to bad banks meant the size of the privatized provincial banks was smaller than pre-privatization (Berger, et al., 2005). Since foreign banks had mainly located in Buenos Aires and tended to finance large-scale manufacturing and utilities firms in that province, commentators questioned foreign banks' commitment to diversify lending to the provinces (Berger, et al., 2005).

Temporarily, privatization and foreign bank entry disrupted credit in the 1990s. Disruptions were most pronounced in provinces that had privatized banks; credit levels fell but quickly returned to pre-privatization levels once privatized banks increased in size. Privatization did not affect the lending of private domestic or foreign banks. For foreign bank acquisitions, lending increased in importance and as a ratio of total assets, with loans growth allocated more toward consumers than manufacturing (Berger, et al., 2005). Fears that foreign banks would concentrate

their lending in Buenos Aires did not materialize. Foreign banks entered provincial markets, aggressively in provinces which had privatized their banks. In contrast, the newly privatized banks decreased lending relative to total assets to control risk through more prudent lending (Berger et al., 2005). In summary, foreign bank penetration caused an increase in provincial lending because foreign banks offset changes in lending of domestic banks (Clarke, Crivelli, and Cull, 2005).

Consolidation and interest rate spreads

Finally, we review the effects that foreign bank penetration and market concentration have had on the evolution of bank interest rate spreads and on the process of financial intermediation. The effects were determined by comparing the spreads charged by foreign banks and domestic banks and the evidence comes from several countries (Argentina, Chile, Colombia, Mexico, and Peru). Generally speaking, foreign banks have operated with lower spreads compared to domestic banks (especially 'de novo' foreign banks), but the main impact of foreign bank penetration has been the inducement for all banks to reduce costs rather than a marked decline in spreads. Concentration, on the other hand, could offset the apparent benefit of foreign bank penetration, since higher concentration could raise operational costs and thereby widen spreads especially for domestic banks (Martinez-Peria and Mody, 2004).

Conclusion

The last two decades have witnessed deep changes in the operation of the banking sector everywhere, but without a doubt these changes have been particularly strong in Latin America. In these twenty years, financial repression was eliminated or drastically attenuated from Mexico to the Southern Cone. The role of state-owned banks was streamlined either by privatization or by increasing specialization in the provision of financial support to special groups of borrowers, such as, small and medium-sized firms, as in the case of Mexico. In a few countries, however, and most notably in Argentina and Brazil, a large sector of state-owned banks survived the financial liberalization process and went on to become leaders in their domestic banking sectors.

A common feature of the financial liberalization process in the whole region was the increasing presence of foreign banks in domestic markets. Led by US and Spanish banks, foreign institutions have aggressively taken advantage of the

relaxation of restrictions on the operation of foreign banks in practically the whole continent.

Liberalization, privatization, and foreign bank entry combined with larger macroeconomic policy changes and strategies to generate a process of consolidation in the banking sector of all countries in the region. Consolidation was actively supported by local government policies aiming at taking advantage of possible economies of scale and scope in the production of banking services. Nevertheless, the results of these efforts are still to appear more clearly, although there is some evidence of efficiency improvements in bank operations in the region.

In the major countries of Latin America, banks faced important difficulties in adapting to the new context of financial deregulation and liberalization. Serious banking sector crises took place in Chile, which pioneered the liberalization process, Argentina, also in the early stages of liberalization, and Mexico. In Brazil, banks suffered strong pressures resulting from the joint impact of deregulation and price stabilization processes in the mid-1990s, forcing the government to create a special crisis-resolution program. Argentina suffered another banking crisis in 2001, connected to the balance of payments crisis that put an end to the Convertibility Plan.

In sum, practically all of the changes in the book have been implemented in the region since the late 1980s. Interest rates are currently market-determined everywhere but Venezuela, where controls still subsist. Privatization advanced strongly everywhere, except Brazil, where the leading banks were kept in the hands of the Federal government. Directed credit was reduced or eliminated across the region, again with the partial exception of Brazil, where a Federal development bank (BNDES) is practically the only provider of long-term credit. Monetary policy in all parts of Latin America is implemented through open-market operations.

The results of the process have been relatively disappointing, given the high expectations that surrounded the liberalization process in the late 1980s. The jury is still out, of course, given the relatively short time during which these changes have been in place and the turbulence that characterized some periods in the 1990s. There is some evidence of improvement in many cases, but still not enough to generate enthusiasm. Banking crises and stresses, however, have not led to reversals in the financial liberalization process so far. On the contrary, most countries in the area have been investing in building regulatory and supervisory institutions while adhering to modern regulatory paradigms, such as the Basel accords. If the assumptions underlying the process of financial liberalization are in fact true, better results should begin to show in the short term in the form of lower cost of capital, wider access to finance, better allocation of resources, while, of course, maintaining a reasonable degree of financial stability. It is a tall order, but financial liberalization promises no less.

Epilogue: The international financial crisis and Latin American Banks

Arguably the economies and financial systems of Latin America are better placed to withstand the effects of exogenous shocks emanating from international financial market distress than at any other time in the recent past. Since 2003, Latin America has enjoyed 'an unprecedented cycle of economic growth with macroeconomic stability... while inflation has fallen and fiscal positions have improved... This period of economic growth was supported by an exceptionally favorable external financing environment' (Bank for International Settlements, 2008b: 2). In many regional economies, current account surpluses have replaced deficits, which together with increasing foreign direct investment and growth in remittances have allowed a substantial accumulation of international reserves. Consequently, external borrowing and debt have fallen, aided by important changes in financial structure—for instance, the development of local currency bond markets. In addition to a reduction in vulnerability over time, Latin American economies are relatively less vulnerable than emerging markets in Asia and Central and Eastern Europe. However, there is concern that, in spite of the improvements, Latin America is vulnerable to a slowdown in the US economy, and also to changes in the positive market sentiment the region has enjoyed since 2003 as global economic conditions worsen (Bank for International Settlements, 2008b).

The current global financial crisis initiated with the collapse of the subprime mortgage market in the US, in 2007, did not have a strong immediate impact on banking markets in the main economies in the region. Although there has been fast-paced growth in securitization in Latin America, the market is relatively nascent, although issuance of domestic asset-backed securities increased fivefold from 2003 to 2006. In 2006, the value of issues of domestic asset-backed securities in Latin America was $13.6 billion, with activity concentrated in Brazil and Mexico (40 percent and 32 percent of total), followed by Argentina (18 percent). Mortgage-backed securitization accounted for 21 percent of market activity in 2006 (Bank for International Settlements, 2008a). The scanty evidence available so far, however, suggests that Latin American banks were not significantly exposed to the US subprime mortgage market as such, contrary to what happened in Europe. As a result, they were spared the credit and market value losses that plagued banking systems in the US, Europe, and, to some extent, Asia. One of the apparent lessons of the crisis, that the survival of independent institutions, particularly those dealing with securities markets, such as investment banks, may be under threat, when compared to universal banks would not be a problem to the region, since its banking industry has long been converted to the universal bank model.

One possible source of transmission of the crisis into Latin American financial markets would be through a retrenchment of claims on the region by international banks. Though the latest official data (see Bank for International Settlements, 2008c) report nominal growth in claims by international banks on Latin America of nearly 30 percent between June 2007 and June 2008—with three-quarters of claims on Brazil and Mexico, respectively—the effects of the crisis began to be felt from the middle of 2008, and mostly in the form of a drastic cut in access to foreign credit and securities markets, which was an important source both for domestic borrowers and for banks operating locally. With this in mind, we next review developments in the region's three largest banking markets in 2008.

Argentina

Argentina's banking sector, like its Brazilian and Mexican counterparts, so far has not suffered a direct impact of the global financial crisis. After the dramatic fall of loans to the private sector during the Convertibility Plan's crisis—from around 20 percent of GDP in 2001 to 8 percent of GDP in the second half of 2003—a slow and gradual credit recovery developed in Argentina: total credit to the private sector over GDP reached 12 percent in October 2008, well above the minimum 7.5 percent of GDP registered in 2004. In October 2008, total loans to the private sector (in pesos and in foreign currency) grew 2.8 percent, accumulating an increase of 22.2 percent in the first ten months of 2008, pushed up mainly by state-owned banks (an increase of 40 percent in the same period of 2008). The credit boom, however, has been losing strength: the annual rate of growth of credit fell from 40 percent until June 2008 to 30.1 percent in October 2008. In terms of the risk of default of outstanding loans, the Argentine financial system seems to be in a reasonably safe position: non-performing loans reached 2.9 percent of total credit to the private sector in September 2008 compared to 3.5 percent in September 2007.[7]

In October 2008, the demand for foreign currency from the private sector increased (private savings in foreign currency increased, on average, by 3.1 percent (Asociación de Bancos Privados de Capital Argentino, 2008: 8), which caused a decrease in the value of peso-denominated private sector deposits of 1.6 percent. Data released by the Central Bank (Banco Central de la República Argentina, 2008), however, shows some recovery of peso deposits in November.

At least until October 2008, banks were significantly liquid. The bank-liquidity indicator—defined by the pesos in cash in banks, reserve deposits in pesos in BCRA, and reverse repos with the Central Bank as a percentage of total deposits in pesos—was, on average, 21 percent in October 2008, above the average for the last

[7] Asociación de Bancos Privados de Capital Argentino (2008) based on data from Central Bank of the Republic of Argentina.

three years, which was 18.9 percent. On the other hand, Argentine banks' Basel coefficients have been around 17 percent since the beginning of 2006, far above the minimum international standard (Banco Central de la República Argentina, 2008: 10).

In spite of the situation of sufficient liquidity at local banks, the call-money market loan rate registered an increase of 3.9 percentage points and averaged 13.1 percent in October. As a result, the cost of credit (both in wholesale and retail markets) increased along with the rates paid on deposits. This upward trend was a consequence of the increase in the rate established by BCRA for repo transactions, and also of more risk-averse behavior by banks in the context of the global financial crisis and of the risk of a prolonged global recession.

Faced with the volatility of the international financial markets, and in order to avoid bank liquidity problems due to a reduction of deposits and the increase of the cost of credit, BCRA has adopted some preventive measures: it stipulated a bimonthly period to carry out the position of minimum cash in pesos in October and November (Communication A4858); and, with the intent to lower the cost of loans, it admitted in a transitory way (as from December 2008), to take into account the totality of cash kept in financial entities (in pesos and in foreign currency) in order to integrate the minimum cash (Communication A4872).

In summary, as of the end of 2008, there were no significant credit, liquidity, and solvency problems in the Argentine banking sector. However, some concern is related to the future of the Argentine economy owing to the fall of the international demand for, and prices of, commodities in international markets. The still subdued decline in domestic industrial output is also a cause for concern. In the third quarter of 2008, the economy grew by 6.5 percent, the lowest rate of growth since 2003, and the outlook in the near future is not optimistic. Under such conditions, it is likely that banks will adopt a more pro-cyclical behavior and reducing the supply of credit for both households and firms.

Brazil

Since 2005 there was a credit boom in Brazil, in part pushed by the economic growth: total credit to private sector over GDP grew from 23.5 percent in January 2005 to 38.0 percent in September 2008. Available data show that, as of October 2008, bank credit was still growing vigorously. The monthly survey published by the Central Bank of Brazil,[8] indicated that total credit had expanded 26.8 percent in the first ten months of 2008, compared to the same period of 2007, when credit had already been growing very quickly. All three segments of the banking system expanded strongly in 2008: state banks' credit grew 30 percent, private domestically

[8] Available on its website < http:www.bcb.gov.br>.

owned banks expanded credit in 26.4 percent, and even foreign banks' credit supply grew 22.7 percent. A widespread feeling of credit rationing prevailed in the period though, leading the federal government to take many initiatives toward easing the pressure felt by borrowers. Required reserves were drastically reduced, public banks were allowed to give support to private institutions, new sources of liquidity were opened, including through the use of international reserves to support the supply of trade credit to exporters. Banks tended to explain the sensation of credit rationing by the virtual closure of domestic as well as international securities markets, which supposedly strongly increased the demands for bank credit, beyond their capacity to meet them.

Another important effect of the crisis could be said to be psychological in nature. Some measure of panic seems to have spread throughout the economy as a reflex of the news coming from the US and Europe. Middle-sized banks suffered some loss of deposits, which tended to be shifted to larger banks, particularly public banks, which always benefit in bouts of panic. It is widely expected that, as a result of such movement, a new consolidation wave will take place in the near future, as mid-sized and small banks lose access to resources and are forced to look for stronger partners to merge. In addition, the Brazilian government has proposed measures to allow public banks to buy threatened institutions. So far, however, only one important merger took place, whereby the second-largest private bank in the country, Itau, merged or acquired (the details of the deal remain fuzzy) the fourth-largest private bank, Unibanco. It is unclear, however, to what extent this merger is already a precocious answer to the crisis or just another move in the complex competitive picture of the Brazilian banking sector. Itau had long been striving to beat Bradesco, which was the largest private bank in the country for decades, and this was finally possible with this merger or acquisition of Unibanco.

In any case, other than additional moves toward consolidation, there does not seem to exist signs as yet of further changes in the structure of the industry in Brazil. Current indicators of fragility remain within safe intervals, although the current crisis should feed some caution in dealing with such data. Non-performing loans have actually decreased in 2008, to 2.9 percent of total credit in October, down from 3.3 percent in January. Provisions, consequently, have also remained stable, at levels above 5 percent, which, in principle, are more than sufficient to absorb expected losses. Basel coefficients also remain significantly above required minimums, although the same caveat applies.

Mexico

Over 2008 (from January to October) domestic financing by commercial banks to the non-bank sector increased by 6.47 percent to $160 billion. On average, domestic bank loans accounted for approximately 97 percent of domestic financing (with the

remainder sourced from commercial bank agencies and loans related to restructuring programs). In terms of private and public sector lending, on average, the former received over 83 percent and the latter roughly 17 percent of bank loans. Public sector bank lending has fallen to 13.95 percent in October 2008 from 18.4 percent in June; correspondingly, private sector bank lending grew by 10.85 percent during 2008.

The composition of bank loans has altered over 2008 with the share of consumer credit in bank lending decreasing from around 26 percent to just over 18 percent by October. Concomitantly, bank lending to corporations and self-employed businesses grew from 38 to nearly 43 percent of bank financing. Mortgage lending remained relatively stable (at around 15 percent) whilst loans to non-bank financial intermediaries had doubled to over 10 percent by September and October.

The stock of non-performing loans (NPL) has increased by 9.87 percent in 2008, which is roughly comparable with the observed growth in private sector bank lending. On average, the quality of the loan portfolio as measured by the ratio of NPL to total loans was 2.25 percent, with above-average observations since July. The bulk of NPL are concentrated in the consumer credit sector and mostly in the credit card segment. Yet, the stock of NPL in consumer credit fell by 8.49 percent to account for 54.35 percent of total NPL by October. However, a more-worrying trend seems to be emerging in the mortgage lending and business lending segments, as stocks of NPL increased by 30.48 and 56.06 percent, respectively. As at October 2008, NPL in the mortgage loans sector accounted for more than 20 percent of total NPL (from February's 16 percent); the comparative figure for business loans is 24.58 percent (from 17 percent in January).

Arguably, the situation outlined above is understandable given that interest rates have revised upwards over 2008. For instance, the twenty-eight-day interbank equilibrium interest rate (TIIE) was 8.6835 percent on 30 December 2008 compared with 7.925 percent one year earlier, whilst credit card interest rates reached 41.78 percent in November 2008 compared with 31.6 percent a year previously. Although rates on fixed rate peso-denominated mortgages have been gradually easing since the end of 2004 to a recent low of 12.10 percent in April 2008, there is the beginning of an upward trend, with rates reaching 12.64 percent in November 2008. In spite of these tensions, the Mexican commercial banking sector is well capitalized with a capital adequacy ratio of 15.18 percent at the end of October 2008, albeit slightly down on January's 17 percent.

References

Afanasieff, T. S., P. M. V. Lhacer, and M. I. Nakane (2002). 'The Determinants of Bank Interest Spreads in Brazil'. *Money Affairs*, 15: 183–207.

Ahumada, A. and J. Marshall (2001). 'The Banking Industry in Chile: Competition, Consolidation and Systemic Stability'. *Bank for International Settlements Papers*, 4, August.

Asociación de Bancos Privados de Capital Argentino (ADEBA) (2008). *Report*, October.

Avalos, M. and F. Trillo (2006). 'Competencia Bancaria en México', UN CEPAL, August 2006.

Banco Central de la República Argentina (BCRA) (2008). *Informe sobre Bancos*, 6/2, October.

Bank for International Settlements (2007). Annex tables. BIS Papers 33, February.

—— (2008a). 'New Financing Trends in Latin America: A Bumpy Road towards Stability'. BIS Papers 36, February.

—— (2008b). 'Monetary and Financial Stability Implications of Capital Flows in Latin America and the Caribbean'. BIS Papers 43, November.

—— (2008c). 'BIS Quarterly Review: International Banking and Financial Market Developments', December.

Barajas, A., R. Steiner, and N. Salazar (1998). 'Interest Spreads in Banking: Costs, Financial Taxation, Market Power, and Loan Quality in the Colombian Case 1974–96'. IMF Working Paper, 98/110.

—— —— —— (2000). 'The Impact of Liberalization and Foreign Investment in Colombia's Financial Sector'. *Journal of Development Economics*, 63: 157–96.

Belaisch, A. (2003). 'Do Brazilian Banks Compete?' IMF Working Paper 03/113.

Berger, A. N. (2007). 'International Comparisons of Bank Efficiency'. New York University Salomon Center, *Financial Markets, Institutions & Instruments*, 16: 119–44.

—— G. R. G. Clarke, R. Cull, L. Klapper, and G. F. Udell (2005). 'Corporate Governance and Bank Performance: A Joint Analysis of the Static, Selection, and Dynamic Effects of Domestic, Foreign, and State Ownership'. *Journal of Banking & Finance*, 29: 2179–221.

Betancour, C., J. De Gregorio and A. Jara (2006). 'Improving the Banking System: The Chilean Experience'. *Bank for International Settlements Papers* 28, August.

Brock, P. and L. Rojas Suarez (2000). 'Understanding the Behavior of Bank Spreads in Latin America'. *Journal of Development Economics*, 63: 113–34.

Buch, C. M. and G. L. DeLong (2001). 'Cross-Border Bank Mergers: What Lures the Rare Animal?' Kiel Institute of World Economics Working Paper 1070.

Cardias Williams, F. and J. Williams (2008). 'Does Ownership Explain Bank M&A? The Case of Domestic Banks and Foreign Banks in Brazil', in P. Arestis and L. F. de Paula (eds.), *Financial Liberalization and Economic Performance in Emerging Countries*, Basingstoke: Palgrave Macmillan, 194–215.

Carvalho, F. (1998). 'The *Real* Stabilization Plan and the Banking Sector in Brazil'. *Banca Nazionale del Lavoro Quarterly Review*, 206: 291–36.

—— (2000). 'New Competitive Strategies of Foreign Banks in Large Emerging Economies: The Case of Brazil'. *Banca Nazionale del Lavoro Quarterly Review*, 213: 135–70.

—— (2002). 'The Recent Expansion of Foreign Banks in Brazil: First Results'. *Latin American Business Review*, 3: 93–119.

—— (2008). 'Financial Liberalization in Brazil and Argentina', in P. Arestis and L. F. Paula (eds.), *Financial Liberalization and Economic Performance in Emerging Countries*, Basingstoke: Palgrave Macmillan, 121–41.

—— and A. Kasman (2005). 'Cost Efficiency in the Latin American and Caribbean Banking Systems'. *Journal of International Financial Markets, Institutions and Money*, 15: 55–72.

Claessens, S., A. Demirgüç-Kunt, and H. Huizinga (2001). 'How Does Foreign Bank Entry Affect Domestic Banking Markets?' *Journal of Banking & Finance*, 25: 891–911.

Clarke, G. R. G. and R. Cull (2000). 'Why Privatize? The Case of Argentina's Public Provincial Banks'. *World Development*, 27: 865–86.

—— J. M. Crivelli, and R. Cull (2005). 'The Direct and Indirect Impact of Bank Privatization and Foreign Entry on Access to Credit in Argentina's Provinces'. *Journal of Banking & Finance*, 29: 5–29.

Coelho, C. A., J. M. P. de Mello, and L. Rezende (2007). 'Are Public Banks Pro-Competitive? Evidence from Concentrated Local Markets in Brazil'. Texto para Discssão 551: Pontifícia Universidade Católica do Rio de Janeiro.

Committee on the Global Financial System (CGFS) (2004). 'Foreign Direct Investment in the Financial Sector of Emerging Market Economies'. Report submitted by a working group established by the CGFS. Basel: BIS, March.

Crystal, J. S., B. G. Dages, and L. Goldberg (2002). 'Has Foreign Bank Entry Led to Sounder Banks in Latin America?' *Current Issues in Economics and Finance*, 8: 1–6.

Dages, B. G., L. Goldberg, and D. Kinney (2000). 'Foreign and Domestic Bank Participation in Emerging Markets: Lessons from Mexico and Argentina'. *Federal Reserve Bank of New York Economic Policy Review*, 17–36.

De la Torre, A. (2000). 'Resolving Bank Failures in Argentina: Recent Developments and Issues'. World Bank Policy Research Working Paper 2295.

De Vries, M. G. (1998). *Balance of Payments Adjustment, 1945 to 1986: The IMF Experience*. IMF: Washington.

Domanski, D. (2005). 'Foreign Banks in Emerging Market Economies: Changing Players, Changing Issues'. *Bank for International Settlements Quarterly Review*, 69–81, December.

Economic Commission for Latin American and Caribbean (ECLAC) (2000). *Foreign Investment in Latin America and the Caribbean: 1999 Report*. Santiago: ECLAC.

—— (2005). *Foreign Investment in Latin America and Caribbean: 2004 Report*. Santiago: ECLAC.

Fanelli, J. M. (2003). *Estrategias para la Reconstrucción Monetaria y Financiera de la Argentina*. Buenos Aires: Siglo XXI.

Focarelli, D. and A. F. Pozzolo (2001). 'The Patterns of Cross-Border Bank Mergers and Shareholdings in OECD Countries'. *Journal of Banking & Finance*, 25: 2305–37.

Foxley, A. (1983). *Latin American Experiments with Neo-Conservative Economics*. Berkeley: University of California Press.

Fry, M. (1995). *Money, Interest, and Banking in Economic Development*. 2nd edn. Baltimore: The Johns Hopkins University Press.

Gelos, R. G. and J. Roldós (2004). 'Consolidation and Market Structure in Emerging Market Banking Systems'. *Emerging Markets Review*, 5: 39–59.

—— (2006). 'Banking Spreads in Latin America'. IMF Working Paper 06/44.

Gerschenkron, A. (1962). *Economic Backwardness in Historical Perspective*. Cambridge, MA: The Belknap Press of Harvard University Press.

Guimarães, P. (2002). 'How Does Foreign Entry Affect Domestic Banking Market? The Brazilian Case'. *Latin American Business Review*, 3: 121–40.

Haber, S. (2005). 'Mexico's Experiments with Bank Privatization and Liberalization, 1991–2003'. *Journal of Banking & Finance*, 29: 2325–53.

—— and A. Musacchio (2005). 'Foreign Banks and the Mexican Economy, 1997–2004'. Stanford Center for International Development Working Paper.

Hawkins, J. and D. Mihaljek (2001). 'The Banking Industry in the Emerging Markets Economies: Competition, Consolidation and Systemic Stability'. Bank for International Settlements Papers 4, August.

Hernandez-Murillo, R. (2007). 'Experiments in Financial Liberalization: The Mexican Banking Sector'. *Federal Reserve Bank of St Louis Review*, 89: 415–32.

Honohan, P. (2007). 'Cross-Country Variation in Household Access to Financial Services'. Paper prepared for World Bank conference on Access to Finance, March 15–16.

Hoshino, T. (1996). 'Privatization of Mexico's Public Enterprises and the Restructuring of the Private Sector'. *Developing Economies*, 34: 34–60.

Inter-American Development Bank (IADB) (2005). *Unlocking Credit: The Quest for Deep and Stable Bank Lending*. Washington, DC: IADB.

International Monetary Fund (IMF) (2001). *International Capital Markets: Developments, Prospects, and Key Policy Issues*. Washington, DC: IMF.

—— *Global Financial Stability Report*. Washington, DC: IMF, September.

Jeanneau, S. (2007). 'Banking Systems: Characteristics and Structural Changes'. Bank for International Settlements Papers 33: 3–16, February.

Martinez-Peria, M. S. and A. Mody (2004). 'How Foreign Participation and Market Concentration Impact Bank Spreads: Evidence from Latin America'. *Journal of Money, Credit and Banking*, 36: 511–37.

Montes-Negret, F. and L. Landa (2001). 'Interest Rate Spreads in Mexico during Liberalization', in G. Caprio, P. Honohan, and J. E. Stiglitz (eds.), *Financial liberalization: How far? How fast?*, Cambridge: Cambridge University Press.

Nakane, M. I. (2001). 'A Test of Competition in Brazilian Banking'. Banco Central do Brasil Working Paper 12.

—— L. S. Alencar, and F. Kanczuk (2006). 'Demand for Bank Services and Market Power in Brazilian Banking'. Banco Central do Brasil Working Paper 107, June.

—— and D. B. Weintraub (2005). 'Bank Privatization and Productivity: Evidence for Brazil'. *Journal of Banking & Finance*, 29: 2259–89.

O'Connell, A. (2005). 'The Recent Crisis — and Recovery — of the Argentine Economy: Some Elements and Background, in G. Epstein (ed.), *Financialization and the World Economy*, Cheltenham: Edward Elgar.

Panzar, J. C. and J. N. Rosse (1987). 'Testing for Monopoly Equilibrium'. *Journal of Industrial Economics*, 35: 443–56.

Paula, L. F. (2002). 'Expansion Strategies of European Banks to Brazil and Their Impacts on the Brazilian Banking Sector'. *Latin American Business Review*, 3: 59–91.

—— and A. J. Alves Jr. (2003). 'Banking Behavior and the Brazilian Economy after the Real Plan: A Post-Keynesian Approach'. *Banca Nazionale del Lavoro Quarterly Review* 227: 337–65.

—— —— (2007). 'The Determinants and Effects of Foreign Bank Entry in Argentina and Brazil: A comparative Analysis'. *Investigación Económica*, 66: 259, 63–102.

Peek, J. and E. S. Rosengren (2000). 'Implications of the Globalization of the Banking Sector: The Latin American Experience'. *New England Economic Review*, September/October, 45–62.

Rojas Suarez, L., 2001. 'Rating Banks in Emerging Markets: What Credit Rating Agencies Should Learn from Financial Indicators'. Institute for International Economics Working Paper 01-06.

Rojas Suarez, L., (2007). 'The Provision of Banking Services in Latin America: Obstacles and Recommendations'. Center for Global Development Working Paper 124, June.

Schulz, H. (2006). 'Foreign Banks in Mexico: New Conquistadors or Agents of Change?' University of Pennsylvania disscussion paper.

Sebastián, M. and C. Hernansanz (2000). 'The Spanish Banks Strategy in Latin America'. Société Universitaire Européenne de Recherches Financières, Working Paper 9, Paris.

Sidaoui, J. (2006). 'The Mexican Financial System: Reforms and Evolution 1995–2005'. Bank for International Settlements Papers 28, August.

Singh, A., A. Belaisch, C. Collyns, P. de Masi, R. Krieger, G. Meredith, and R. Rennhack (2005). 'Stabilization and Reform in Latin America: A Macroeconomic Perspective on the Experience Since the Early 1990s'. IMF Occasional Paper 238.

Stallings, B. and R. Studart (2006). *Finance for Development: Latin America in Comparative Perspective*. Washington, DC: The Brookings Institution.

Studart, R. and J. Hermann. (n.d.) 'Sistemas Financeiros Argentino e Brasileiro', manuscript.

Tabak, B. M., S. M. Guerra, E. J. A. Lima, and E. J. Chang (2007). 'The Stability-Concentration Relationship in the Brazilian Banking System'. Banco Central do Brasil Working Paper Series 145.

Vasconcelos, M. R. and J. R. Fucidji (2002). 'Foreign Entry and Efficiency: Evidence from the Brazilian Banking Industry'. State University of Maringá, Brazil.

Yeyati, E. L. and A. Micco (2007). 'Concentration and Foreign Penetration in Latin American Banking Sectors: Impact on Competition and Risk'. *Journal of Banking & Finance*, 31: 1633–47.

Yildirim, H. S. and G. C. Philippatos (2007). 'Restructuring, Consolidation and Competition in Latin American Banking Markets'. *Journal of Banking & Finance*, 31: 629–39.

CHAPTER 35

BANKING IN JAPAN

HIROFUMI UCHIDA
GREGORY UDELL[1]

Introduction

This chapter focuses on the Japanese banking industry. We examine its structure, its performance, and some of its defining characteristics. There are a number of reasons why an analysis of the banking industry in Japan may be particularly interesting. First, it is an essential part of one of world's largest economies—an economy that is now second only to the US in terms of the size of its GDP. Second, like some other developed economies such as Germany it has historically been a banking-oriented financial system. Third, the banking industry has some very interesting idiosyncratic features related to the nature of the Japanese corporate environment such as its 'main banking system'. Fourth, like other countries, the Japanese banking system has been in a period of significant transition, some of which is idiosyncratic to Japan such as the banking crisis of the 1990s.

In the next section of the chapter we will provide an overview of the Japanese banking system. Then in the third section we will address specific topics related to the Japanese banking system that are particularly interesting: the main bank system (including relationship banking), the Japanese banking crisis during the 1990s, and

[1] The authors would like to thank the editors, Kozo Harimaya, Hikaru Fukanuma, Hiroshi Fujiki, Takeo Hoshi, Tae Okada, Arito Ono, Kenji Fujii, Wako Watanabe, and Yoshiaki Ogura for their helpful comments.

the impact of the subprime crisis in 2008 on Japanese banks. The fourth section concludes.[2]

Overview of the Japanese banking system

In this section we provide an overview of the banking industry in Japan including discussions of its market structure, efficiency, permissible activities, and regulation.

The importance of banking and intermediated finance in Japan

The importance of banking and intermediated finance in Japan can be seen in Table 35.1. This breakdown of Japanese financial assets shows that of the 6,400 trillion yen of financial assets in Japan, 46 percent are held by Japanese financial institutions. Depository financial institutions are by far the largest component, holding 23 percent of the country's financial assets. Among them, banks, which are our main focus in this chapter, comprise the largest segment of depository institutions, holding 19 percent of the country's financial assets. They can be broken down into domestically licensed banks (63 percent of banks assets), foreign banks (4 percent), 'financial' (banking) institutions for agriculture, forestry, and fisheries (17 percent), and 'financial' (banking) institutions for small businesses (16 percent).

We can also see the importance of the banking system in terms of the dependency of the household sector and the corporate sector, which are respectively the largest creditor and debtor sectors in Japan. Table 35.2 presents the asset and liability composition by financial instruments for these sectors. Of the 15.4 trillion yen of financial assets held by households, 47.1 percent is allocated to demand (transferable) and time deposits. In the corporate sector, loans from private financial institutions comprise 36.5 percent of firms' total debt financing (total assets minus shares and equity). This dwarfs other forms of debt, including, in particular, commercial paper (0.9 percent) and corporate bonds (i.e., industrial securities) (6.9 percent). By way of comparison, in the US, bank loans consist of 8.8 percent of corporate liabilities with commercial paper providing 0.8 percent and

[2] For a more comprehensive analysis of corporate finance and the Japanese banking industry we suggest that the reader refer to Hoshi and Patrick (2000); and Hoshi and Kashyap (2001). For an analysis of the Japanese economy we refer the reader to Flath (2000). And, for a comprehensive evaluation of economic policies related to the recent recession we refer the reader to Ito, Patrick, and Weinstein (2005).

Table 35.1. Financial assets in Japan (by holder)

Financial institutions	29,636,683 (46%)			
Central Bank		1,204,335 (2%)		
Depository corporations		14,940,029 (23%)		
Banks			12,516,669 (19%)	
Domestically licensed banks				7,939,431 (63% among 'Banks')
Foreign banks in Japan				526,360 (4% among 'Banks')
Financial institutions for agriculture, forestry, and fisheries				2,083,195 (17% among 'Banks')
Financial institutions for small businesses				1,967,683 (16% among 'Banks')
Postal savings			2,301,880 (4%)	
Collectively managed trusts			121,480 (0%)	
Insurance and pension funds		5,126,374 (8%)		
Other financial intermediaries		8,233,741 (13%)		
Securities investment trusts			1,077,554 (2%)	
Non-banks			1,235,946 (2%)	
Finance companies				789,599
Structured-financing special purpose companies and trusts				446,347
Public financial institutions			4,449,924 (7%)	
Fiscal Loan Fund				2,904,723
Government financial institutions				1,545,201
Financial dealers and brokers			1,504,312 (2%)	
Financial auxiliaries (financial institutions other than intermediaries)		132,204 (0%)		
Non-financial corporations	10,147,392 (16%)			
General government	5,224,741 (8%)			
Households	15,361,628 (24%)			
Private non-profit institutions serving households	499,434 (1%)			
Overseas	3,495,646 (5%)			
Total	64,365,524 (100%)			

Source: Flow of Funds Account (the Bank of Japan) (100 million yen, March 31, 2007).

Table 35.2. Assets and liabilities of the Japanese corporate and household sectors

	Private non-financial corporations				Households			
	Assets		Liabilities		Assets		Liabilities	
Currency and deposits	1,967,222	(19.7%)			7,698,512	(58.1%)		
Currency	272,962	(2.7%)			426,629	(2.8%)		
Transferable deposits	1,154,181	(11.6%)			2,255,977	(14.7%)		
Time and savings deposits	397,149	(4.0%)			4,974,303	(32.4%)		
Certificates of deposits	132,250	(1.3%)			575	(0.0%)		
Foreign currency deposits	10,680	(0.1%)			41,028	(0.3%)		
Loans	392,261	(3.9%)	3,416,028	(23.2%)	287	(0.0%)	3,268,959	(82.6%)
Call loans and money	27,658	(0.3%)						
Loans by private financial institutions			2,596,277	(17.6%)			2,634,274	(66.5%)
Housing loans							*1,448,581*	*(36.6%)*
Consumer credit							*382,445*	*(9.7%)*
Loans to companies and governments			*2,596,277*	*(17.6%)*			*803,248*	*(20.3%)*
Loans by public financial institutions			307,751	(2.1%)			557,840	(14.1%)
Of which: housing loans							*423,209*	*(10.7%)*
Loans by the non-financial sector	321,061	(3.2%)	352,737	(2.4%)	287	(0.0%)	65,335	(1.7%)
Installment credit (not included in consumer credit)			159,142	(1.1%)			11,510	(0.3%)
Repurchase agreements and securities lending transactions	43,542	(0.4%)	121	(0.0%)				
Securities other than shares	394,999	(4.0%)	741,439	(5.0%)	1,117,193	(7.3%)		
Central government securities and FILP bonds	21,730	(0.2%)			333,795	(2.2%)		
Local government securities	17,669	(0.2%)			12,291	(0.1%)		
Public corporation securities	44,517	(0.4%)	23,108	(0.2%)	6,118	(0.0%)		

Bank debentures	21,996	(0.2%)			27,531	(0.2%)		
Industrial securities	16,777	(0.2%)	488,043	(3.3%)	645	(0.0%)		
External securities issued by residents			134,243	(0.9%)				
Commercial paper	38,657	(0.4%)	63,258	(0.4%)				
Investment trust beneficiary certificates	98,094	(1.0%)	32,787	(0.2%)	684,285	(4.5%)		
Trust beneficiary rights	19,238	(0.2%)			51,728	(0.3%)		
Structured-financing instruments	115,788	(1.2%)						
Mortgage securities	533	(0.0%)			800	(0.0%)		
Shares and other equities	3,141,069	(31.4%)	7,619,671	(51.7%)	1,874,530	(12.2%)		
Of which: shares	1,259,532	(12.6%)	4,915,909	(33.4%)	1,112,185	(7.2%)		
Financial derivatives	14,267	(0.1%)	20,749	(0.1%)	1,600	(0.0%)	1,229	(0.0%)
Insurance and pension reserves					4,018,540	(26.2%)		
Insurance reserves					2,289,842	(14.9%)		
Pension reserves					1,728,698	(11.3%)		
Deposits money	274,945	(2.8%)	340,890	(2.3%)	83,086	(0.5%)		
Trade credits and foreign trade credits	2,578,354	(25.8%)	2,863,529	(14.0%)			568,232	(14.4%)
Accounts receivable/payable	70,869	(0.7%)	111,662	(0.8%)	334,831	(2.2%)	66,009	(1.7%)
Outward direct investment	327,578	(3.3%)						
Outward investment in securities	526,270	(5.3%)			98,410	(0.6%)		
Other external claims and debts	145,628	(1.5%)	37,698	(0.3%)				
Others	128,281	(1.3%)	377,423	(2.6%)	142,639	(0.9%)	55,232	(1.4%)
Total	9,988,601	(100.0%)	14,729,089	(100.0%)	15,361,628	(100.0%)	3,959,661	(100.0%)
(Difference between financial assets and liabilities)			−4,738,146	(debtor)			11,401,967	(creditor)

Source: Flow of Funds Account (the Bank of Japan) (100 million yen, March 31, 2007).

corporate bonds providing 21.5 percent (2006 Flow of Funds Accounts of the US). Overall, these data indicate that Japan is very much a bank-intermediated financial system compared to market-oriented systems such as the US

It has been suggested that high growth in the post-war period may have been facilitated by the banking-oriented nature of Japan's financial system, and some have further argued that this type of financial system with its emphasis on financial intermediation may be a good model for developing economies (Aoki and Patrick, 1994). However, it should also be noted that Japan suffered a banking crisis that lasted virtually the entire decade of the 1990s, the cause of which has been linked to the banking system. Today, the dependence on banking appears to be diminishing. Specifically, the fraction of financial assets held in banks has been on a general downward trend for more than a decade (see Figure 35.1).

Segmentation in the Japanese banking market

The Japanese banking sector can best be described as segmented. In particular, since World War II, the Japanese financial industry had been segmented by the nature of the financial services that each type of financial institution could provide. The origin of this regulatory segmentation dates back to the crisis-mode wartime system. Its purpose was to limit competition in order to promote banking

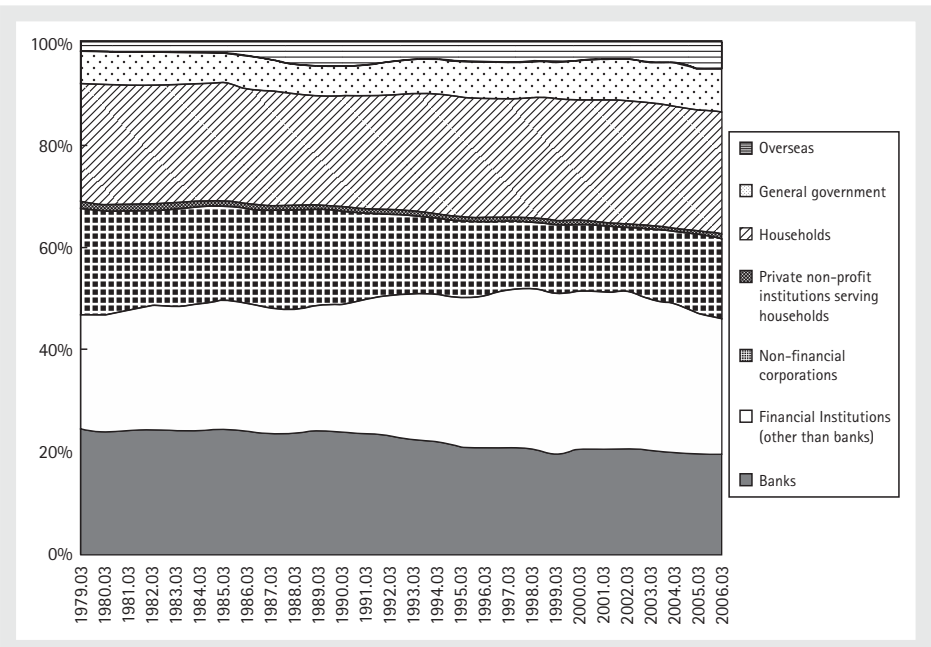

Fig. 35.1. Composition of financial assets in Japan (by holder)

Source: Flow of Funds Account (the Bank of Japan).

profitability, thereby enhancing financial system safety and soundness. Although financial liberalization in the 1980s and 1990s has blurred the wall between these institutions, there still remain some boundaries (Hoshi and Kashyap, 2001: chap. 4).

In Japan today commercial banks and commercial banking are defined under the 1981 Banking Law. The Law defines banking as either the simultaneous provision of lending and deposit taking, or just the provision of payments and settlements services. Among financial institutions regulated under the Law, ordinary banks (futsuu ginko) are the most common, although there are other financial institutions and public banks that engage in the business of banking, as we will see below.

The Banking Law also permits banks to engage in other activities, including investing in bonds and stocks—that is, in Japan, banks are allowed to own equity in non-financial corporations.[3] As we will see below, bank equity ownership gives banks an important role in corporate governance not only as a creditor but also as a stockholder. In the case of some activities such as factoring and leasing, banks must engage in them indirectly through affiliates. With this background in mind, we now turn to a description of the various types of banks in Japan.

City banks

City banks are a type of ordinary bank. Although there are now only five city banks (Mitsui-Sumitomo, Mitsubishi-Tokyo-UFJ, Mizuho, Mizuho-Corporate, and Resona), they are the largest single category (Table 35.3). City banks grew quite rapidly in absolute and relative importance during the 1980s (Figures 35.2 and 35.3). All five city banks are universal banks, all offer nationwide branch banking, and three city banks have extensive foreign bank networks (Mitsui-Sumitomo, Mitsubishi-Tokyo-UFJ, and Mizuho-Corporate).

Regional banks and second regional banks

Regional banks are medium-sized banks whose banking operations are regionally focused. There are sixty-four of these banks in Japan, and while they are individually substantially smaller than the city banks, they collectively comprise the second-largest category (Table 35.3). Like the regional banks, the second (-tier) regional banks also operate regionally, but they tend to be smaller in size. Historically, these banks were established as mutual (Sogo) banks whose purpose was to provide financing to small and medium-size enterprises (SMEs). While they are no longer restricted to this sector of the economy, they still tend to focus on SMEs. As a group, the second regional banks are considerably smaller than city banks and regional banks but they still play an important role in providing SME financing. Economists have often grouped regional and second regional banks together because both categories tend to focus on local and retail banking.

[3] There are now two key restrictions on this ownership. Banks cannot hold more than 5% of one company's equity to prevent predominating influence and they cannot own equity in aggregate that exceeds the bank's own capital for safety and soundness reasons.

Table 35.3. Descriptive statistics for different bank types in Japan

	# of banks	Assets	Loans	Deposits	Loans/ assets	Loans/ deposits	Supervisory authority
Private banks							
City banks	5	3,855,503	1,860,370	2,507,624	0.48	0.74	FSA (Financial Services Agency)
Regional banks	64	2,228,711	1,445,409	1,888,910	0.65	0.77	FSA
Second regional banks	47	602,318	419,377	541,266	0.70	0.77	FSA
Trust banks	21	587,756*	317,319*	332,295*	0.54*	0.95*	FSA
Foreign banks	69	577,750	86,980	174,284	0.15	0.50	FSA
Shinkin banks*	292	1,180,074	626,706	1,092,212	0.53	0.57	FSA
Credit cooperatives*	172	168,095	93,078	159,430	0.55	0.58	FSA
Labor banks*	13	150,554	97,095	141,804	0.64	0.68	Ministry of Health, Labor and Welfare and FSA
Agricultural cooperatives*	865	805,558	213,185	788,653	0.26	0.27	Ministry of Agriculture, Forestry and Fishery
Other banks under the Banking Law*	11	251,682#	131,503#	168,395#	0.52#	0.78#	FSA
Government financial institutions and postal savings							
Government Housing Loan Corporation	1	478,097	436,327	0	0.91	NA	Ministry of Land, Infrastructure and Transport
National Life Finance Corporation	1	84,191	83,435	0	0.99	NA	Ministry of Health, Labor and Welfare and Ministry of Finance
Shoko Chukin Bank	1	109,968	93,553	25,399	0.85	3.68	Ministry of Economy, Trade, and Industry

Japan Finance Corporation for Small and Medium Enterprises	1	74,984	68,965	0	0.92	NA	Ministry of Economy, Trade, and Industry
Development Bank of Japan	1	133,185	121,974	0	0.92	NA	Ministry of Finance
Japan Bank for International Cooperation*	1	209,341	192,023	0	0.92	NA	Ministry of Finance
Postal savings (Japan Post)	1	2,316,282	43,761	1,858,226	0.02	0.02	Ministry of Internal Affairs and Communications

Source and dates:
[# of banks]: From FSA (for its licensed banks) and Nikkin homepage (for others), as of March 31, 2006.
[Balance sheet figures]: From the Bank of Japan, from the Japanese Bankers Association (for trust banks), and from respective banks (for Other banks under the Banking Law and the government financial institutions), as of March 31, 2007 (For banks with*, as of March 31, 2006).
[Unit (for assets, loans, and deposits)]: 100 million yen.

Notes: Deposits do not include CDs (certificates of deposits) and financial bonds issued by some banks (similar to time deposits).
* These figures for trust banks are for 7 banks that are full members of the Japanese Bankers Association.
# These figures for Other banks under the Banking Law are for 8 banks (Aozora Bank, Shinsei Bank, Seven Bank, e-Bank corporation, Japan Net Bank, Sony Bank, ShinGinko Tokyo and Saitama Resona Bank). The Resolution and Collection Bank and the Second Bridge Bank of Japan are excluded because they are not commercial banks in the usual sense. The Incubator Bank of Japan is excluded because of data unavailability.

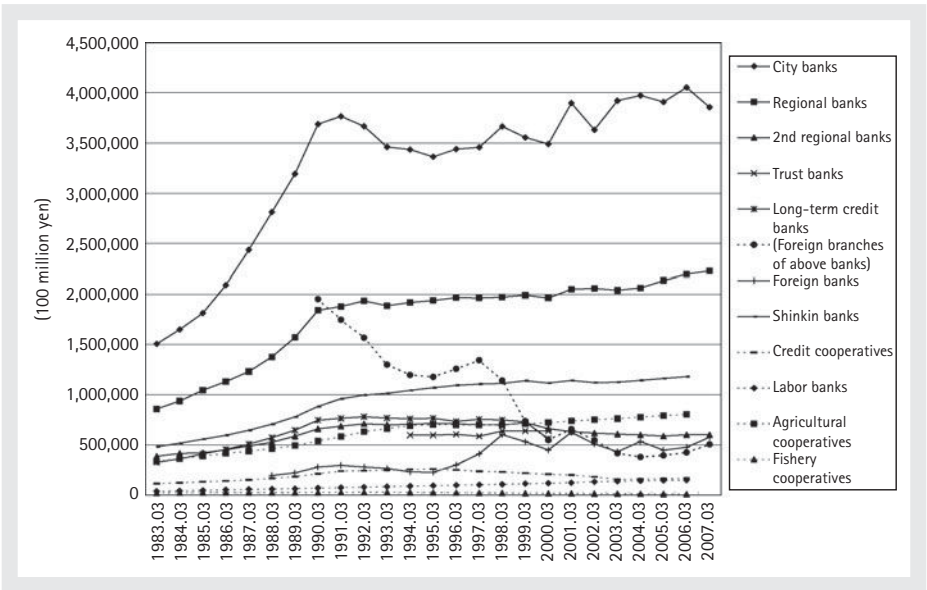

Fig. 35.2. Private bank assets

Source: Financial and Economic Statistics Monthly (the Bank Japan).

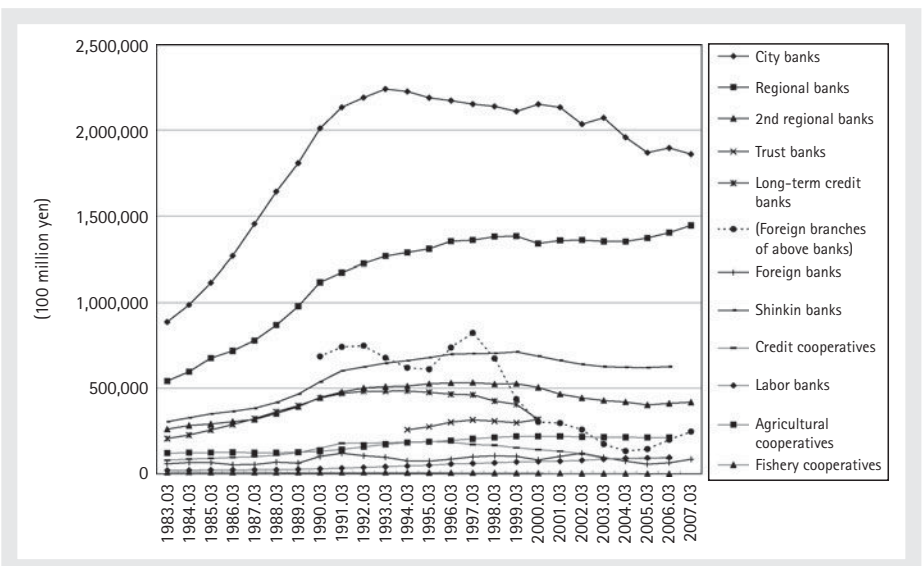

Fig. 35.3. Loans outstanding of private banks

Source: Financial and Economic Statistics Monthly (the Bank Japan).

Trust banks

Under the 1943 Act on Provision of Trust Business by Financial Institutions, 'trust banks' are allowed to offer trust services. They are, nevertheless, commercial banks under the Banking Law with respect to their provision of normal banking services (deposits, loans, and payments and settlements). These banks offer 'money trusts' (kinsen shintaku) to their customers which are essentially a form of medium- to long-term time deposit. These money trusts enable the banks to make long-term commercial loans and investments in bonds and equities. That is, from an asset liability perspective these banks specialize in long-term lending funded by long-term liabilities. As a result, the trust banks have played an important role in providing long-term funding to Japanese corporate borrowers. This role is especially important given that the domestic post-war Japanese corporate bond market had been underdeveloped.

Long-term credit banks

Historically, long-term credit banks also played an important role in providing long-term corporate funding in the post-war Japanese financial system. They operated under the 1952 Long-Term Credit Banking Law until the banking crisis of the 1990s. They no longer exist today in their original form. They were initially designed to complement ordinary banks, which were (supposed to be) restricted to short-term lending (based on the asset-liability-management principle that commercial banks whose funding comes from short-term demand deposits should be prohibited from investing in long-term loans). These banks could issue bonds that were historically more attractive to investors than time deposits because of deposit-rate ceilings and other attributes.

There were only three long-term credit banks. During the banking crisis, the Industrial Bank of Japan was merged with two city banks and consolidated to become the Mizuho Bank and the Mizuho Corporate Bank. The other two, the Long-Term Credit Bank of Japan and the Nippon Credit Bank, both went into bankruptcy in 1998, marking a key event in Japan's banking crisis. After being temporarily nationalized, these two banks now operate under the Banking Law but under different names: the Shinsei Bank and the Aozora Bank, respectively. The disappearance of the long-term credit banks and the long-term debt financing that they provided to large Japanese corporations is partly a consequence of the emergence of a domestic corporate bond market in Japan and increased access by Japanese corporations to the Eurobond and other international bond markets.

Shinkin banks and credit cooperatives

Shinkin banks (Shin-you Kinko) and credit cooperatives (Shin-you Kumiai) are both cooperative banks that specialize in providing commercial banking services to member SMEs and individuals. Although they are not legally 'banks', because they

operate under a special set of laws, they engage in the same activities as banks do under the Banking Law—that is, lending, deposit taking, and payments and settlements. However, with respect to lending, they are restricted to lending to their member firms. Deposits can be offered to non-members by Shinkin banks, but only to members by credit cooperatives. Firms that are members of a Shinkin bank must have fewer than 300 employees or a capitalization less than 900 million yen, while members of credit cooperatives must have fewer than 300 employees or a capitalization less than 300 million yen. These banks are also restricted geographically, typically to an area no larger than a single prefecture. Since financial deregulation in the 1980s, both types of banks have been permitted to expand their business scope to activities that include, for example, offering loans to non-members and offering mutual funds.

Foreign banks

There are a large number of foreign banks that have branch offices or agencies in Japan. These branches require a banking license, and are regulated in the same manner as domestic banks under the Banking Law. Overall, focusing primarily on providing foreign exchange-related services, they play only a minor role in Japanese financial intermediation as indicated in Table 35.3, and Figures 35.2 and 35.3.

Public banks

As in some other countries postal savings have played a historically important role in Japan. Postal savings had long been provided by the government (the Ministry of Posts and Telecommunications). Funds collected through postal savings had flowed to the Ministry of Finance that, in turn, allocated the funds to government financial institutions and other official accounts through the Fiscal Investment and Loan Program (FILP) (see, e.g., Cargill and Yoshino, 2000).

The Program is now in the process of being restructured. In January 2001, the Postal Services Agency assumed the operation of the three postal businesses (postal savings as well as postal insurance and postal mail services), and in April 2001 stopped sending its funds to the Ministry of Finance and started to allocate its funds at its own discretion. In 2003, these three businesses were further transferred to Japan Post, a state-owned company. Finally, in October 2007, the operation of postal saving business was succeeded by a newly established private bank, the Japan Post Bank. All of the equity of the Japan Post Bank is still owned by a government holding company. The equity is supposed to be sold in the market in a stepwise manner, and the completion of the privatization process is scheduled for 2017.

Although in the past they had favorable tax and/or institutional advantages, postal savings today are almost the same as deposits provided by other private banks, and Japan Post Bank is effectively the single largest depository institution in Japan (see Table 35.3). Owing to historical inertia, most of the assets of the Japan

Post Bank are now invested in very low-risk instruments, making it effectively a narrow (100 percent reserve) bank, although there is some interest in expanding its asset composition into such categories as housing loans.

There are also a number of other government financial institutions, some of which are not, technically, banks (Table 35.3). They had been users of the FILP funds raised through postal savings (and related sources). Now, however, they raise funds themselves by issuing special government-guaranteed bonds. They are also in the process of being privatized and consolidated. These institutions include several that have focused on SME lending: the National Life Finance Corporation, the Shoko Chukin Bank, and the Japan Finance Corporation for Small and Medium Enterprises (formerly the Japan Finance Corporation for Small Business) (see Fukanuma, Nemoto, and Watanabe, 2006). Another institution, the Government Housing Loan Corporation, had provided housing loans, although it has already stopped investing in these loans and now focuses on securitizing housing loans by private banks. The Development Bank of Japan is a public bank that has been providing long-term funds to corporations. It played an important role in the post-war development of Japan. The Japan Bank for International Cooperation provides support for the Japanese government's foreign economic policy initiatives and economic cooperation programs.[4]

Other financial institutions providing commercial banking services

As shown in Table 35.3, there are also a number of other financial institutions that provide commercial banking services. Other banks under the Bank Law include two former long-term credit banks (explained above), Internet banks, a resolution bank to manage, collect, and dispose of assets of failed financial institutions, bridge banks provisionally to assume the assets of failed financial institutions until they are transferred to assuming financial institutions, and recently established banks which do not fit into the classification above.

There are also cooperative banks that operate for the benefit of their members: labor banks, agricultural cooperatives, fishery cooperatives, and forestry cooperatives. Among these, the agricultural cooperatives (commonly called collectively as JA Bank (Japan Agriculture Bank) are relatively large (see Table 35.3). Similar to Shinkin banks and credit cooperatives, some of the restrictions on these cooperatives have been lifted to allow, for example, the provision of services to non-members and sales of mutual funds. Thus, to a certain degree, these banks are becoming similar to banks that operate under the Banking Law.

[4] It should also be noted that government credit guarantees are also an important instrument of public policy. During the fiscal year 2006 (April 2006–March 2007), the Credit Guarantee Corporations provided SME loan guarantees totaling 13,659,133 million yen. For more information about the credit guarantee system in Japan, see annual reports of the Japan Finance Corporation for Small and Medium Enterprises <http://www.jasme.go.jp/jpn/summary/disclosure/annualreport.html>.

Market structure and competition of the Japanese banking industry

The segmentation explained above makes it difficult to describe the market structure of the Japanese banking industry. Regarding the deposit market, it does not seem to be segmented by bank type because no product differentiation is likely to exist, even between bank deposits and postal savings. However, the lending market is more complicated. On the one hand, financial deregulation has probably promoted integration (overlap) of the markets of different bank types. On the other hand, as we noted above, different types of banks are likely to have different comparative advantages with respect to different types of borrowers. Moreover, geographical segmentation may still be important, particularly for certain types of lending, such as relationship loans that probably have a spatial dimension. Thus, two types of segmentation are likely to exist: spatial segmentation and bank-type segmentation. Defining the scope of lending markets in Japan is an important empirical question.

Very few empirical studies have investigated market segmentation in Japan. A rare exception is Kano and Tsutsui (2003), who investigate segmentation by prefecture (i.e., spatial segmentation). They find that the lending market for Shinkin banks is segmented by prefecture (probably due to geographical restrictions on their operating areas), whereas prefectural segmentation is only weakly confirmed for regional banks. However, they do not investigate segmentation by bank type. Instead, they implicitly *assume* that lending markets for Shinkin banks and for regional banks are segmented. Thus, bank-type segmentation remains untested. Segmentation by prefecture was also found by Ishikawa and Tsutsui (2006).

Whether or not markets are segmented by type or by region, banks can compete via branching. However, there used to be strict regulation on branching in Japan. To open a new bank branch office, banks had to satisfy regulatory criteria and obtain approval by the Ministry of Finance, the bank regulatory authority at that time. The criteria were relaxed in a stepwise manner from the 1980s through the 1990s, and banks are now virtually unconstrained in opening branch offices following a 2002 amendment to the Banking Law. Also, banks are now able to provide banking services through their agents, such as other banks, insurance companies, securities companies, and non-financial companies.

Table 35.4 shows the number of branch offices of the four major types of banks. As we explained above, city banks have many branches in their nationwide operations (except for Mizuho Corporate, which focuses on big businesses). Typically, there are one or two regional banks and one or two second regional banks in a prefecture (there are forty-seven prefectures in Japan). These banks typically have branch offices in and around their own prefecture and in large cities such as Tokyo and Osaka. Regional banks have a larger number of branch offices than second regional banks.

Table 35.4. Four main types of banks in Japan

	# of branches			# of employees
	Total	Domestic	Foreign	
5 City banks total	2,357	2,251	106	82,792
Mizuho Bank	456	456	—	17,455
Bank of Tokyo-Mitsubishi UFJ	842	780	62	30,876
Sumitomo Mitsui Banking Corporation	623	605	18	20,380
Resona Bank	377	377	—	9,432
Mizuho Corporate Bank	59	33	26	4,649
64 Regional banks total	7,484	7,470	14	124,274
Hokkaido Bank	134	134	—	1,724
Aomori Bank	111	111	—	1,484
Michinoku Bank	116	116	—	1,182
Akita Bank	104	104	—	1,507
Hokuto Bank	85	85	—	1,051
Shonai Bank	71	71	—	751
Yamagata Bank	79	79	—	1,380
Bank of Iwate	110	110	—	1,474
Tohoku Bank	56	56	—	588
77 Bank	141	141	—	2,716
Toho Bank	115	115	—	1,916
Gunma Bank	145	144	1	3,009
Ashikaga Bank	150	150	—	2,278
Joyo Bank	173	173	—	3,486
Kanto Tsukuba Bank	85	85	—	1,098
Musashino Bank	91	91	—	2,002
Chiba Bank	163	160	3	3,733
Chiba Kogyo Bank	71	71	—	1,257
Tokyo Tomin Bank	78	78	—	1,600
Bank of Yokohama	193	193	—	3,418
Daishi Bank	125	125	—	2,365
Hokuetsu Bank	89	89	—	1,525
Yamanashi Chuo Bank	91	91	—	1,688
Hachijuni Bank	155	154	1	2,991
Hokuriku Bank	185	185	—	2,570
Toyama Bank	35	35	—	392
Hokkoku Bank	129	129	—	1,908
Fukui Bank	100	100	—	1,319
Shizuoka Bank	184	181	3	3,419
Suruga Bank	119	119	—	1,530
Shimizu Bank	81	81	—	1,127
Ogaki Kyoritsu Bank	142	142	—	2,442
Juroku Bank	150	150	—	2,668
Mie Bank	74	74	—	1,117
Hyakugo Bank	128	128	—	2,310
Shiga Bank	131	130	1	2,186
Bank of Kyoto	130	130	—	2,685
Kinki Osaka Bank	137	137	—	2,391
Senshu Bank	64	64	—	1,153

(cont.)

Table 35.4. (*Continued*)

	# of branches			# of employees
	Total	Domestic	Foreign	
Bank of Ikeda	72	72	—	1,262
Nanto Bank	126	126	—	2,690
Kiyo Bank	94	94	—	1,868
Tajima Bank	72	72	—	626
Tottori Bank	69	69	—	699
San-in Godo Bank	153	153	—	2,090
Chugoku Bank	167	166	1	3,018
Hiroshima Bank	168	168	—	3,030
Yamaguchi Bank	156	153	3	2,793
Awa Bank	94	94	—	1,469
Hyakujushi Bank	119	119	—	2,170
Iyo Bank	151	150	1	2,574
Shikoku Bank	121	121	—	1,670
Bank of Fukuoka	167	167	—	3,346
Chikuho Bank	42	42	—	589
Bank of Saga	110	110	—	1,550
Eighteenth Bank	101	101	—	1,468
Shinwa Bank	141	141	—	2,039
Higo Bank	124	124	—	2,039
Oita Bank	107	107	—	1,567
Miyazaki Bank	97	97	—	1,358
Kagoshima Bank	139	139	—	2,264
Bank of the Ryukyus	65	65	—	1,203
Bank of Okinawa	62	62	—	1,087
Nishi-Nippon City Bank	247	247	—	4,365
47 Second regional banks total	3,312	3,312	—	48,542
North Pacific Bank	166	166	—	3,052
Sapporo Bank	65	65	—	762
Yamagata Shiawase Bank	63	63	—	759
Shokusan Bank	55	55	—	677
Kita-Nippon Bank	83	83	—	962
Sendai Bank	71	71	—	766
Fukushima Bank	60	60	—	570
Daito Bank	63	63	—	622
Tewa Bank	91	91	—	1,584
Tochigi Bank	94	94	—	1,610
Ibaraki Bank	61	61	—	914
Keiyo Bank	115	115	—	1,838
Higashi-Nippon Bank	76	76	—	1,398
Tokyo Star Bank	34	34	—	957
Kanagawa Bank	32	32	—	454
Taiko Bank	70	70	—	1,037
Nagano Bank	58	58	—	811
First Bank of Toyama	69	69	—	753
Fukuho Bank	51	51	—	580

(cont.)

Table 35.4. (*Continued*)

	# of branches			# of employees
	Total	Domestic	Foreign	
Shizuoka Chuo Bank	41	41	–	466
Gifu Bank	48	48	–	589
Aichi Bank	107	107	–	1,681
Bank of Nagoya	111	111	–	1,929
Chukyo Bank	95	95	–	1,297
Daisan Bank	96	96	–	1,499
Biwako Bank	70	70	–	874
Kansai Urban Banking Corporation	104	104	–	1,729
Taisho Bank	30	30	–	297
Wakayama Bank	31	31	–	463
Minato Bank	108	108	–	1,903
Shimane Bank	34	34	–	429
Tomato Bank	59	59	–	848
Momiji Bank	130	130	–	2,321
Saikyo Bank	64	64	–	712
Tokushima Bank	73	73	–	992
Kagawa Bank	86	86	–	1,222
Ehime Bank	93	93	–	1,455
Bank of Kochi	71	71	–	1,025
Fukuoka Chuo Bank	41	41	–	474
Saga Kyoei Bank	34	34	–	387
Bank of Nagasaki	33	33	–	334
Kumamoto Family Bank	77	77	–	1,121
Howa Bank	49	49	–	655
Miyazaki Taiyo Bank	53	53	–	671
Minami-Nippon Bank	63	63	–	835
Okinawa Kaiho Bank	50	50	–	561
Yachiyo Bank	84	84	–	1,667
7 Trust banks total	294	285	9	20,715
Mitsubishi UFJ Trust and Banking	101	96	5	7,124
Mizuho Trust & Banking	38	38	–	3,216
The Chuo Mitsui Trust and Banking	84	84	–	3,825
Sumitomo Trust & Banking	63	59	4	5,230
Nomura Trust and Banking	1	1	–	136
Mitsui Asset Trust and Banking	3	3	–	708
Resona Trust & Banking	4	4	–	476
Other banks total	170	169	1	6,315
Shinsei Bank	39	38	1	2,095
Aozora Bank	18	18	–	1,481
Saitama Resona Bank	113	113	–	2,739
Total	13,617	13,487	130	282,638
Cf) Japan Post (postal savings)	24,126	24,126	–	55,410

Note: Only the full members of the Japanese Bankers Association are included.
Source: Japanese Bankers Association and Japan Post Bank homepage. (As of March 31, 2006).

How competitive is the Japanese banking market? Again, there is a scarcity of research. Pooling city and regional banks, Molyneux, Thornton, and Lloyd-Williams (1996) report that Japanese banks were uncompetitive in 1986 and 1988. Another study, using a sample of city and regional banks from 1974 to 2000, estimates the degree of competition by bank type using a marginal price (Lerner index) approach (Uchida and Tsutsui, 2005). They find that competition improved throughout the sample period, especially in the 1970s and in the first half of the 1980s, when financial deregulation began. They also find that city banks had been facing more competitive pressure than regional banks.

Efficiency of Japanese banks

Most of the studies on the efficiency of Japanese banks focus on ordinary banks (plus long-term credit banks and trust banks).[5] On balance, they find evidence of economies of scale, at least for the average bank. Most earlier studies found evidence of scale economies until the early 1990s regardless of bank size (Fukuyama, 1993; and McKillop, Glass, and Morikawa, 1996), although one study found evidence of diseconomies (Tadesse, 2006). Fukuyama (1993) finds that regional banks are scale-inefficient, second regional banks are more efficient than regional banks, and city banks are close to efficient, exhibiting constant returns to scale.

However, subsequent studies do not uniformly find evidence of scale economies in later periods. Altunbas, et al. (2000) find during the period 1993–6 scale economies for smallest banks only (1–2 trillion yen of assets, or less), but find scale diseconomies for large banks. They also find that when bank asset quality and liquidity risk are not controlled for, the optimal size is much larger. However, their proxy for asset quality, a non-performing loan ratio, might be problematic during the period 1993–6 owing to imprecise disclosure. Drake and Hall (2003) find similar results in 1996, but the optimal bank size in their study is larger: 6–10 trillion yen in terms of loans outstanding. Also, whether the disappearance of universal scale economies after the 1990s in empirical studies is due to an underlying environmental change, or due to methodological improvements, is an open question.

Drake and Hall (2003) also find that in comparing bank types ordinary banks are scale inefficient, whereas long-term credit banks and trust banks are scale efficient. However, the efficiency results for long-term credit banks and trust banks found in this and other studies may be due to the lack of appropriate controls for the difference between these banks and ordinary banks with respect to their asset/liability structure.

[5] Fukuyama (1996) examines the efficiency of Shinkin banks, and Fukuyama, Guerra, and Weber (1999) investigate the efficiency of credit cooperatives.

Studies based on data envelopment analysis investigate pure technical inefficiency—that is, adequate/excessive use of inputs. Fukuyama (1993) finds pure technical inefficiency for ordinary banks with the magnitude being the greatest for regional banks. Drake and Hall (2003) find that the magnitude of pure technical inefficiency is greater than that of scale inefficiency—that is, banks can reduce cost more by adopting a technology requiring less inputs than by increasing the scale of their operation. Drake and Hall (2003) also show that, in terms of pure technical inefficiency, regional banks, and next second-regional banks, are inefficient, while city banks are almost efficient and trust and long-term credit banks are efficient. They also report that the larger the bank size, the smaller the pure technical inefficiency becomes.

There are only a few studies that investigated scope economies in Japanese banks and their results are not consistent with each other.[6] Tachibanaki, Mitsui, and Kitagawa (1991) find cost complementarity between lending and securities investment in 1987 for city, regional, long-term credit, and trust banks. However, McKillop, Glass, and Morikawa (1996) find for the five city banks from 1978 to 1991 no global economies of scope among lending, liquid asset holdings, and securities investments. Rather, they find cost anti-complementarity between lending and holding liquid assets and between lending and securities investment, and cost complementarity between holding liquid assets and securities investment.

Most recently, Harimaya (2008), using the sample of regional banks from 1994 to 2003, finds that there is cost anti-complementarity between lending and securities investment, and between lending and trust businesses, but cost complementarity between securities investment and trust businesses. He also reports that although scale economies are observed on average, product-specific scale diseconomies are found for banks' trust business, which casts doubt on the prospect of banks increasing their profitability by focusing on fee businesses. On balance, it is still unclear whether there are scope economies in banking.

Interestingly, a recent study revealed that the efficiency results might vary depending on modeling methodologies (Drake, Hall, and Simper, 2009). Using bank data from the 1995–2002 period, they show that the efficiency ranking of different types of Japanese banks varies considerably across modeling approaches, casting some doubt on earlier studies.

Commercial vs. universal banking in Japan

Historically, Japan's regulation of universal banking has mirrored regulation in the US (with the exception of equity ownership). Like the Glass-Steagall Act in the US,

[6] As Berger, Hunter, and Timme (1993) point out, the measurement of the economy of scope is methodologically challenging.

Article 65 of the 1948 Securities and Exchange Law in Japan separated investment banking from commercial banking. As in the US, there was a sequential dismantling of this separation beginning significantly in the 1980s when city, long-term credit, trust, and regional banks were allowed to underwrite and deal in public bonds. In addition to the banking activities discussed above, corporate underwriting was allowed in affiliates beginning in 1993, and in 1998 banks could form financial holding companies. These measures were part of a ('big bang') financial system liberalization similar to the UK and the US (see Horiuchi, 2000; and Royama, 2000). Investment banking and trust activities, however, must still be conducted in affiliate organizations that are separate from the banking entity. As of 1 October 2007, there were thirteen bank financial holding companies in Japan, including those of all five city banks and that of the Japan Post Bank.

There have been some studies that have examined issues related to universal banking in Japan, including the conflict of interest issue and relationship building across commercial and investment banking services. The results are mixed. Hamao and Hoshi (2000) show that the new-issue corporate bond yield spread does not depend on whether the underwriter is a bank subsidiary. However, Takaoka and McKenzie (2004) find that underwriting commissions are smaller when the lead underwriter is a bank-owned securities company. They also find that, after the entry of bank-affiliated securities companies, underwriting commissions and yield spreads decreased. Takaoka and McKenzie (2004) further find that commissions and spreads do not vary depending on the strength of the bank–issuer relationship, but Yasuda (2007), using a more elaborate methodology, finds that the bank–issuer relationship does have a beneficial effect.

Regulation of the Japanese banking system

Up until the banking crisis in the 1990s, the Ministry of Finance (MOF) played the dominant role in prudential supervision of most of the banking system. In 1998, bank regulatory responsibility was shifted to the new Financial Supervisory Agency, which was reorganized as the Financial Services Agency (FSA) in 2000. As can be seen in Table 35.3, the FSA supervises and charters most of the banking system, including, most importantly, the ordinary banks and financial holding companies. Some of the other private banking institutions are supervised by various ministries of the government as are all of the government financial institutions.

As the central bank, the Bank of Japan also has the ability to monitor banks (an 'on-site examination') in order to discharge its responsibilities in determining and executing monetary policy and in providing liquidity to the banking system including its role as lender of last resort (LOLR). The deposit insurance system in Japan was established in 1971 and is provided by the Deposit Insurance Corporation.

Selected topics in Japanese banking

The main bank system and relationship banking

The main bank system in Japan can be more precisely defined as a 'system of corporate financing and governance involving an informal set of practices, institutional arrangements, and behaviors among industrial and commercial firms, banks of various types, other financial institutions, and the regulatory authorities. At its core there is the relationship between the main bank and the firm' (Aoki, Patrick, and Sheard, 1994). Originally set up in wartime Japan based on zaibatsus to help coordinate wartime production, these industrial groups allegedly became a driver of economic growth during the post-war period when they became known as keiretsu (see Hoshi and Kashyap, 2001; and Teranishi, 1994).

This relationship has many dimensions including: reciprocal shareholdings, the supply of management resources and directors, and the provision of various financial services (including loans, guarantees, trustee administration, operation of settlement accounts, foreign exchange dealings, securities underwriting, and investment banking advisory services). Also important is the relationship between the main bank and the firm's other financiers (see Sheard, 1994b) and the relationships among the regulatory authorities and all of these actors. The financial institutions and firms tied together under this system are referred to as the 'financial keiretsu' (horizontal keiretsu). This can be distinguished from the concept of a 'corporate keiretsu' (vertical keiretsu) that mainly focuses on ties through vertical relationships among suppliers and sellers (see Aoki and Patrick, 1994 for more on the main bank system).

Before the empirical studies in the mid-1990s on relationship banking in the US (e.g., Petersen and Rajan, 1994; and Berger and Udell, 1995) and later in Europe (e.g., Angelini, DiSalvo, and Ferri 1998), the practitioner and academic analysis of banking and commercial lending in Japan had focused primarily on the role of Japan's main bank system. As we will see below, these earlier main-bank-focused studies share much of their theoretical foundation with the newer literature on relationship lending.

However, two distinctions need to be made between the newer literature on relationship lending and the study of the role of the main bank in the financial keiretsu. First, the main bank literature has a corporate governance component which is lacking in the newer relationship lending literature. Second, the main bank literature focuses, for the most part, on large companies, while most of the newer relationship lending literature focuses on SMEs. These distinctions are important because corporate governance issues are much less relevant in the SME sector where there is usually no separation of ownership and management. It is important to note that there have been some studies on SME lending in Japan

conducted in the spirit of the newer literature on relationship lending that focus on the relationship between SMEs and their *main banks*.

Traditional main bank studies

A large number of academic studies since the 1980s have examined the role of main banks in financial keiretsus. The early focus of these studies was on risk sharing among keiretsu members (e.g. Nakatani, 1984; and Osano and Tsutsui, 1985). The literature then gradually shifted to the role of banks as providers of corporate governance (see Aoki, 1994). Some studies emphasize a special corporate governance role of the main bank in periods of firm distress (Sheard, 1994a; and Osano, 1998). Other studies emphasize a contingent governance role for the main bank: the main bank plays a minimal role in *normal* times but in financial distress the bank assumes managerial control (e.g. Berglöf and Perotti, 1994). To a certain extent, the Japanese main banking system in terms of corporate governance can be viewed as similar to Germany's historical *hausbank* system, each standing in contrast to markets-oriented economies such as the UK and the US, where the market for corporate control and shareholder activism play a more important role (Prowse 1995).

Empirically, the issue of the very definition of 'the main bank' is challenging. A common approach is to use the information provided by the data source, such as (i) keiretsu affiliation in *Keiretsu no Kenkyu* (Studies on keiretsu) data; (ii) a first-listed bank in *Quarterly Corporate Report* (Japan Company Handbook); (iii) an affiliation in *Dodwell Marketing Consultants' Industrial Groupings in Japan*. Alternatively, the main bank is defined as a bank that (iv) dispatches a director to the borrower; (v) is the largest lender; (vi) is the largest shareholder; or (vii) is both (v) and (vi) (plus other characteristics).

Empirical studies on this subject can also be broadly classified based on their focus on the role of the main bank. Some studies emphasize the role of the main bank in mitigating liquidity constraints (e.g., Hoshi, Kashyap, and Scharfstein, 1990a; Hoshi, Kashyap, and Scharfstein, 1990b; Hoshi, Kashyap, and Scharfstein, 1991; and Ogawa and Suzuki, 2000).[7] Other studies investigate managerial intervention by main banks. Kaplan and Minton (1994), Kang and Shivdasani (1995), Kang and Shivdasani (1997), and Morck and Nakamura (1999) find that during the late 1980s, banks assigned new board members in a timely and effective manner. Shin and Korali (2004) found that main banks played a unique role in information production specific to the 1995–7 period during the banking crisis.

However, to some extent the tone of the main bank studies may reflect the timing of the studies themselves and the changing view of the Japanese economy from the more positive pre-crisis perspective to the more critical post-crisis perspective.

[7] Although the methodology in the Hoshi, Kashyap, and Scharfstein papers has been criticized (Kaplan and Zingales, 1997; and Hayashi, 2000), similar findings have been reported in a subsequent study using an improved methodology (Hori, Saito, and Ando, 2006).

There are other later studies that find that firms with a main bank exhibit weaker performance (e.g., Weinstein and Yafeh, 1998; Hanazaki and Horiuchi, 2000; and Wu and Xu, 2005), suggesting that main banks extract excessive rents from their borrowers. Kang and Shivdasani (1999) and Kang and Stultz (2000) find similar results for bank-dependent (though not necessarily main-bank-dependent) firms. Some authors even argue that the main bank system and the importance of keiretsus is a 'myth' and that many of the empirical results in this literature cannot be reproduced (Miwa and Ramseyer, 2002; and Miwa and Ramseyer, 2005).

One interpretation of these seemingly conflicting results is that the benefits from the main bank system (e.g., liquidity provision) may come at the cost of extracted rents. Interestingly, Weinstein and Yafeh (1998) find (in effect) evidence of a trade-off in that they also find that a main bank relationship may mitigate financing constraints even though it reduces firm performance. Also, the possibility of such a trade-off between the benefits and costs of a main-bank relationship had already been noted in a much earlier study (Nakatani, 1984).

It is important to add here that recently there has been a fundamental change in corporate governance in Japan—a change that has been associated with a dismantling of keiretsu ties (Aoki, Jackson, and Miyajima 2007). This suggests that for large businesses the main bank relationship may be significantly less important in the future. However, it seems unlikely that this trend has altered the importance of the main bank relationship for SMEs. This recent trend also suggests the possibility that if access to capital markets had not been constrained in post-war Japan, the role of the keiretsus and the main bank would not have been as prominent.

SME relationship lending

Relatively recently there has been growing interest in research that examines relationship lending in the context of Japanese SMEs. A recent increase in the availability of SME data in Japan has spawned new empirical work that has investigated financing constraints in the SME sector including the impact of bank–borrower relationships and the extent to which these relationships benefit borrowers through the production of soft information.[8] This coincides with increased practitioner and policy interest in SME financing and the FSA's adoption of measures that are intended to promote relationship banking between SMEs and smaller banks (the *Action Program Concerning Enhancement of Relationship Banking Functions* (2003 and 2004), and its successor program, *Ensuring Further Promotion of Regionally Based Relationship Banking* (2005 and 2006).

Findings in the academic literature on SME lending practices in Japan are interesting from an international perspective. Kano, et al. (2006) find that a lower loan interest rate and enhanced credit availability that are associated with long-term relationships between banks and borrowers are observed only when

[8] See, e.g., Boot (2000) for a summary of the theoretical and empirical work on relationship lending.

hard information is unavailable for the borrower and the bank is small and faces stiff competition. However, the lending relationship in this sample is quite long (32.2 years on average) compared with that in other countries. Uchida, Udell, and Watanabe (2008) find that the *mode* of relationship building in Japan is different from that in the US as reported by Berger, et al. (2005). Also, findings in Uchida, Udell, and Yamori (2008b) suggest that the role of the loan officer in Japan may be different from that in the US.

The Japanese banking crisis

The banking crisis in Japan spanned the entire decade of the 1990s. Given our space limitation, we offer a brief overview of the crisis, its causes, and its effects on bank behavior.

Brief review of the crisis

The visible beginning of the crisis is associated with the failures of two credit cooperatives in 1994. Ultimately there were 171 bank failures in Japan from 1994 to 2003 involving one city bank, two long-term credit banks, one regional bank, twelve second regional banks, twenty-three Shinkin banks, and 132 credit cooperatives (Nikkin, 2005). The enormity of the banking crisis was revealed in stages that progressed from the early 1990s to the early 2000s.[9] The regulatory response to the crises can best be described as one of catching up with rapidly unfolding events. In particular, the regulatory policies and infrastructure in place at the beginning of the crisis were simply not capable of handling a crisis of this magnitude. New policies and infrastructure were created to address the problem—but these were implemented with a significant lag.

A limited number of bank failures prior to 1994 were resolved in a conventional manner using arranged mergers. However, the crisis moved to a more visibly serious stage in late 1994 with failures of Tokyo Kyowa and Anzen, the two urban credit cooperatives. They were too large to be resolved by an arranged merger, and the deposit insurance fund was insufficient to cover the unprecedented losses. Concern about contagion effects persuaded regulators to avoid a payoff resolution in which depositors would take a 'haircut'. The ultimate 'hand made' (Nakaso, 2001: p. 7) nature of the resolution of these two failures involved the establishment of a new successor bank, capitalized by funds from the deposit insurance agency, the Bank of Japan, and private financial institutions, including those that had no relationship with the two cooperatives.

In the following year a number of other banks failed, including a much-larger urban cooperative. In addition, a group of real estate finance companies known as

[9] The following discussion of the stages of the banking crisis is based on Nakaso (2001).

the *jusen* failed. The jusen were initially founded by commercial banks to augment their residential mortgage lending, but by the time that they failed they had shifted their focus to financing real estate developers. Because of the collective size of these institutions, the resolution could not be handled without the use of taxpayer funding. Additional emergency measures were also undertaken at this time, including the creation of the Resolution and Collection Bank, and the temporary implementation of a 100 percent deposit insurance guarantee.[10]

Following several other bank failures, the financial crisis escalated in 1997 as it became apparent that problem loans were threatening the viability of Japan's largest banks. Nippon Credit Bank, one of the three long-term credit banks, was bailed out, two securities firms failed, and major bank failures began occurring on a regular basis in the fall, including that of Hokkaido Takushoku Bank (a city bank) and some second regional banks. Belatedly, in 1997, the government implemented emergency and permanent measures to cope with the crisis. These measures resulted in capital injections into twenty-one large banks in 1998.

Nevertheless, problems in the banking system continued to mount in 1998, including the failure of two of the long-term credit banks—the Nippon Credit Bank (bailed out earlier) and the Long Term Credit Bank of Japan, each being resolved by temporary nationalization. Early in 1998, legislation was passed that provided for further injection of $230 billion of public funds, part of which was allocated to the Deposit Insurance Corporation, the remainder being allocated to direct capital injections. Later, in 1998, the Diet passed two pieces of legislation that significantly expanded the regulatory infrastructure to handle the disposition of failed banks and to inject capital into viable banks. Also, available funds were doubled from the original $230 billion. Additionally, in June 1998, responsibility for prudential supervision of banks was shifted from the Ministry of Finance to the Financial Supervisory Agency, which was later reorganized as the Financial Services Agency, in 2000. Following these and subsequent sporadic capital injections, the crisis subsided. No major bank failure has occurred since the failure and nationalization of Ashikaga Bank (a regional bank) in 2003.

The resolution of the crisis over the decade saw the government and bank regulators deploy a variety of tools that had been used elsewhere in the world. These included establishing a bridge bank for segregating non-performing loans, and temporarily nationalizing large banks. Ultimately, these measures were associated with the injection of massive amounts of government funding to back up the 100 percent deposit insurance coverage.[11]

[10] Even before this temporary measure, deposits had been implicitly 100% guaranteed under the convoy system (Hoshi, 2002). See Hoshi (2002) for a discussion of the 'convoy' system in use until the early stages of the crisis in which the MOF protected and kept alive all financial institutions, including the most inefficient.

[11] Studies on policy responses to the crisis include Hoshi and Patrick (2000); Hoshi and Kashyap (2001); Spiegel and Yamori (2003); and Montgomery and Shimizutani (2009).

Causes of the crisis

Cargill (2000) argues that there were five underlying causes of the crisis: a rigid financial regime, the failure of the Bank of Japan's monetary policy, a slow and indecisive regulatory response to emerging problems, a lack of public support to deal with troubled financial institutions with public funds (and a lack of a political will to do so), and the intransigence of financial institutions in accepting criticism of management policies. Hoshi (2001) and Hoshi and Kashyap (2001) make a related argument that a fundamental cause of the banking crisis was the slow and incomplete deregulation of the financial system in the 1980s, which brought about the flight of good borrowers from the bank loan market (disintermediation), forcing banks to lend to borrowers with whom they were unfamiliar.

Ueda (2000) points out that excessive lending to real estate and related industries might have contributed to the accumulation of non-performing loans.[12] This factor is essentially a consequence of the financial deregulation. As discussed above, disintermediation encouraged banks to lend to unfamiliar borrowers, particularly those in the real estate industry (Hoshi and Kashyap, 2001). Banks lent aggressively to this industry both directly and indirectly through affiliates.

Ueda (2000) speculates that rising land prices decreased the perceived risk in real estate lending, prompting more lending *ex ante*. He also finds evidence that the big drop in land prices *ex post* was an important determinant of deteriorating non-performing loans. A common lending practice in Japan—the so-called 'collateral principle' (a tendency to lend against real estate even in general-purpose business lending)—might also have exacerbated this pathology. Ogawa, et al. (1996) and Ogawa and Suzuki (2000) found that large firms with more real estate were less financially constrained than those with less real estate. However, post-crisis analysis suggests that the collateral principle is no longer a dominant SME loan underwriting technology (see Uchida, Udell, and Yamori, 2008a).

Anecdotal evidence indicates that euphoric lending practices driven by the formation of the asset price bubble ultimately led to problem loans in the banking system.[13] This is partly consistent with the 'lazy bank' hypothesis (Manove, Padilla, and Pagano, 2001). Irrational herding behavior may also have occurred (Uchida and Nakagawa, 2007).[14] Kashyap (2002), however, argues that the loan problems that were ultimately revealed in the later stages of the crisis were simply too large to

[12] He also finds evidence of inefficient ('lax') bank management and the safety-net-driven moral hazard problem as factors driving poor loan performance.

[13] 'Reckless lending' more broadly defined could include the practice of evergreening (the behavior of banks to keep zombie firms alive), as we discuss below.

[14] A practice in Japan, known as *amakudari*, of hiring retired government officials as board members may also have been a factor that leads to non-performing loans. Horiuchi and Shimizu (2001) find that for regional banks from 1979 to 1991 a greater number of amakudari was associated with lower bank capital asset ratios and higher non-performing loan ratios. However, another study using a more elaborated methodology did not find this association (Konishi and Yasuda, 2004).

be entirely attributable to euphoric or reckless lending during the formation of the bubble. The arguments about euphoric and irrational lending notwithstanding, the evidence seems clear that the initial cause of problems in the banking sector was a pricing bubble in real estate that burst around 1990 (Hoshi and Kashyap, 2008).[15]

The effects of the crisis

Did the banking crisis lead to a 'credit crunch'? It is clear that the policymakers at the Bank of Japan in January of 1998 concluded that the country was suffering from a 'credit crunch' noting in the minutes of their meeting that the 'prospects for a more restrictive lending attitude of financial institutions and its possible effects were discussed in detail'. Also, a quarterly survey on business expectations conducted by the Bank of Japan—the Tankan (short-term economy) survey—showed a significant shift in the perception of credit tightening by corporate Japan beginning in late 1997.[16]

A number of research papers have shown that bank capital deterioration and the decline in the health of the banking industry reduced bank lending (Ito and Sasaki, 2002), firm capital expenditures, and firm performance (Gibson, 1997; Hosono and Masuda, 2005; Fukuda, Kasuya, and Nakajima, 2005; and Miyajima and Yafeh, 2007). Woo (2003) and Watanabe (2007) find that the negative impact was the greatest in fiscal year 1997, when the MOF became stricter on bank asset valuation.[17] The view that the 'credit crunch' had significant real effects, however, is not universal. Hayashi and Prescott (2002a) provide evidence showing that the stagnation of the Japanese economy in the 1990s was not due to a breakdown of the financial system, but rather due to low-productivity growth (measured as total factor productivity (TFP)) in the real economy. Motonishi and Yoshikawa (1999) also report that, at least as far as large firms are concerned, tighter lending attitudes in banks did not constrain corporate investments, although they also found that small firms were constrained by tighter lending behavior.

Nevertheless, there are some studies that claim that the low TFP was a consequence of banking sector problems. Peek and Rosengren (2005) found that the practice of 'evergreening' (where banks roll over or renew problem loans to make them appear performing) contributed to a decrease in TFP. The practice of evergreening facilitates the perpetuation of economically unviable firms ('zombies') who should otherwise be liquidated. The evidence indicates that the perpetuation

[15] There are a number parallels between the Japanese crisis in the 1990s and the recent financial crisis in the US including the root cause—the bursting of a real estate bubble. Other similarities include the failure of investment banks and other non-banking financial institutions and the 'handmade'/ad hoc nature of the regulatory response. For a more complete comparison of the Japanese banking crisis and the current financial crisis in the US, see Hoshi and Kashyap (2008); and Udell (2009).

[16] For a more detailed discussion of the minutes of the Bank of Japan meeting in January 1998 and the Tankan survey, see Hoshi and Kashyap (2008).

[17] Some studies also found that the introduction of Basel capital standards may have reduced bank lending (Hall (1993); Honda (2002); and Konishi and Yasuda (2004)).

of these zombies caused real economic distortion and led to lower productivity (Caballero, Hoshi, and Kashyap, 2008). There is also some evidence suggesting that the Darwinian natural selection process by which good firms survive and bad firms disappear was inoperative during the crisis (e.g., Nishimura, Nakajima, and Kiyota, 2005).

It should be noted that the 'capital crunch' story and the evergreening story have the opposite predictions on lending. The former predicts a decrease in lending, while the latter predicts an increase. Further research on which effect dominated would be helpful. Also, the issue of real vs. financial stagnation and the chicken and egg (i.e., causality) problem between economic stagnation and the banking crisis have not been fully investigated.

Japanese banks and the subprime crisis

Soon after the persistent adverse effect from the banking crisis subsided in Japan, the turmoil in the US subprime residential mortgage market beginning in early 2007 triggered an unprecedented worldwide financial crisis (hereafter called the 'subprime crisis'). Described by some as a 'once in a century credit tsunami',[18] the crisis has caused serious direct and collateral damage to financial institutions worldwide. This section provides an overview of the impact of the subprime crisis on Japanese banks as of December 2008.

Residential mortgage markets in Japan

Before turning to the impact of the subprime crisis, we take a brief look at the residential mortgage market in Japan.[19] First, we note that there is no subprime component to the market in Japan. That is, there is no mortgage market in Japan that specifically targets low-income borrowers. Second, Japan did not experience a real estate bubble during this past decade when the US experienced its bubble. On average, real estate prices nationwide, based on the land price index of the Japan Real Estate Institute (an average of commercial, residential, and industrial land indices), have been steadily decreasing since the Japanese real estate bubble burst in 1990. The prices in the six largest city areas also fell continuously until 2004. However, since then they have risen by about 25 percent from 68.6 in September 2004 to 84.8 in March 2008. Nevertheless, real estate prices in these six cities have still not reached the 2000 level (= 100).[20]

[18] Testimony of Alan Greenspan (the former chairman of the Board of Governors of the Federal Reserve System) before the US Committee of Government Oversight and Reform (23 October 2008).

[19] In Japan, the term '*ju-taku rohn*' (loans for housing, or housing loans) is more commonly used than 'residential mortgages'. Although they may have some different nuance, we use the terms interchangeably.

[20] The index in March 1990, around the peak of the bubble, was 276.8.

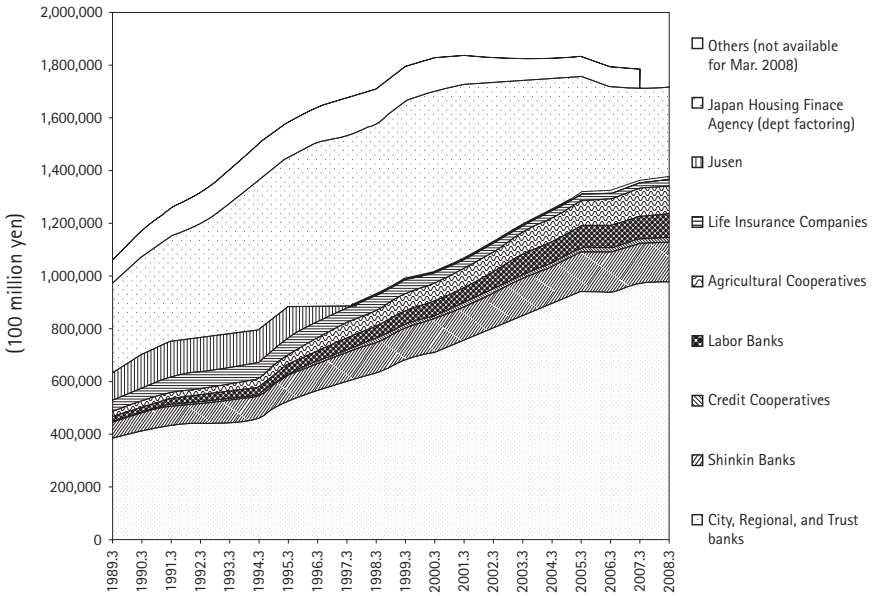

Fig. 35.4. Mortgage loans outstanding in Japan (by lender type)

Figure 35.4 shows housing loans outstanding (loans to individuals to purchase homes and land) in Japan by lender type. As shown in the figure, the primary supplier of residential mortgages had historically been the Japan Housing Finance Agency (formerly the Government Housing Loan Corporation). Today, however, the largest providers of housing loans in Japan are city, regional, and trust banks, with city banks being the most important. The Government Housing Loan Corporation had been engaged in direct lending using the funds supplied to it through the Fiscal Investment and Loan Program (FILP). When the corporation was reorganized into the Japan Housing Finance Agency, on 1 April 2007, it then ended its direct lending, and focused instead on facilitating securitization of residential mortgages by private banks.

Despite these efforts, the securitization market for residential mortgages, as well as for other financial instruments, is still underdeveloped in Japan. Figure 35.5 shows the amount of securitized products issued in the last four years. Although residential mortgages compose more than half of the securitized assets, the amount is still quite small compared to the amount of outstanding mortgages in the US.

To summarize, over the past decade, Japan's residential mortgage market did not suffer from the same fate as that in the US: subprime mortgages do not exist in Japan; securitization is relatively nascent in Japan; and Japan did not experience a real estate bubble.

Unfortunately, a sound domestic housing loan market did not insulate Japan from the global financial crisis. Indeed, one of the characteristics of this current

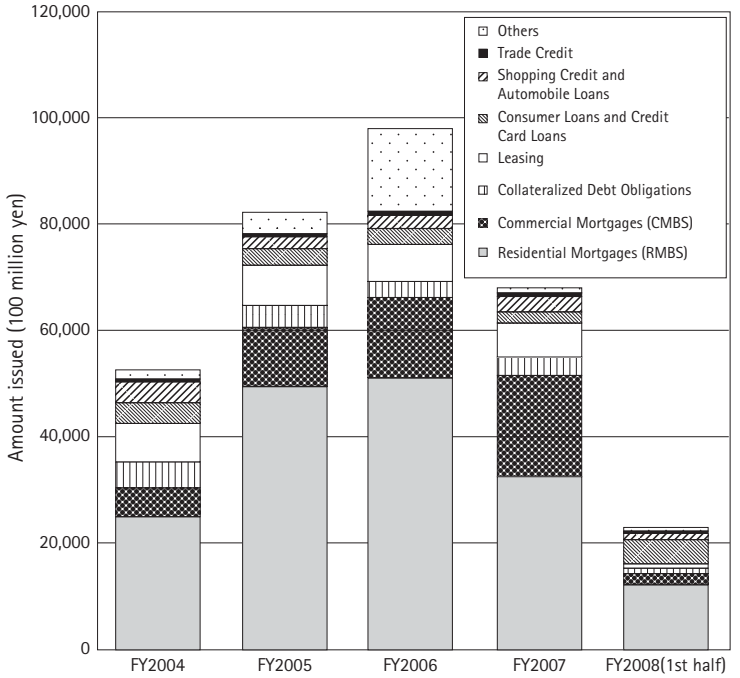

Fig. 35.5. Securitized products issued in Japan (categorized by underlying asset)

subprime crisis is the contagious effect of turmoil in the US on other countries, including Japan. We turn to this topic next.

Japanese bank exposure to the subprime crisis

The magnitude of the damage to Japanese financial institutions from the current crisis is still too difficult to assess with any accuracy. Available information, however, indicates that the exposure of Japanese banks to US subprime securities was on average low. According to statistics compiled by the Financial Services Agency, the book value of outstanding subprime products for major banks, regional banks, and cooperative banks are respectively 719 billion, 46 billion, and 31 billion yen on 30 September 2008 (versus 1,246 billion, 115 billion, and 47 billion yen on 30 September 2007).[21] These figures are small relative to the size of their core

[21] Subprime products here are: ABS backed by subprime loans or CDOs and other financial products constructed from these ABS (excluding the exposures to subprime-related products through mutual funds). Major banks in this context include not only the banks normally considered to be the 'major banks' (Mizuho Bank, Mizuho Corporate Bank, Mizuho Trust Bank, Bank of Tokyo–Mitsubishi UFJ, Mitsubishi UFJ Trust Bank, Sumitomo Mitsui Banking Corporation, Resona Bank, Chuo-Mitsui Trust Bank, and Sumitomo Trust Bank (including their group securities companies)), but also Norinchukin Bank, Shinsei Bank, Aozora Bank, Citibank Japan, 'new' types of banks, foreign trust banks, and others. Regional banks include regional, second regional banks, and Saitama Resona

net business profits (3,499 billion, 1,799 billion, and 795 billion yen for fiscal year 2007, respectively). For example, for the largest three banking groups (Mizuho, Mitsui Sumitomo, and Mitsubishi-Tokyo UFJ), investment in subprime real estate 'mortgage-backed securities' (MBS) and subprime 'asset back securities' (ABS) or 'collateralized debt obligations' (CDOs) at the end of March 2008 consisted of less than 1 percent of their banking assets.

The Bank of Japan's *Financial System Report* (September 2008) notes that '[w]hile Japanese banks' losses stemming from the US subprime mortgage problem increased as the problem became more serious, such losses seem to have been contained within their current profit levels and capital strength, since Japanese banks' related exposures were mainly in the form of investments in structured credit products'. Significantly, no banks have yet failed in Japan due to this crisis. The only failure of a financial institution since the beginning of the crisis has been a medium-sized life insurance company. Also, as in the US, volatility in the Japanese stock market spiked upward around the failure of Lehman Brothers, and large firms encountered increasing difficulty in raising funds in malfunctioning capital markets. Bank lending may thus have become relatively more important.[22] Interestingly, during this acute period of turmoil in September 2008, when Lehman Brothers failed, Mitsubishi UFJ financial group (the holding company of the Bank of Mitsubishi Tokyo UFJ) injected equity into a floundering Morgan Stanley.[23]

Again, this does not mean, however, that banks in Japan were entirely insulated from the crisis. They have certainly suffered some collateral damage. As explained above, banks in Japan are able to invest in securities. Owing to the worldwide fall in asset prices Japanese banks have been hit by losses in their securities portfolios. This is reflected in Table 35.5, where the eleven largest banks reported losses on securities portfolios of 179 billion yen in the first half of fiscal year 2008. Also, the slowdown in the world economy has affected loan losses. Disposal of bad loans has increased significantly for both the largest banks and the regional banks (see Table 35.5).[24]

Larger banks, including Mitsubishi UFJ Financial Group, which invested in Morgan Stanley, are planning to raise billions in capital in the equity markets to offset decreasing profits and to increase their lending capacity to cope with the increasing demand (*Nikkei*, 26 November 2008). The Bank of Japan cut interbank lending rates from 0.5 percent to 0.2 percent on 31 October 2008, and further to 0.1

Bank. Cooperative banks include not only Shinkin banks and credit cooperatives but also Shinkin Central Bank, Shinkumi Federation Bank, Labour Banks, Rokinren Bank, Prefectural Banking Federations of Agricultural Cooperatives, and Prefectural Banking Federations of Fishery Cooperatives.

[22] Nikkei newspaper reports (10 December 2008) that some large firms are setting new lines of credit, while other large firms increased borrowing, due to the absence of investors in corporate bonds and commercial papers.

[23] In the same period, Nomura Security, the largest securities company in Japan, acquired Asian and European divisions of failed Lehman Brothers.

[24] Disposal of bad loans includes the provision for allowances for loan losses, net charge-offs, losses from loan sales, and provisions for other allowances.

Table 35.5. Latest profits and losses for major banks and regional banks

(Bank type)	(period)	Disposal of bad loans (billion yen)	Losses on sales and depreciation losses for equity and bond holdings (billion yen)	Net profits (billion yen)
11 Major Banks*	1st half of Fiscal 2008	−779.8	−179.0	334.4
	1st half of Fiscal 2007	−404.2	2.0	830.6
	1st half of Fiscal 2006	187.2	143.7	1,683.0
Regional and Second Regional Banks**	1st half of Fiscal 2008	−537.3	16.3	139.4
	1st half of Fiscal 2007	−377.2	77.8	374.7
	1st half of Fiscal 2006	−347.9	93.5	402.8

Source: Financial Services Agency (basd on disclosure information by Individual banks).

*Mizuho Bank, Mizuho Corporate Bank, Mizuho Trust Bank, Bank of Mitsubishi-Tokyo UFJ, Mitsubishi UFJ Trust Bank, Mitsui Sumitomo Bank, Resona Bank, Chuo Mitsui Bank, Sumitomo Trust Bank, Shinsei Bank, and Aozora Bank.
**64 regional banks, 45 second regional banks, and Saitama Resona Bank.

percent on 19 December, and has undertaken measures for 'monetary easing' in order to prevent a 'credit crunch'. On 12 December 2008, the diet revived the already expired Act on Special Measures for Strengthening Financial Functions, which enables preventive capital injection into banks.

Judging from the evidence above, the direct damage to Japanese banks from the crisis appears relatively small because they substantially avoided investing in subprime real estate. Nevertheless, they have certainly suffered non-negligible collateral damage. Japan's 'institutional memory' of its own crisis may have been a positive restraining factor here (Berger and Udell, 2004). However, it is also possible that recovering Japanese banks simply could not afford to purchase securitized US subprime mortgages and other risky securitized products as did US and European banks. When the subprime market took off in 2003–4, Japanese banks were still fighting through the after-effect of non-performing loan problems from the prior decade. Japanese banks did not return to adequate capitalization until 2006 and 2007 following improvement in macroeconomic conditions and changes in bank-supervision policies (Hoshi and Kashyap, 2008).

Conclusions

In this chapter we examined the structure, the performance, and some of the defining characteristics of the Japanese banking industry. In addition to this overview, we reviewed three interesting topics related to the Japanese banking

system: the Japanese main bank system, the Japanese banking crisis of the 1990s, and the impact of the subprime crisis on Japanese banks.

We conclude this chapter by pointing out the scarcity of research on Japanese banking. Even on the selected topics discussed above there remain many open questions. For example, how and to what extent are banking markets in Japan segmented? Are there economies of scale or scope in banks in Japan, and to what extent? What were, and what are, the pros and cons of the main bank system? Did the banking crisis cause the prolonged stagnation in the Japanese economy, or vice versa?

The future of the Japanese banking industry is also an open question. The market structure described in the second section has persisted since the late 1990s. But, is it a temporary artifact of the banking crisis, or will it persist in the future? Will the keiretsu-driven ties between firms and their main banks weaken over time, or will they persist? Will the collateral damage to the Japanese banking system caused by the subprime crisis level off or intensify? Will less damaged large Japanese banks regain their stature in the international banking market?

The banking-oriented Japanese financial system has been a critical component of the country's economy—an economy that has grown to one of the largest in the world. Despite the idiosyncratic nature of Japanese banking, its seems quite likely that there is much we can learn from the Japanese experience that will inform us more generally about the role of banks in the global financial system architecture. More research on the banking industry in Japan is clearly called for.

References

Altunbas, Y., Liu, M.-H., Molyneux, P., and Seth, R. (2000). 'Efficiency and Risk in Japanese Banking', *Journal of Banking & Finance*, 24: 1605–28.

Angelini, P., DiSalvo, R., and Ferri, G. (1998). 'Availability and Cost of Credit for Small Businesses: Customer Relationships and Credit Cooperatives', *Journal of Banking & Finance*, 22: 925–54.

Aoki, M. (1994). 'Monitoring Characteristics of The main Bank System: An Analytical and Developmental View', in *The Japanese Main Bank System*, Aoki, M. and Patrick, H. T. (eds.), New York, Oxford University Press.

—— Jackson, G., and Miyajima, H. (2007). *Corporate Governance in Japan,* Oxford: Oxford University Press.

—— and Patrick, H. T. (1994). *The Japanese Main Bank System: Its Relevance for Developing and Transforming Economies*, New York: Oxford University Press.

—— —— and Sheard, P. (1994). 'The Japanese Main Bank System: An Overview', in *The Japanese Main Bank System*, Aoki, M. and Patrick, H. T. (eds.), Oxford: Oxford University Press.

Bank of Japan (2008), *Financial System Report* (September 2008), available at http://www.boj.or.jp/en/type/ronbun/fsr/fsr08b.htm.

Berger, A. N., Hunter, W. C., and Timme, S. G. (1993). 'The Efficiency of Financial Institutions: A Review and Preview of Research Past, Present and Future', *Journal of Banking & Finance*, 17: 221–49.

—— Miller, N. H., Petersen, M. A., Rajan, P. G., and Stein, J. C. (2005). 'Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks', *Journal of Financial Economics*, 76: 237–69.

—— and Udell, G. F. (1995). 'Relationship Lending and Lines of Credit in Small Firm Finance', *Journal of Business*, 68: 351–81.

—— —— (2004). 'The Institutional Memory Hypothesis and the Procyclicality of Bank Lending Behavior', *Journal of Financial Intermediation*, 13: 458–95.

Berglöf, E. and Perotti, E. (1994). 'The Governance Structure of the Japanese Financial Keiretsu', *Journal of Financial Economics*, 36: 259–84.

Boot, A. W. A. (2000). 'Relationship Banking: What Do We Know?', *Journal of Financial Intermediation*, 9: 7–25.

Caballero, R. J., Hoshi, T., and Kashyap, A. K. (2008). 'Zombie Lending and Depressed Restructuring in Japan', *American Economic Review*, 98: 1943–77.

Cargill, T. F. (2000). 'What Caused Japan's Banking Crisis?', in *Crisis and Change in the Japanese Financial System*, Hoshi, T. and Patrick, H. T., New York: Kluwer Academic Publishers, 37–58.

—— and Yoshino, N. (2000). 'The Postal Savings System, Fiscal Investment and Loan Program, and Modernization of Japan's Financial System', in *Crisis and Change in the Japanese Financial System*, Hoshi, T. and Patrick, H. T., New York: Kluwer Academic Publishers.

Drake, L. and Hall, M. J. B. (2003). 'Efficiency in Japanese Banking: An Empirical Analysis', *Journal of Banking & Finance*, 27: 891–917.

—— Hall, M. J. B., and Simper, R. (2009). 'Bank Modeling Methodologies: A Comparative Non-Parametric Analysis of Efficiency in the Japanese Banking Sector', *Journal of International Financial Markets, Institutions and Money*, 1: 1–15.

Flath, D. (2000). *The Japanese Economy*, Oxford: Oxford University Press.

Fukanuma, H., Nemoto, T., and Watanabe, W. (2006). 'Do Governmental Financial Institutions Help Startups Grow? Evidence from Japan', National Life Finance Corporation, Chuo University and Keio University.

Fukuda, S.-I., Kasuya, M., and Nakajima, J. (2005). 'Bank Health and Investment: An Analysis of Unlisted Companies in Japan', Bank of Japan Working Paper Series 05-E-5.

Fukuyama, H. (1993). 'Technical and Scale Efficiency of Japanese Commercial Banks: A Non-Parametric approach', *Applied Economics*, 25: 1101–12.

—— (1996). 'Returns to Scale and Efficiency of Credit Associations in Japan: A Nonparametric Frontier Approach', *Japan and the World Economy*, 8: 259–77.

—— Guerra, R., and Weber, W. L. (1999). 'Efficiency and Ownership: Evidence from Japanese Credit Cooperatives', *Journal of Economics & Business*, 51: 473–87.

Gibson, M.S. (1997). 'More Evidence on the Link between Bank Health and Investment in Japan', *Journal of the Japanese and International Economies*, 11: 296–310.

Hall, B. J. (1993). 'How Has the Basel Accord Affected Bank Portfolios?', *Journal of the Japanese and International Economics*, 7: 408–40.

Hamao, Y. and Hoshi, T. (2000). 'Bank-Owned Security Subsidiaries in Japan: Evidence after the 1993 Financial System Reform', in *Finance, Governance, and Competitiveness in Japan*, Aoki, M. and Saxonhouse, G. R., Oxford: Oxford University Press.

HANAZAKI, M. and HORIUCHI, A. (2000). 'Is Japan's Financial System Efficient?', *Oxford Review of Economic Policy*, 16: 61–73.

HARIMAYA, K. (2008). 'Impact of Nontraditional Activities on Scale and Scope Economies: A Case Study of Japanese Regional Banks', *Japan and the World Economy*, 20: 175–93.

HAYASHI, F. (2000). 'The Main Bank System and Corporate Investment: An Empirical Reassessment', in *Finance, Governance, and Competitiveness in Japan*, Aoki, M. and Saxonhouse, G. R., Oxford: Oxford University Press.

—— and PRESCOTT, E.C. (2002). 'The 1990s in Japan: A Lost Decade', *Review of Economic Dynamics*, 5: 206–35.

HONDA, Y. (2002). 'The Effects of the Basel Accord on Bank Credit: The Case of Japan', *Applied Economics*, 34: 1233–39.

HORI, K., SAITO, M., and ANDO, K. (2006). 'What Caused Fixed Investment to Stagnate during the 1990s in Japan? Evidence from Panel Data of Listed Companies', *Japanese Economic Review*, 57: 283–306.

HORIUCHI, A. (2000). 'The Big Bang: Idea and Reality', in *Crisis and Change in the Japanese Financial System*, Hoshi, T. and Patrick, H. T., Dordrecht: Kluwer Academic Publishers.

—— and SHIMIZU, K. (2001). 'Did Amakudari Undermine the Effectiveness of Regulator Monitoring in Japan?', *Journal of Banking & Finance*, 25: 573–96.

HOSHI, T. (2001). 'What Happened to Japanese Banks?', *Bank of Japan Monetary and Economic Studies*, 19: 1–30.

—— (2002). 'The Convoy System for Insolvent Banks: How it Originally Worked and Why it Failed in the 1990s', *Japan and the World Economy*, 14: 155–80.

—— and KASHYAP, A. (2001). *Corporate Financing and Corporate Governance in Japan*, Cambridge, MA: MIT Press.

———— A. (2008). 'Will the U.S. Bank Recapitalization Succeed? Lessons from Japan'. Mimeo. University of California San Diego and University of Chicago.

———— and SCHARFSTEIN, D. (1990a). 'Bank Monitoring and Investment: Evidence from the Changing Structure of Japanese Corporate Banking Relationships', in *Asymmetric Information, Corporate Finance, and Investment*, Hubbard, R. G., Chicago: University of Chicago Press.

———— ———— (1990b). 'The Role of Banks in Reducing the Costs of Financial Distress in Japan', *Journal of Financial Economics*, 27: 67–88.

———— ———— (1991). 'Corporate Structure, Liquidity, and Investment: Evidence from Japanese Industrial Groups', *Quarterly Journal of Economics*, 106: 33–60.

—— and PATRICK, H. T. (2000). 'The Japanese Financial System: An Introductory Overview', in *Crisis and Change in the Japanese Financial System*, Hoshi, T. and Patrick, H. T., Dordrecht: Kluwer Academic Publishers, 1–36.

HOSONO, K. and MASUDA, A. (2005). 'Bank Health and Small Business Investment: Evidence from Japan', Research Institute of Economy, Trade, and Industry Discussion Paper 05-E-030.

ISHIKAWA, D. and TSUTSUI, Y. (2006). 'Has the Credit Crunch Occurred in Japan in 1990s?' Research Institute of Economy, Trade, and Industry *Discussion Paper* 06-E-012.

ITO, T., PATRICK, H. T., and WEINSTEIN, D. E. (2005). *Reviving Japan's Economy: Problems and Prescriptions,* Cambridge, MA: MIT Press.

—— and SASAKI, Y. N. (2002). 'Impacts of the Basel Capital Standard on Japanese Banks' Behavior', *Journal of the Japanese and International Economies*, 16: 372–97.

Kang, J.-K. and Shivdasani, A. (1995). 'Firm Performance, Corporate Governance, and Top Executive Turnover in Japan', *Journal of Financial Economics*, 38: 29–58.
—— and Shivdasani, A. (1997). 'Corporate Restructuring during Performance Declines in Japan', *Journal of Financial Economics*, 46: 29–65.
—— —— (1999). 'Alternative Mechanisms for Corporate Governance in Japan: An Analysis of Independent and Bank-Affiliated Firms', *Pacific-Basin Finance Journal*, 7: 1–22.
—— and Stulz, R. (2000). 'Do Banking Shocks Affect Borrowing Firm Performance? An Analysis of the Japanese Experience', *Journal of Business*, 73: 1–23.
Kano, M. and Tsutsui, Y. (2003). 'Geographical Segmentation in Japanese Bank Loan Markets', *Regional Science and Urban Economics*, 33: 157–74.
Kano, M., Uchida, H., Udell, G. F., and Watanabe, W. (2006). 'Information Verifiability, Bank Organization, Bank Competition and Bank-Borrower Relationships', Research Institute of Economy, Trade, and Industry Discussion Paper 06-E-003.
Kaplan, S. N. and Minton, B. (1994). 'Appointments of Outsiders to Japanese Boards: Determinants and Implications for Managers', *Journal of Financial Economics*, 36: 225–57.
—— and Zingales, L. (1997). 'Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?', *Quarterly Journal of Economics*, 112: 169–215.
Kashyap, A. (2002). 'Sorting out Japan's Financial Crisis', *Federal Reserve Bank of Chicago Economic Perspectives*, 26: 42–55.
Konishi, M. and Yasuda, Y. (2004). 'Factors Affecting Bank Risk Taking: Evidence from Japan', *Journal of Banking & Finance*, 28: 215–32.
Manove, M., Padilla, A. J., and Pagano, M. (2001). 'Collateral versus Project Screening: A Model of Lazy Banks', *RAND Journal of Economics*, 32: 726–44.
McKillop, D. G., Glass, J. C., and Morikawa, Y. (1996). 'The Composite Cost Function and Efficiency in Giant Japanese Banks', *Journal of Banking & Finance*, 20: 1651–71.
Miwa, Y. and Ramseyer, J. M. (2002). 'The Fable of the Keiretsu', *Journal of Economics & Management Strategy*, 11: 169–224.
—— and Ramseyer, J. M. (2005). 'Does Relationship Banking Matter? The Myth of the Japanese Main bank'. *Journal of Empirical Legal Studies*, 2: 261–302.
Miyajima, H. and Yafeh, Y. (2007). 'Japan's Banking Crisis: An Event-Study Perspective', *Journal of Banking & Finance*, 31: 2866–85.
Montgomery, H. and Shimizutani, S. (2009). 'The Effectiveness of Bank Recapitalization Policies in Japan', *Japan and the World Economy*, 21: 1–25.
Molyneux, P. J., Thornton, P., and Lloyd-Williams, D. M. (1996). 'Competition and Market Contestability in Japanese Commercial Banking', *Journal of Economics & Business*, 48: 33–45.
Morck, R. and Nakamura, M. (1999). 'Banks and Corporate Control in Japan', *Journal of Finance*, 54: 319–39.
Motonishi, T. and Yoshikawa, H. (1999). 'Causes of the Long Stagnation of Japan during the 1990's: Financial or Real?', *Journal of the Japanese International Economies*, 13: 181–200.
Nakaso, H. (2001). 'The Financial Crisis in Japan during the 1990s: How the Bank of Japan Responded and the Lessons Learnt'. Bank for International Settlements Papers 6.
Nakatani, I. (1984). 'The Economic Role of Financial Corporate Grouping', in *The Economic Analysis of the Japanese Firm*, Aoki, M., Amsterdam: Elsevier.
Nikkin (2005). *Nikkin Data Annual* (in Japanese), http.//www.nikkin.co.jp/.

Nishimura, K. G., Nakajima, T., and Kiyota, K. (2005). 'Does the Natural Selection Mechanism Still Work in Severe Recessions? Examination of the Japanese Economy in the 1990s', *Journal of Economic Behavior and Organization*, 58: 53–78.

Ogawa, K., Kitasaka, S.-I., Yamaoka, H., and Iwata, Y. (1996). 'Borrowing Constraints and the Role of Land Asset in Japanese Corporate Investment Decision', *Journal of the Japanese and International Economies*, 10: 122–49.

—— and Suzuki, K. (2000). 'Demand for Bank Loans and Investment under Borrowing Constraints: A Panel Study of Japanese Firm Data', *Journal of the Japanese and International Economies*, 14: 1–21.

Osano, H. (1998). 'Default and Renegotiation in Financial Distress in the Multiple Bank Model: An Analysis of the Main Bank System', *Japanese Economic Review*, 49: 138–57.

—— and Tsutsui, Y. (1985). 'Implicit Contracts in the Japanese Bank Loan Market', *Journal of Financial and Quantitative Analysis*, 20: 211–30.

Peek, J. and Rosengren, E. S. (2005). 'Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan', *American Economic Review*, 95: 1144–66.

Petersen, M. A. and Rajan, R. G. (1994). 'The Benefits of Lending Relationships: Evidence from Small Business Data', *Journal of Finance*, 49: 3–37.

Prowse, S. (1995). 'Corporate Governance in an International Perspective: A Survey of Corporate Control Mechanisms among Large Firms in the USA, UK, Japan and Germany', *Financial Markets, Institutions & Instruments*, 4: 1–63.

Royama, S. (2000). 'The Big Bang in Japanese Securities Markets', in *Crisis and Change in the Japanese Financial System*, Hoshi, T. and Patrick, H. T., Dordrecht: Kluwer Academic Publishers.

Sheard, P. (1994a). 'Main Banks and the Governance of Financial Distress', in *The Japanese Main Bank System*, Aoki, M. and Patrick, H. T., New York: Oxford University Press.

—— (1994b). 'Reciprocal Delegated Monitoring in the Japanese Main Bank System', *Journal of the Japanese & International Economies*, 8: 1–21.

Shin, G. H. and Kolari, J. W. (2004). 'Do Some Lenders Have Information Advantages? Evidence from Japanese Credit Market Data', *Journal of Banking & Finance*, 28: 2331–51.

Spiegel, M. M. and Yamori, N. (2003). 'The Impact of Japan's Financial Stabilization Laws on Bank Equity Values', *Journal of the Japanese and International Economies*, 17: 263–82.

Tachibanaki, T., Mitsui, K., and Kitagawa, H. (1991). 'Economies of Scope and Shareholding of Banks in Japan', *Journal of the Japanese and International Economics*, 5: 261–81.

Tadesse, S. (2006). 'Consolidation, Scale Economies and Technological Change in Japanese Banking', *Journal of International Financial Markets, Institutions and Money*, 16: 425–45.

Takaoka, S. and McKenzie, C. R. (2004). 'The Impact of Bank Entry in the Japanese Corporate Bond Underwriting Market', *Journal of Banking & Finance*, 30: 59–83.

Teranishi, J. (1994). 'Loan Syndication in War-Time Japanese and the Origins of the Main Bank System', in *The Japanese Main Bank System*, Aoki, M. and Patrick, H., New York: Oxford University Press.

Uchida, H. and Nakagawa, R. (2007). 'Herd Behavior in the Japanese Loan Market: Evidence from Bank Panel Data', *Journal of Financial Intermediation*, 16: 555–83.

—— and Tsutsui, Y. (2005). 'Has Competition in the Japanese Banking Sector Improved?' *Journal of Banking & Finance*, 29: 419–39.

—— Udell, G. F., and Watanabe, W. (2008). Bank Size and Lending Relationships in Japan, *Journal of the Japanese and International Economies* 22: 242–67.

Uchida, H. and Yamori, N. (2006). 'SME Financing and the Deployment of Lending Technology', Wakayama University, Indiana University, and Nagoya University.

——— Udell, G. F. (2008a) 'How Do Japanese Banks Discipline Small and Medium - Sized Borrowers?: An Investigation of the Deployment of Lending Technologies', *International Finance Review*, vol. 9 (Institutional Approach to Global Corporate Governance), 2008.

——— ——— (2008b). 'Loan Officers and Relationship Lending to SMEs', Federal Reserve Bank of San Francisco, Working Paper Series 2008–17.

Udell, G. F. (2009). 'Wall Street, Main Street, and a Credit Crunch: Thoughts on the Current Financial Crisis'. *Business Horizons* 2: 117–25.

Ueda, K. (2000). 'Causes of Japan's Banking Problems in the 1990s', in *Crisis and Change in the Japanese Financial System*, Hoshi, T. and Patrick, H. T., Dordrecht: Kluwer Academic Publishers.

Watanabe, W. (2007). 'Prudential Regulation and the "Credit Crunch": Evidence from Japan', *Journal of Money, Credit and Banking*, 39: 639–65.

Weinstein, D. E. and Yafeh, Y. (1998). 'On the Costs of a Bank-Centered Financial System: Evidence from the Changing Main Bank Relations in Japan', *Journal of Finance*, 53: 635–72.

Woo, D. (2003). 'In Search of "capital crunch": Supply Factors behind the Credit Slowdown in Japan', *Journal of Money, Credit, and Banking*, 35: 1019–38.

Wu, W. and Xu, L. L. (2005). 'The Value Information of Financing Decisions and Corporate Governance during and after the Japanese Deregulation', *Journal of Business*, 78: 243–80.

Yasuda, A. (2007). 'Bank Relationships and Underwriter Competition: Evidence from Japan', *Journal of Financial Economics*, 86: 369–404.

CHAPTER 36

BANKING IN DEVELOPING NATIONS OF ASIA

AN OVERVIEW OF RECENT CHANGES IN OWNERSHIP STRUCTURE

LEORA KLAPPER
MARIA SOLEDAD MARTINEZ-PERIA
BILAL ZIA[1]

Introduction

Developing economies in East and South Asia have seen their banking sectors go through important transformations. Throughout the last decade, pressures brought on by the 1997 East Asian crisis and by the global trend toward increasing financial integration have resulted in significant reforms and structural changes. In particular, countries in East and South Asia have embarked in efforts to clean up

[1] We thank Ariel BenYishay, Soledad Lopez, and Subika Farazi for excellent research assistance.

their banking systems, reduce state ownership, and allow greater foreign participation. However, progress has not been even and some economies have made strides in these areas more quickly than others. While countries like Pakistan and Korea have been fairly aggressive in their reform agendas, others like India and China have proceeded more slowly. In particular, the banking sectors in India and China remain dominated by government-owned banks and are still fairly closed to foreign interests. Also, relative to other regions such as Latin America and Eastern Europe, reform in East and South Asia—in particular privatization and foreign entry—has been less comprehensive and far-reaching.

After presenting some basic statistics on the banking sectors of East and South Asia, this chapter focuses on characterizing the regions' banking sector ownership structure, discussing the recent reforms and changes that have given way to the current structure, and summarizing the evidence available so far on the effects of the reform process. Finally, in light of the recent global financial turmoil, the chapter ends by discussing the likely impact of the crisis on the reform process and the banking sectors in Asia. The chapter is organized as follows. The second section presents some basic statistics on the size, depth, efficiency, and outreach of banking sectors in East and South Asia. In the case of East Asia, we focus on China, Indonesia, Korea, Malaysia, Philippines, and Thailand, while for South Asia we center our description on Bangladesh, India, and Pakistan. The third section illustrates the degree of public and foreign ownership across East and South Asian economies and compares it to the experience of developing countries in other regions. The fourth section describes the recent reforms and transformations that have led to the current banking structure. The fifth section reviews the existing evidence on the implications from the recent ownership changes and the sixth section concludes and discusses the probable consequences of the current financial turmoil on developing countries in Asia.

Characterizing banking sectors in Asia

Banking sectors in East and South Asia have many similarities, but are also quite different from each other in a number of respects. Also, within each of these regions there are important differences across countries. For example, in terms of size, relative to GDP, the banking sector is twice as large in East Asia as it is in South Asia (see Table 36.1). On average, in East Asia, bank assets account for roughly 81 percent of GDP, while they only represent approximately 43 percent of GDP in South Asia. But, within East Asia there is significant variation. While the share of

Table 36.1. Banking sector statistics in developing countries of Asia

	Bank assets to gross domestic product (%) 2005	Private credit to gross domestic product (%) 2005	Overhead costs to assets (%) 2005	Net interest margins to assets (%) 2005	Return on assets (%) 2005	Branches (per 100,000 people) (%) 2003–4
China	122.80	112.18	1.08	2.27	0.57	1.33
Indonesia	34.24	21.81	2.86	4.75	1.91	8.44
Korea (Republic of)	93.47	89.15	1.58	3.00	1.25	13.40
Malaysia	109.73	102.96	1.42	2.23	0.82	9.80
Philippines	44.76	26.42	5.71	3.95	1.16	7.83
Thailand	80.35	73.18	1.76	2.96	1.3	7.18
East Asia and Pacific (average)	80.89	70.95	2.40	3.19	1.17	8.00
Bangladesh	35.67	27.65	2.12	2.71	0.60	4.47
India	56.54	36.80	2.09	2.86	0.98	6.30
Pakistan	36.79	26.75	2.19	3.56	1.84	4.73
South Asia (average)	43.00	30.40	2.13	3.04	1.14	5.17
Europe and Central Asia (average)	42.78	32.55	3.13	3.64	1.74	12.45
Latin America (average)	31.52	16.06	6.33	8.17	1.91	9.16
Middle East & North Africa (average)	60.80	37.62	2.16	3.16	0.61%	n/a
Sub-Saharan Africa (average)	46.68	39.57	5.78	6.89	1.79	2.99

Source: World Bank Financial Structure Database and Bankscope <http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:20696167~pagePK:64214825~piPK:64214943~theSitePK:469382,00.html>.

bank assets in China and Malaysia exceeds 100 percent of GDP, countries like Indonesia (34 percent) and Philippines (45 percent) have significantly smaller banking sectors than other countries in the region and, in fact, are more similar in size to the banking sectors of Bangladesh (36 percent), India (57 percent), and Pakistan (37 percent). Relative to other regions, banking sectors in East Asia are much larger than those of all other regions, while the size of South Asian banking sectors is smaller than that of the largest economies of Africa (47 percent) and the Middle East (61 percent), comparable to that of countries in Eastern Europe (43 percent), and exceeds those in Latin America (32 percent).[2]

A similar pattern to that described for banking sector size holds for a commonly used measure of banking sector depth, the share of bank credit to the private sector relative to GDP. At 71 percent of GDP, banking sector depth in East Asia is more than twice that for South Asia, where it averages 30 percent, and also far exceeds that for other regions like Eastern Europe, Latin America, Middle East, and North and South Africa. Again, Indonesia and Philippines are an exception, where the share of bank credit to the private sector—respectively at 22 percent and 26 percent of GDP—approaches banking sector depth measures for South Asia. In turn, banking sector depth measures for South Asia are smaller than those observed for the top economies in Africa (40 percent) and the Middle East (38 percent), are similar to those observed for Eastern Europe (33 percent) and exceed those in Latin America (16 percent).

When it comes to measures of efficiency, the averages for East and South Asia are similar to each other and are in general better than those observed in other regions. Overhead costs as a share of assets average approximately 2.4 percent in East Asia and 2.1 percent in South Asia, while they exceed 3 percent in Eastern Europe and average close to 6 percent in Latin America and Sub-Saharan Africa. Within East Asia, banks in Indonesia and the Philippines seem relatively less efficient than those in other countries. Overhead costs are 2.9 percent in Indonesia and 5.7 percent in the Philippines, while they stand at 1.1 percent for China, 1.6 percent for Korea and 1.4 percent for Malaysia. In South Asia, the share of overhead costs to bank assets is 2.1 percent for Bangladesh and India, and 2.2 percent for Pakistan.

The cost of financial intermediation is similar in East and South Asia and it is also lower to that observed for most other regions. Net interest margins average 3 percent of assets in both regions compared to 3.6 percent in Eastern Europe, 8.2 percent in Latin America, and 6.9 percent in Sub-Saharan Africa. Cross-country analyses of the determinants of net interest margins indicate that margins in Latin America and Sub-Saharan Africa tend to be higher primarily because of higher

[2] The following countries are included in the regional averages: in East Europe, we consider the Czech Republic, Hungary, Poland, and Russia; in Latin America, we include Argentina, Brazil, Mexico, and Venezuela; in the Middle East, we include Algeria, Egypt, and Morocco; and in Sub-Saharan Africa, we include Kenya, Nigeria, and South Africa.

overhead costs (see Gelos, 2006; and Honohan and Beck, 2007). Within East Asia, net interest margins are lowest in Malaysia (2.2 percent) and China (2.3 percent) and highest in Indonesia (4.8 percent) and the Philippines (3.9 percent). Net interest margins are 2.7 percent in Bangladesh, 2.9 percent in India, and 3.6 percent in Pakistan.

Not surprisingly, comparing East and South Asia to other regions like Eastern Europe, Latin America, and Sub-Saharan Africa, lower intermediation margins for East and South Asia translate into lower profitability ratios. Return on assets average 1.17 percent in East Asia and 1.14 percent in South Asia, while they stand at 1.74 percent for Eastern Europe, 1.91 percent for Latin America, and 1.79 percent for Sub-Saharan Africa. Within Asia, profit margins are highest for Indonesia (1.91 percent), Pakistan (1.84 percent), and Korea (1.25 percent) and lowest in China (0.57 percent) and Bangladesh (0.60 percent), two of the countries with the largest share of government-owned institutions.

Countries in East and South Asia seem to rank below most other regions (except for Sub-Saharan Africa) in terms of one measure of banking outreach: the number of branches per person.[3] Across East and South Asia, per capita branch penetration averages 7.9 and 5.2 branches per 100,000 people, respectively. In contrast, there are on average 12.5 branches per 100,000 people in Eastern Europe and 9.2 in Latin America.

Another area where East and South Asia seem to be lagging behind other regions is in the degree of private and foreign ownership in the banking sector. Because this is an area where important changes are taking place, the rest of the chapter is devoted to describing these changes and exploring the evidence on their implications.

Banking sector structure in Asia: Evolution and comparison with other regions

Over the last decade, developing countries in Asia have witnessed noticeable changes in their banking sectors' structure. In general, as shown in Figure 36.1,

[3] Indicators of financial outreach—the extent to which the population has access to financial services—are hard to come by. Beck, Demirgüç-Kunt, and Martinez Peria (2007) collected information on the number of branches, ATMs, loans, and deposits across a number of countries. However, for the purpose of comparison across regions and countries, information on branches is the most comprehensive.

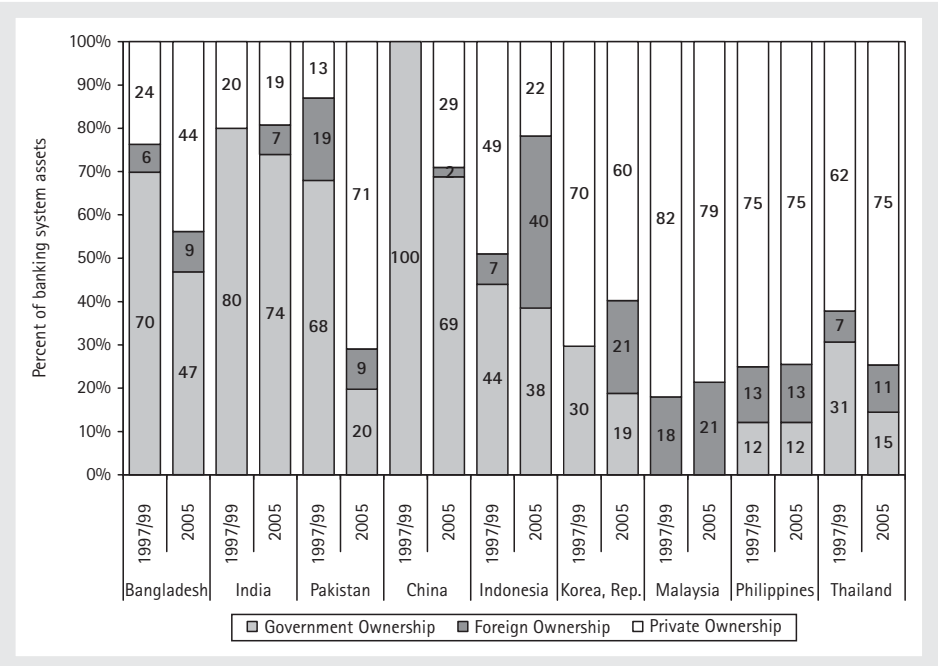

Fig. 36.1. Changes in bank ownership structure in East and South Asia

Source: Barth, Caprio, and Levine; Supervision and Regulation Database. <http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:20345037~pagePK:64214825~piPK:64214943~theSitePK:469382,00.html>.

the trend has been toward less government ownership of banks and increased foreign participation in the sector. However, there are significant differences across countries in the extent to which reforms have taken place. In South Asia, Pakistan has led the way in reducing government participation in the banking sector. Between 1997 and 2005, government ownership fell from 68 percent to 20 percent. On the other hand, though some progress was made by Bangladesh and less so by India, in both countries, but especially in India, government banks still play a dominant role. Over this period, Bangladesh reduced the share of assets held by government banks from 70 percent to 47 percent, while in India government participation declined only from 80 percent to 74 percent. In terms of foreign bank participation, progress in South Asia has been slow. While India had practically no foreign presence in 1997, today foreign banks account for close to 7 percent of the system. In Bangladesh, the share of bank assets held by foreign banks rose from 6 percent to 9 percent over the period 1997–2005.

In East Asia, government ownership has declined across the board but continues to be significant in Indonesia and especially in China. Though since 2001 the Chinese government has introduced important banking sector reforms (see

below), which have resulted in an increase in private ownership of banks, government banks still account for 69 percent of the sector. In Indonesia, the share of government-owned banks fell from 44 percent to 38 percent between 1997 and 2005. Korea and Thailand also witnessed a decline in government ownership, but the levels of government participation were much lower in these countries to begin with. In both economies, government banks controlled roughly 30 percent of banking sector assets in 1997. By 2005, the shares of government participation declined to 19 percent in Korea and 15 percent in Thailand. In the Philippines, government participation has remained roughly unchanged at 12 percent of banking sector assets.

With the exception of China, where foreign ownership has only recently been allowed and accounts for close to 2 percent of bank assets, foreign bank participation in East Asia is much higher than that observed in South Asia and the changes have been more significant. In Indonesia, for example, the share of assets held by foreign banks increased from 7 percent to almost 40 percent. Similarly, Korea also experienced a significant increase in foreign bank participation. The share of assets held by foreign banks in this country rose from close to zero to 21 percent. Malaysia, Philippines, and Thailand also underwent an increase in foreign bank presence, but the changes in these countries were less pronounced. In Malaysia, the share of assets held by foreign banks rose from 18 percent to 21 percent, while in the Philippines foreign bank participation rose marginally from 12.7 percent to 13.4 percent. Finally, Thailand experienced an increase in foreign participation from 7 percent to 11 percent.

Despite recent changes toward more private and foreign ownership in the banking sectors of South and East Asia, private and foreign bank presence in these regions is significantly lower than that in most other regions of the world (see Figure 36.2). While on average 33 percent of assets are held by government banks in Asia, government bank participation is on average less than 1 percent in Sub-Saharan Africa (Kenya, Nigeria, and South Africa), 15 percent in Eastern Europe (Czech Republic, Hungary, Poland, and Russia), and 21 percent in Latin America (Argentina, Brazil, Mexico, and Venezuela). Only countries in North Africa and the Middle East (Egypt and Morocco) exhibit higher levels of government ownership (48 percent of assets). In terms of foreign ownership, Asia ranks at the bottom of all other regions. On average, 15 percent of assets are held by foreign banks in Asia vis-à-vis 22 percent in North Africa and the Middle East, 24 percent in Sub-Saharan Africa, 40 percent in Latin America, and 65 percent in Eastern Europe.

While the Asian economies have undergone significant changes in banking sector ownership structure, there is less evidence of noticeable shifts in the degree of concentration (see Figure 36.3). With the exception of Korea and Malaysia, where the share of assets held by the top five banks rose from 48 percent to 69 percent in the first case, and from 30 percent to 56 percent in the

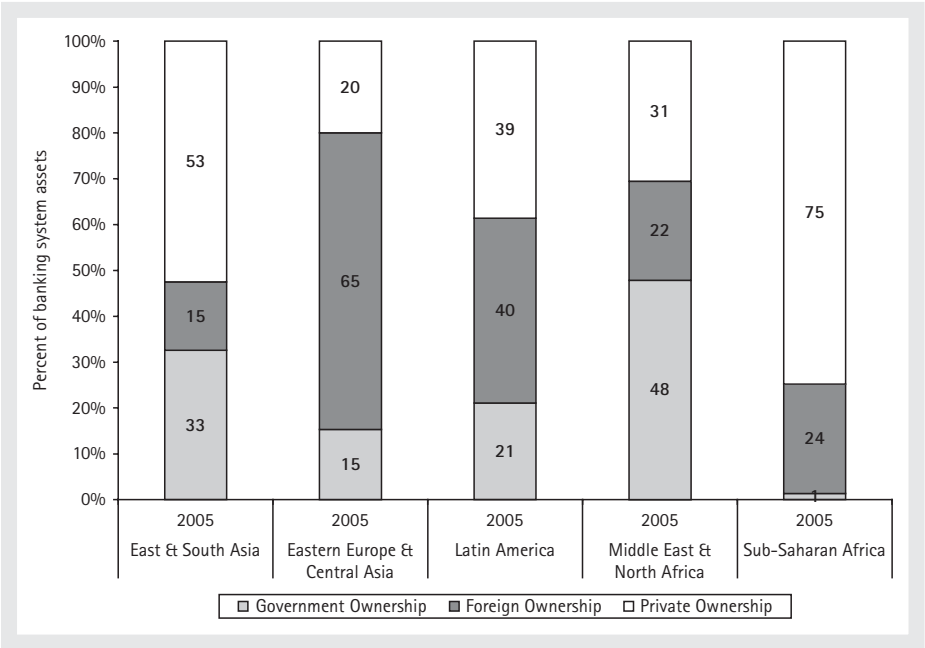

Fig. 36.2. Bank ownership structure across regions

Source: Barth, Caprio, and Levine; Supervision and Regulation Database: 64. <http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:20345037~pagePK 214825~piPK:64214943~theSitePK:469382,00.html>.

second, in the other countries concentration levels remained constant or declined slightly.

Explaining the recent changes in Asia's banking sector structure

While, as described above, most banking sectors in Asia have witnessed a decline in government participation and an increase in foreign bank presence, there are significant differences in the ways and speed in which these changes are being implemented. In addition, the development of the banking sector in the region is rooted in historical events, both political and economic. For instance, the legacy of bank nationalizations in India followed populist socialist movements, the state ownership of banks in China was established under the Communist regime, and the reorganization and domestic and foreign sales of banks in East Asia was in

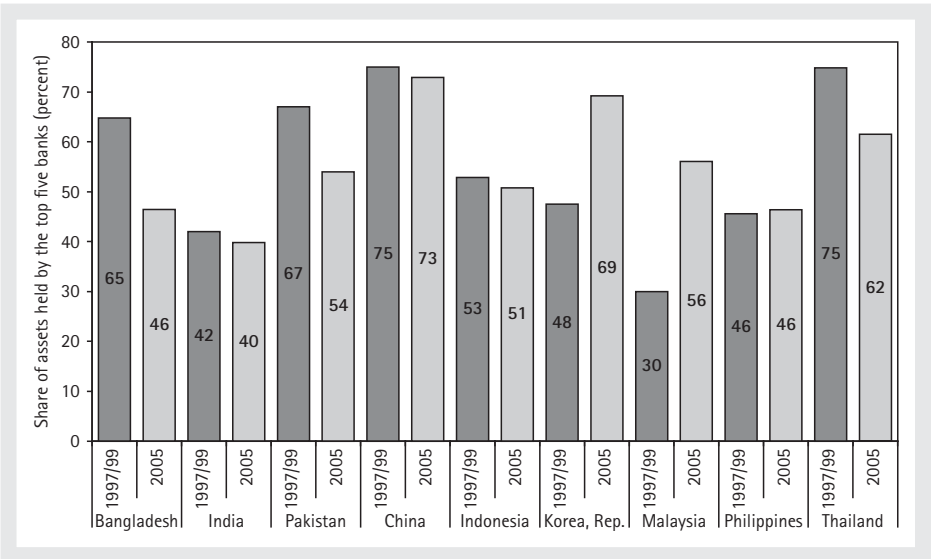

Fig. 36.3. Banking sector concentration in East and South Asia

Source: Barth, Caprio, and Levine Supervision and Regulation Database. <http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:20345037~pagePK:64214825~piPK:64214943~theSitePK:469382,00.html>.

response to the economic crisis. This section discusses the reform process in the main economies in East and South Asia.

China

China's entry into the World Trade Organization (WTO) in 2001 brought pressure to reform the largely government-owned and almost completely closed Chinese banking sector. During the 1990s, foreign banks were prohibited from conducting consumer banking in local currency with Chinese residents and, by the end of 1999, only twenty-five foreign banks had permission to conduct local currency business with Chinese enterprises. Following accession to the WTO, China allowed foreign banks to provide foreign currency services for Chinese residents and enterprises and in 2004 China opened its local currency market and allowed foreign banks to provide local currency services to Chinese enterprises in selected cities and areas. The retail market was supposed to be opened for foreign banks in December 2006; however, in late 2006, the Chinese government imposed the additional requirement that banks had to incorporate locally in order to cater to Chinese residents, delaying foreign banks' access to consumers. However, in April 2007, four foreign banks (Citigroup, HSBC, Standard Chartered, and Bank of East Asia) received approval from Chinese regulators and began accepting deposits in renminbi (RMB) from Chinese residents (Berger, Hasan, and Zhou, 2009).

Chinese regulators have also relaxed the rules concerning foreign acquisition of domestic banks. Since 2003, foreigners can collectively own up to 25 percent of any domestic bank but individual investors are limited to between 5 percent and 20 percent ownership, subject to regulatory approval. Following these changes, foreign strategic investors have acquired stakes in four of the top five banks in China, as well as in other smaller banks. In August 2004, the Hong Kong Shanghai Banking Corporation (HSBC) acquired a 19.9 percent stake of Bank of Communications. On 17 June 2005, Bank of America reached a deal to buy a 9 percent stake in China Construction Bank. Also in 2005, Royal Bank of Scotland and Temasek each acquired a 10 percent stake in Bank of China, while in 2006, Goldman Sachs, Allianz AG, and American Express agreed to buy a total 10 percent stake in the Industrial and Commercial Bank of China.

In an effort to improve the management and governance of Chinese banks, regulators have also recently allowed banks to list on stock exchanges. Between 2005 and 2006, Bank of Communications, Bank of China, China Construction Bank, and the Industrial and Commercial Bank of China had successful initial public offerings in the Hong Kong and Shanghai stock markets raising over $40 billion in total.

Rest of East Asia

The East Asian crisis began in Thailand in mid-1997, when an ailing financial sector, an export slowdown, and large increases in Central Bank credit to weak financial institutions triggered a run on the local banks. The crisis quickly spread to Indonesia and Korea, as common vulnerabilities and changes in international sentiment resulted in large capital outflows. East Asian governments were initially slow to address the growing distress in the banking sector (Kho and Stulz, 2000). At first, the governments tried to keep insolvent institutions afloat by injecting liquidity. This strategy, however, incurred large (and unpopular) fiscal costs. The governments' delayed and sometimes partial responses to the crisis caused financial turbulence and runs on financial institutions. The governments responded to the crisis in public confidence (in Indonesia and Thailand) and foreign currency outflows (in Korea) by issuing unlimited guarantees on their financial systems' liabilities. These guarantees stemmed the confidence crisis, but weakened governments' need to act comprehensively (Djankov, Jindra, and Klapper, 2005).

Responses and progress on financial restructuring varied considerably across the three crisis countries. Korea moved aggressively to strengthen its banking system through recapitalizations, nationalizations, removal of bad debt, and mergers. Although Korean officials closed over 100 non-bank financial institutions, no banks were shut down. Instead, eleven banks were merged with other domestic banks and four banks were nationalized (Delhaise, 1999).

Thailand adopted a market-based approach, allowing banks to raise capital over a longer time period. The Thai authorities closed down two-thirds of the finance companies but allowed banks a transitional period to raise capital through phased-in tighter loan provisioning requirements. At the same time, the government offered to inject Tier-1 capital, subject to the condition that any bank accepting public money would have to satisfy certain stringent conditions—for example, meeting strict loan-loss provisioning and making management changes. As a result, the Thai government was only required to shut down one bank. In addition, three banks were merged with other domestic banks and four banks were nationalized.

Of the three crisis countries, Indonesia has made the least progress in reforming its banking sector. By October 1999, sixty-four small banks were closed down, twelve banks were nationalized, and nine of the large banks were recapitalized. However, most financial institutions remained insolvent or undercapitalized. In response, the government guaranteed deposits of all Indonesian banks and nominated the Central Bank to act as a paying agent for depositors of the closed banks.

In addition, all three crisis countries reformed existing bank regulations to permit foreign banks to purchase domestic banks shortly following the crisis. The expected benefits of this measure were to infuse foreign capital and bring banking expertise. For example, Moody's stated:

> Foreign acquisition is the option of choice for the government because it will bring in badly needed foreign capital, thus reducing the burden on the public sector, while providing technical expertise and enhancing the quality of management. In Moody's view, improving management quality that is free from political influence is an especially critical aspect of the banking sector reform. *Banking System Outlook: Korea* (1998)

However, as discussed above and in previous chapters, foreign bank participation in these countries is still low, relative to other developing regions.

India

Following independence in 1945, the Reserve Bank of India (RBI) was formed as the central bank and high priority was given to increasing credit to rural areas and small businesses. In 1955, the government took over the largest bank, the Imperial Bank of India, to form the State Bank of India (SBI). The State Bank of India Act in 1959 directed SBI to take over regional banks that were associated with local governments and make them subsidiaries of SBI, which were later named 'associates'. SBI is now the largest commercial banking organization in the country—and one of the largest in the world. SBI and its seven regional associates have a substantial rural branching footprint—of about 14,000 branches of these banks, 74 percent are located in rural and semi-urban areas (*India Banking Yearbook*, 2003).

Given continued pressure to extend bank credit to the agricultural and small business sectors, the Indian government nationalized fourteen large banks in 1969 and another six in 1980 to redirect credit to 'underserved' sectors and populations. Unlike SBI, nationalized banks remained corporate entities and retained most of their management and staff. Although their boards of directors were replaced by the state, appointees included representatives from both the government and private industry (Banerjee, Cole, and Dufflo, 2005). RBI continued to fix interest rates on loans, and a significant portion of nationalized banks' deposit bases were redirected to support government expenditures through statutory measures that required banks to maintain specified fractions of their total deposits as cash balances with RBI and additional fractions in government and quasi-government securities. These lending restrictions and other RBI regulations affected the extension of credit by Indian banks (Bhaumik and Piesse, 2005). However, the nationalized banks continue to the present day to maintain relationships with large firms that begun prior to nationalization.

Banking sector liberalization and deregulation in India started in the early 1990s as part of a comprehensive reform agenda. This included permission to establish 'de novo' banks and the entry of foreign banks, the deregulation of branch expansion, and the privatization of some state-owned banks. Interest rates were also liberalized and banks were permitted to invest in equity. However, commercial banks are still required to make loans to 'priority sectors' at below market rates. These sectors consist largely of agriculture, exporters, and small businesses.[4]

Most foreign banks began operating in the 1990s under a license to open branches and are permitted to take deposits and provide credit in accordance with local banking laws and RBI regulations.[5] As of 2005, 33 foreign banks operated in India, but accounted for less than 0.5% of bank branches, five percent of deposits, and less than seven percent of assets (Federal Reserve Bank of San Francisco, 2005; and *Hindu Times*, 7 March 2007, respectively). Foreign banks have generally not purchased shares of local Indian banks, since, prior to 2006, foreign banks were restricted to a ceiling of 10 percent of voting rights, even though foreign banks could legally own up to 74 percent of equity. Foreign banks have typically focused their operations in the top 25 cities in the country, probably due in part to restrictions on branch expansion;[6] they

[4] Reforms in 2007 permit banks to hold securitized portfolios of priority loans, but this still effectively directs credit away from private commercial lending.

[5] A few foreign banks, such as Standard Chartered, have had limited operations in India for decades.

[6] Foreign banks currently operate only on a branch-license basis under which they are required to keep locally US $25 million in capital for the first three branches. Further expansion does not require additional capital, but requires RBI approval, which is often difficult to receive.

also generally use more modern equipment, pay higher salaries, and attract better-trained employees (IndiaMart, 2007).

In 2005, the government announced reforms to foreign banking laws that will be gradually implemented between 2005 and 2009 and will allow foreign banks to establish or convert existing operations into wholly owned subsidiaries. In addition, the RBI raised the limit of foreign direct investment in private banks to 74 percent from 49 percent and announced 'roadmap' plans to amend the Banking Regulation Act to allow for the voting rights of foreign banks to reflect their ownership level, eliminating the current 10 percent cap. In June 2009, RBI announced that permission for foreign banks to acquire any private bank will be delayed because of the global financial crisis. At that point, wholly owned subsidiaries will be allowed to list or dilute their stake to 74 percent through an IPO or offer for sale. Regardless of the decline in Indian markets and bank stock prices in 2008, there remains popular support for reform. For example, the Raghuram Rajan 2008 Committee on Financial Sector Reform strongly endorses the abolition of branch licensing and more liberal permission for bank takeovers and mergers.

Private banks are primarily 'de novo' entrants that were granted banking licenses during the financial liberalization in the early 1990s. A total of twenty-five 'de novo' private banks began operations between 1994 and 2000. There are also a small number of incumbent private banks that existed before 1990 and some state-owned institutions that have been successfully privatized. An example of the latter is ICICI, which was formed in 1955 as a state-owned institution at the initiative of the Government of India and the World Bank to create a development financial institution for providing medium- and long-term project financing to Indian businesses. During the 1990s, ICICI was privatized and evolved into a private, full-service bank and is now India's second largest bank, offering a wide range of services to retail and corporate customers.

To summarize, although there are numerous foreign banks in India, they have relatively few branches and accounts and also have fewer deposits and assets than the other types. Nationalized banks are the largest type, as measured by number of branches, accounts, deposits, and assets. The state-owned banks combined—SBI plus the nationalized banks—dominate the banking sector with about 80 percent of deposits and assets.

Pakistan

In 1974, the banking industry in Pakistan—which was predominately owned by the private sector—was nationalized following political pressure to break up the large

and powerful ownership holdings (in both the industrial and banking sectors) held by a few business groups. For more than twenty-seven years the banking sector remained closed to the private sector. In addition, the foreign exchange market was highly regulated through a system of direct exchange control over suppliers and users of foreign exchange.

In late 1996, the banking system was on the verge of a crisis. A World Bank report summarized the problems that plague many state-dominated banking systems:

> In late 1996, Pakistan's banking system was on the verge of a crisis . . . Political interference vitiated the financial intermediation function of the banking system and borrowers expected not to repay loans they took, especially from the state-owned banks. Overstaffing and over-branching and undue interference by labor unions in bank personnel and operations resulted in large operating losses. Poor disclosure standards abetted corruption by window-dressing the true picture of banks. (World Bank, 2000)

In 2000, the government of Pakistan began a process of banking reform. The objective of the reform package was to increase competition among financial institutions and markets, revise banking laws to strengthen bank governance and supervision, and adopt a market-based indirect system of monetary, exchange, and credit management for better allocation of financial resources (Husain, 2003). This sector, which had been fully dominated by nationalized commercial banks, was opened up to the private sector: fourteen new domestic private commercial banks and sixteen private investment banks were established, and nineteen foreign commercial banks began operating in the country. In addition, four out of the five nationalized commercial banks were privatized. Following the privatization of Habib Bank—the largest Pakistani bank—in 2004, the market share of public sector banks was reduced to 20 percent, with the remaining 80 percent of banking assets in the hands of private banks. This is a remarkable accomplishment for a country which only a decade ago had almost 90 percent of banking assets under government ownership and control and, as previously discussed, few countries in the region have accomplished such a feat.

Evidence on the implications of recent banking sector structure changes in Asia

While changes in the ownership structure of the Asian banking sectors are recent and are still under way, it is interesting to survey the evidence available so far. For the most part, studies have tested for differences in performance among bank

ownership types and evaluated the implications of recent ownership changes for bank efficiency and bank credit availability.

Cross-country studies on the performance of banks in Asia have generally found a positive association between foreign bank ownership and bank performance, though some empirical issues remain. Using data on commercial banks in Hong Kong, Indonesia, Korea, Malaysia, the Philippines, Singapore, and Thailand between 1994 and 2004, Laeven (2005a) finds that foreign-owned banks exhibited significantly stronger performance than state-owned banks in terms of their operating income relative to total assets. In fact, Laeven finds that even after controlling for a variety of country-level policies and bank-level diversification measures, there exists a roughly one-to-one correspondence between increases in foreign ownership and increases in bank performance. A caveat to Laeven's measure of bank performance is that it assumes that all banks are profit maximizers, and that higher operating income relative to assets means better performance. State-owned banks, however, may have certain social objectives that are not tied to operating income, such as lending to certain developing sectors. Nonetheless, Laeven does find that, based on the same measure of performance, privately owned banks did not outperform state-owned ones in terms of their operating income relative to total assets. Moreover, this work builds on additional evidence provided by Laeven (2005b) that foreign-owned banks not only enjoyed stronger profit efficiency in East Asia, but they did so while maintaining portfolios with less risk. Using data-envelopment analysis for a sample of commercial banks in Indonesia, Korea, Malaysia, the Philippines, and Thailand between 1992 and 1996, Laeven (1999) shows that those banks whose portfolios reflected larger risks—primarily domestic banks—were also more likely to be restructured at the time of the Asian financial crisis.

Focusing on profit efficiency as their measure of performance, Williams and Nguyen (2005) study a similar cross-country set of banks between 1990 and 2003, finding that foreign banks did enjoy stronger overall performance, but with significant lags before these performance gains were realized. In addition, the authors temper their results by pointing out that banks selected for foreign acquisition had the greatest profit efficiencies prior to acquisition, hence highlighting a selection bias. Bank privatization, on the other hand, did yield some efficiency gains that were realized much sooner than was the case with foreign acquisitions. While there is similar evidence that governments did privatize their best-performing banks, these banks increased their profit efficiency in the period after their privatization significantly more than non-privatized banks.

While cross-country studies have shown a somewhat consistent relationship between bank ownership and bank performance, the evidence from within-country studies in Asia is very mixed. This is especially the case when measuring the impact of banking sector deregulation, where even studies within the same country have shown contradictory results.

Studies of the Chinese banking sector have mostly focused on comparing the performance of the 'Big Four' banks—the Agricultural Bank of China, the Bank of China, the China Construction Bank, and the Industrial and Commercial Bank of China—relative to majority state-owned, majority foreign-owned, and other joint-equity banks. Studies focusing on the cost efficiency of banks (Chen, Skully, and Brown, 2005; Kumbhakar and Wang, 2005; and Fu and Heffernan, 2007) have yielded mixed evidence: Chen, Skully, and Brown (2005) find that the Big Four banks outperformed medium-sized joint-equity banks in terms of cost efficiency, while others (Kumbhakar and Wang 2005; and Fu and Heffernan 2007) find that the Big Four are less cost efficient than the joint-equity banks over a similar time-period. These cost-efficiency studies also identify contradictory results on the effects of in financial deregulation the mid-1990s. Chen, Skully, and Brown (2005) find that deregulation had positive efficiency implications, while Kumbhakar and Wang (2005) find no significant improvements in efficiency due to the deregulation.

Moving away from cost-efficiency measures, papers such as Berger, et al. (2007) and Yao and Jiang (2007) focus instead on profit efficiency, arguing that this measure captures both cost and revenue performance. Analyzing a panel of thirty-eight Chinese banks over the 1994–2003 period, Berger, Hasan, and Zhou (2009) find significant differences in profit efficiency due to the identity of banks' majority owners. Banks with majority foreign ownership were the most efficient, followed in order by private domestically owned banks, non-Big Four majority state-owned banks, and the Big Four. The identity of minority owners yields even more dramatic effects on profit efficiency—for example, the presence of a minority foreign stake in non-Big Four state-owned banks is associated with an efficiency level almost 20 percentage points higher than enjoyed by other non-Big Four state-owned banks. To address selection issues, Berger, et al. (2007) use a difference-in-difference approach to examine the change in efficiency after minority foreign investments are made, finding significant efficiency improvements for both private domestic banks and non-Big Four state-owned banks. Specifically, they compare the four-year pre- and post-average efficiency change for banks with and without foreign ownership. Hence, they are able to control for any industry-wide efficiency improvements that affected all banks during their sample period. Nonetheless, as in other studies, the possibility of selection due to foreign investors' expectations of future efficiency improvements by these banks remains a potential source of upward bias.

Yao and Jiang (2007) offer complementary findings in studying Chinese commercial banks between 1995 and 2005, noting that state-owned banks continue to lag behind in terms of efficiency, despite overall improvements in banking efficiency. Moreover, Yao and Jiang find that foreign acquisition is associated with long-term efficiency improvements, but find no such evidence for banks that undergo an IPO.

In the Philippines, Unite and Sullivan (2003) find that while the relaxation of entry regulations for foreign banks generated cost-efficiency gains (reductions in operating expenses), the relaxation of limits on foreign ownership stakes did not produce significant benefits. Studying Thailand's banking system between 1990 and 1997, Williams and Intarachote (2002) find that, following liberalization, entry by foreign banks did increase, but, over the entire period, the efficiency of the entire banking system actually regressed. This is particularly striking because Williams and Intarachote examine alternative profit efficiency[7] and not only the cost side. While it appears that Japanese-owned banks fared somewhat better in terms of their profit efficiency, there is little evidence that foreign-owned banks more generally enjoyed any efficiency advantages over domestic banks. Looking at foreign bank efficiency after the Thai reforms of foreign entry regulations in 1997, Rajan and Montreevat (2001) do offer some casual albeit unsystematic evidence that greater entry by foreign banks did involve the introduction of new technologies and lower operating costs. However, none of these within-country studies can entirely distinguish whether the contemporaneous improvements in efficiency among domestic banks were related to increased competition from foreign banks or other facets of the reform process.

In India, Sanyal and Shankar (2005) offer evidence that the expansion of private ownership in India's banking sector has not been causally connected to greater bank efficiency. They find that following the mid-1990s' deregulations, state-owned banks gained as much as private banks in most measures of productivity and performance. In contrast, other research on the impact of deregulation in India has yielded some evidence of a positive relationship between foreign or private ownership and bank performance. Sahoo, Sengupta, and Mandal (2007) examine the productivity performance trends of Indian commercial banks between 1997 and 2005 using data envelopment analysis, finding that private banks enjoy greater cost efficiency than nationalized banks. In addition, the authors find that foreign banks outperform nationalized banks in both cost-based and price-based measures of performance. Similar results are offered by Reddy (2005), who finds that foreign and newly formed private banks enjoyed advantages in overall technical efficiency between 1996 and 2002. Ghosh, Harding, and Phani (2006), meanwhile, look at changes in stock market prices of shares of both private and public sector banks around the Indian Government's removal of limits on foreign ownership in commercial banks. Since they condition these stock market prices on the actual assets and income of the banks, Ghosh, Harding, and Phani can assess the extent to which valuation gains were due to expectations over each bank's improvement under foreign ownership (i.e., a 'takeover premium'). Their results show that the

[7] Alternative profit efficiency captures the profit rather than simply the cost-based efficiency story, but it can be estimated even when certain conditions required for standard profit efficiency estimation are not met.

benefits of foreign takeover were indeed concentrated among inefficient, poorly managed banks with lower existing market valuations. Ataullah and Le (2004), meanwhile, also use data envelopment analysis in studying commercial banks in both India and Pakistan between 1988 and 1998, finding that India's financial liberalization led to efficiency improvements for both public sector and private sector banks. In Pakistan, however, public sector banks did not improve their pure technical efficiency following liberalization.

More specific studies of the relationship between banks and firms in Asia have produced intriguing evidence that foreign-owned and private domestic banks operate quite differently from state-owned banks. For example, Berger, et al. (2008) study the role of bank ownership type in shaping relationships among firms and banks in India, finding that firms which maintain relationships with foreign banks differ from those with ties to state-owned banks in a number of ways: they are more likely to maintain multiple relationships, interact with a larger number of banks, and diversify their relationships across bank ownership types. Further, papers such as Mian (2006) in Pakistan and Gormley (2007) in India show that foreign banks can be limited in how far they can lend to local firms. Mian (2006) shows evidence that foreign banks have a tougher time lending to informationally opaque firms and are less likely to lend based on 'soft' information. This characteristic, in turn, makes it difficult for foreign banks to renegotiate and recover bad loans, a process that requires strong soft information skills.

In a similar vein, Gormley (2007) finds evidence that foreign bank entry in India yielded a reallocation of loans toward the most profitable firms. He finds that following the deregulation of foreign entry in 1994, the entry of a foreign bank in a particular location was associated with a significant reduction in the likelihood that the average firm in that location obtained a long-term loan. This was particularly true among firms associated with business groups for which moral hazard related to tunneling and informational asymmetries may be most pronounced, and thus those for which foreign banks may be at a particular disadvantage in serving. Based on this evidence, Gormley points out that the growth in foreign bank ownership in India may generate further market segmentation and may not be sufficient for—and in fact may counteract—the expansion of access to credit.

Apart from studying the general and specific effects of bank ownership type on performance, a number of studies in Asia have examined the differential response of foreign banks to the Asian financial crisis. The bulk of the evidence suggests that these banks increased the stability of the banking system rather than weakened it. Looking at twenty-six domestic banks and fifty-nine foreign banks operating in Korea between 1994 and 1999, Jeon and Miller (2005) find that while the returns on assets and equity of Korean banks deteriorated dramatically in 1998, foreign-owned banks experienced no such drop-off (with a statistically significant difference in their performance). Extending this sample through 2005, Jeon, Miller, and Yi (2007) use a generalized Bennet dynamic decomposition to assess the extent to

which aggregate bank performance was due to adjustments within banks, reallocations between banks, and entry or exit of banks. The authors confirm that foreign banks enjoyed a higher return on equity both before and during the crisis, although domestic Korean banks experienced an advantage after the crisis. Nonetheless, there is not much evidence that restructuring among either foreign or domestic banks played a significant role in driving aggregate performance, which instead mostly reflected within-bank changes. In a similar vein, Detragiache and Gupta (2004) study thirty-nine Malaysian banking institutions between 1996 and 2000 and find no evidence that foreign banks scaled down their operations faster than domestic banks during the crisis, contradictory to concerns that foreign banks are likely to flee volatile emerging markets during crises. The authors do find that while the foreign-owned banks mainly active in Asia performed similarly to domestic banks during the crisis, non-Asia-oriented foreign banks enjoyed profits that were significantly higher than domestic banks. On the other hand, evidence on the entry decisions of foreign banks suggests that Asian banks do in fact enjoy an advantage over non-Asian banks when operating in or entering other countries in the region. For example, Leung, Digby, and Young (2003) use survival analysis to study the entry decisions of banks into China between 1985 and 1996, finding that Asian banks were much more likely to establish a branch in China than non-Asian banks (and to do so earlier).

In conclusion, research on the effects of bank ownership in Asia has produced mixed results. Foreign and private ownership in general is associated with stronger bank performance, particularly in the context of cross-country samples. However, the effects on financial access are less clear, with many studies in the region identifying limited positive effects, and some even identifying significant negative effects on access. It is worth noting that although much of the research on the effects of foreign and private ownership in banking has focused on efficiency gains, the contribution of such ownership types can often happen by changing the nature of services being offered, not only by providing existing services more efficiently. He and Fan (2004), for example, point out that foreign entry into China's banking sector has introduced the use of credit card services, financial derivatives, and cross-selling of investment and insurance products, among other features. However, further study of the effects of these new product lines for Asian banks remains warranted.

Conclusions so far and looking ahead

This chapter described the recent changes in bank ownership structure in Asia, offered a detailed account of the reform process in the main economies in the

region, and evaluated the evidence available so far on the implications of these changes.

Overall, important changes in banking sector structure have recently taken place in Asia. Across the board, we have witnessed a trend toward less government ownership of banks and greater foreign bank presence. However, the largest countries in the region such as China, India, and Indonesia still have a way to go in terms of fostering private and foreign bank ownership and offering a level playing field for all types of banking institutions.

The existing literature analyzing the implications of recent ownership changes in Asia is mixed when it comes to performance improvements and quite negative in terms of the effects of foreign ownership on access to finance. These findings stand in contrast to the evidence from regions like Eastern Europe and Latin America, where many studies have documented the benefits from private and foreign ownership, especially in terms of performance and efficiency improvements (see Chapters 33 and 34; and Cull and Martinez-Peria, 2007). What explains these differences across regions? This is an important question beyond the scope of this chapter. However, we offer two potential explanations. First, while reforms promoting foreign and private ownership were implemented in the early to mid-1990s in Latin America and Eastern Europe, these changes are very new to Asia. It may take time for economies in Asia to see the benefits that the other two regions have already experienced. Also, in practical terms, more data might be needed for empirical studies to detect whatever benefits these changes might bring. Second, the extent of reforms (i.e., how far they have gone in producing ownership changes) has been quite different in Asia relative to Latin America and Eastern Europe. While these regions drastically altered the structure of their banking sectors, promoting majority private and foreign ownership, Asia so far has taken a more piecemeal approach. We hope that future research will establish the merit of these proposed explanations and shed light on the question of why foreign and private ownership seemed to have had different effects in Asia relative to other regions.

What will the future bring for Asian banking sectors? In the midst of the deepest financial crisis to hit the global economy in decades, it is hard to predict what will happen to the banking sectors in Asia. Most analysts agree that the direct exposure of Asian banks to the US subprime mortgage market is limited.[8] Furthermore, the consensus is that the region is far better placed to withstand the present shock than it was during the 1997–8 crisis. Banks are better capitalized and have lower levels of non-performing loans; currency regimes are more

[8] Of the estimated US $300 billion total write-downs and credit losses worldwide, write-downs in Asia are expected to be less than 6% (estimates by Crisil consulting firm, an S& P subsidiary, reported by United News of India, 9 May 2008).

flexible; and foreign exchange reserves are bigger (Asian Development Bank, 2008).[9]

On the other hand, after a decade of reform and accelerated globalization, the region is more closely integrated to the world economy. Hence, most analysts worry about the second-round effects of the financial crisis. As the US, Europe, and Japan enter into recessions, growth in developing countries in Asia is likely to slow down (in particular, because these economies are heavily dependent on exports to developed countries) and this will have implications for the health of banks' balance sheets via a rise in loan delinquencies.[10] Indian exports are already suffering, with the country's prized textile export sector firms having to downsize operations and shut down, at least temporarily, some of their mills.

In addition, it is unclear if the process of increasing foreign bank participation and bank privatization will continue, at least in the short run. As banks in the US and Europe face difficult situations at home, it is unlikely that they will continue to be able to grow or be interested in expanding their overseas operations. In fact, recently announced job cuts by banks such as HSBC and Citigroup suggest that overseas operations might even contract. At the same time, it might be harder for governments to undertake bank privatizations in an environment of diminished profit prospects in the banking sector and declining stock market prices.

Furthermore, an important second-round effect of the financial crisis will probably be a substantial curtailment of lending operations by banks in Asia. The global financial squeeze will probably make banks more selective in choosing loan clients, which will have consequences for local corporate and micro-enterprise growth. Banks in India, for instance, have already shut down their subprime loan market and have significantly reduced their secured and unsecured lending. Only time will tell whether the recent changes in banking sector ownership structure will continue in Asia.

References

Asian Development Bank (2008). *Asian Development Outlook Update*. Manila, Philippines.

Ataullah, A., Cockerill, T., and Le, H. (2004). 'Financial Liberalization and Bank Efficiency: A Comparative Analysis of India and Pakistan', *Applied Economics*, 36: 1915–24.

[9] This is not to say that Asia has not suffered some of the repercussions of the global credit crunch. Interbank lending rates have risen, stock markets have experienced severe drops in value, foreign exchange reserves have fallen, and Asian central banks have been required to pump liquidity into their financial systems. Pakistan has been among the hardest hit countries and has been forced to enter into a Stand-By Arrangement with the IMF for US $7.6 billion.

[10] G3 economies represent 60% of final demand for Asian exports. The IMF World Economic Outlook Update (2008) predicts that growth in developing Asia will drop from 10% in 2007 to 8.3% in 2008, and 7.1% in 2009. The Asian Development Bank predicts an even slower growth rate for 2008 and 2009: 7.5% and 7.2%, respectively (ADB, Asian Development Outlook Update, 2008).

Banerjee, A., Cole, S., and Duflo, E. (2005). 'Banking Reform in India', National Council for Applied Economic Research (NCAER), New Delhi and Brookings Institution, Washington, DC India Policy Forum 1.

Bhaumik, S. K. and Piesse, J. (2005). 'The Risk Aversion of Banks in Emerging Credit Markets: Evidence from India', William Davidson Institute Working Paper 774.

Beck, T., Demirguc-Kunt, A., and Martinez-Peria, M. S. (2007). 'Reaching Out: Access to and Use of Banking Services across Countries', *Journal of Financial Economics,* 85: 234–66.

—— and Honohan, P. (2007). *Making Finance Work for Africa.* Washington DC: World Bank.

Berger, A. N., Clarke, G., Cull, R., and Klapper, L. (2005). 'Corporate Governance and Bank Performance: A Joint Analysis of the Static, Selection, and Dynamic Effects of Domestic, Foreign and State Ownership', *Journal of Banking & Finance*, 29: 2179–221.

—— Hasan, I., and Zhou, M. (2009). 'Bank Ownership and Efficiency in China. What will Happen in the World's Largest Nation?', *Journal of Banking & Finance*, 33: 113–30.

—— Klapper, L., Martinez-Peria, M. S., and Zaidi, R. (2008). 'Bank Ownership Type and Banking Relationship', *Journal of Financial Intermediation*, 17: 37–6.

Chen, X., Skully, M., and Brown, M. (2005). 'Banking Efficiency in China: Application of DEA to Pre- and Post-Deregulations Era: 1993–2000', *China Economic Review,* 16: 229–45.

Cull, R. and Martinez-Peria, M. S. (2007). 'Crises as Catalysts for Foreign Bank Activity in Emerging Markets', in Justin Robertson (ed.), *Power and Politics After Financial Crises: Rethinking Foreign Opportunism in Emerging Markets.* New York: Palgrave Macmillan.

Delhaise, P. (1999). *Asia in Crisis: The Implosion of the Banking and Finance Systems* (first ed.). New Jersey: John Wiley & Sons.

Detragiache, E. and Gupta, P. (2004). 'Foreign Banks in Emerging Market Crises: Evidence from Malaysia', IMF Working Paper 04/129.

Djankov, S., Jindra, J., Klapper, L. (2005). 'Corporate Valuation and the Resolution of Bank Insolvency in East Asia', *Journal of Banking & Finance*, 29: 2095–118.

Federal Reserve Bank of San Francisco (2005). 'Indian Banking Reforms: A Changing Landscape for State-Owned and Foreign Banks', *Asia Focus*, May.

Fu, X. and Heffernan, M. S. (2006). 'Cost X-Efficiency in China's Banking Sector, *China Economic Review,* 18: 35–53.

Gelos, G. (2006). 'Bank Spreads in Latin America', IMF Working Paper 06/44.

Ghosh, A., Harding, J., and Phani, B. V. (2006). 'The Effect of Liberalization of Foreign Direct Investment (FDI) Limits on Domestic Stocks: Evidence from the Indian Banking Sector', mimeo.

Gormley, T. (2007). 'Banking Competition in Developing Countries: Does Foreign Bank Entry Improve Credit Access?', Working Paper, University of Washington, St. Louis.

He, L. and Fan, X. (2004). 'Foreign Banks in Post-WTO China: An Intermediate Assessment', *China and the World Economy,* 12: 3–16.

Hindu Times (2007). 'Banking Reform: A Balancing Act', March 7.

Husain, I. (2003). 'Reforms of Public Sector Banks: Case Study of Pakistan', prepared for World Bank Conference on Transforming Public Sector Banks, Washington, DC, April 9, 2003.

Indian Banks' Association (2003). *Indian Banking Yearbook 2003*, Mumbai: Indian Banks' Association.

IndiaMart (2007). 'India Finance and Investment Guide', http://finance.indiamart.com/investment_in_india/index.html.

International Monetary Fund (2008). *World Economic Outlook Update*. Washington, DC.

Jeon, Y. and Miller, S. (2005). 'Performance of Domestic and Foreign Banks: The Case of Korea and the Asian Financial Crisis', *Global Economic Review*, 34: 145–65.

Jeon, Y., Miller, S., and Yi, I. (2007). 'Performance Comparisons and the Role of Restructuring for Foreign and Domestic Banks', mimeo.

Kho, B. C. and Stulz, R. (2000). 'Banks, the IMF, and the Asian Crisis', *Pacific-Basin Finance Journal*, 8: 177–216.

Kumbhakar, S. C. and Wang, D. (2005). 'Economic Reforms, Efficiency, and Productivity in Chinese banking', State University of New York, Binghamton, Working Paper.

Laeven, L. (1999). 'Risk and Efficiency in East Asian Banks', World Bank Policy Research Working Paper 2255.

—— (2005a). 'Banking Sector Performance in East Asian Countries: The Effects of Competition, Diversification, and Ownership', mimeo.

Leung, M., Digby, D., and Young, T. (2003). 'Entry of Foreign Banks in the People's Republic of China: A Survival analysis', *Applied Economics*, 35: 21–31.

Mian, A. (2006). 'Distance Constraints: The Limit of Foreign Lending in Poor Economies', *Journal of Finance*, 61: 1465–505.

Moody's Investors Service (1998). 'Global Credit Research', October.

Podpiera, R. (2006). 'Progress in China's Banking Sector Reform: Has Bank Behavior Changed?', IMF Working Paper 06/71.

Rajan, R. and Montreevat, S. (2001). 'Financial Crisis, Bank Restructuring, and Foreign Bank Entry: An Analytic Case Study of Thailand', Adelaide University Centre for International Economic Studies Working Paper 131.

Reddy, A. (2005). 'Technical Efficiency and Its Decomposition in Indian Banks in Post Liberalization', Indian Institute of Capital Markets Paper.

Sahoo, B., Sengupta, J., and Mandal, A. (2007). 'Productive Performance Evaluation of the Banking Sector in India Using Data Envelopment Analysis', *International Journal of Operations Research*, http.//papers.ssrn.com/sd3/papers.cfm?abstract_id=956812.

Sanyal, P. and Shankar, R. (2005). 'Financial Sector Reforms and Bank Efficiency in Developing Countries: Lessons from India', Brandeis University Working Paper.

Unite, A. and Sullivan, M. (2003). 'The Effect of Foreign Entry and Ownership Structure on the Philippine Domestic Banking Market', *Journal of Banking & Finance* 27: 2323–45.

Williams, J. and Intarachote, T. (2002). 'Financial Liberalization and Profit Efficiency in the Thai Banking System, 1990–1997: The Case of Domestic and Foreign Banks', mimeo.

—— and Nguyen, N. (2005). 'Financial Liberalization, Crisis, and Restructuring: A Comparative Study of Bank Performance and Bank Governance in South East Asia', *Journal of Banking & Finance*, 29: 2119–54.

World Bank (2000). 'Financial Sector Update', Washington, DC, May 31, 2000.

Yao, S. and Jiang, C. (2007). 'The Effects of Governance Changes on Bank Efficiency in China: A Stochastic Distance Function Approach', University of Nottingham Research Paper 2007/19.

Index