The EY Global Financial Services Institute brings together world-renowned thought leaders and practitioners from top-tier academic institutions, global financial services firms, public policy organizations and regulators to develop solutions to the most pertinent issues facing the financial services industry.

The Journal of Financial Perspectives aims to become the medium of choice for senior financial services executives from banking and capital markets, wealth and asset management and insurance, as well as academics and policymakers who wish to keep abreast of the latest ideas from some of the world’s foremost thought leaders in financial services. To achieve this objective, a board comprising leading academic scholars and respected financial executives has been established to solicit articles that not only make genuine contributions to the most important topics, but are also practical in their focus. The Journal will be published three times a year.

gfsi.ey.com

The articles, information and reports (the articles) contained within The Journal are generic and represent the views and opinions of their authors. The articles produced by authors external to EY do not necessarily represent the views or opinions of EYGM Limited nor any other member of the global EY organization. The articles produced by EY contain general commentary and do not contain tailored specific advice and should not be regarded as comprehensive or sufficient for making decisions, nor should be used in place of professional advice. Accordingly, neither EYGM Limited nor any other member of the global EY organization accepts responsibility for loss arising from any action taken or not taken by those receiving The Journal.
Editorial

Editor
Shahin Shojai
EY UAE

Advisory Editors
Dai Bedford
EY U.K.
Shaun Crawford
EY U.K.
David Gittleson
EY U.K.

Michael Inserra
EY U.S.

Antony M. Santomero
The Wharton School
Nick Silitch
Prudential Financial

Special Advisory Editors
Ben Golub
BlackRock
Anthony Neoh
Bank of China
Steve Perry
Visa Europe

Takao Kobayashi
Aoyama Gakuin University
Deborah J. Lucas
Massachusetts Institute of Technology
Massimo Massa
INSEAD
Tim Morris
University of Oxford
Patrice Poncet
ESSEC Business School
Michael R. Powers
Tsinghua University
Philip Rawlings
Queen Mary, University of London
Roberta Romano
Yale Law School
Hato Schmeiser
University of St. Gallen
Peter Swan
University of New South Wales
Marno Verbeek
Erasmus University
Bernard Yeung
National University of Singapore

Editorial Board
Viral V. Acharya
New York University
John Armour
University of Oxford
Philip Booth
Cass Business School and IEA
José Manuel Campa
IESE Business School
Kalok Chan
Hong Kong University of Science and Technology
J. David Cummins
Temple University
Allen Ferrell
Harvard Law School
Thierry Foucault
HEC Paris
Roland Füss
University of St. Gallen
Giampaolo Gabbi
SDA Bocconi
Boris Groysberg
Harvard Business School
Scott E. Harrington
The Wharton School
Jun-Koo Kang
Nanyang Business School
The Journal of Financial Perspectives

Part 1: Strategic

13 The science of winning in financial services – competing on analytics: opportunities to unlock the power of data
Hyong Kim, FSO Global and Americas Data and Analytics Leader, EY U.S.
Errol Gardner, EMEIA FSO Data and Analytics Leader, EY U.K.

25 Cross-border bank resolution
José Viñals, Financial Counsellor and Director of the Monetary and Capital Markets Department, International Monetary Fund
Sean Hağan, General Counsel and Director of the Legal Department, International Monetary Fund

31 Shadow banking: policy challenges for central banks
Thorvald Grung Moe, Special Adviser, Norges Bank

43 The consequences of exit from non-conventional monetary policy
Philip Turner, Deputy Head, Monetary and Economic Department, Bank for International Settlements

61 The audit mandatory rotation rule: the state of the art
Mara Cameran, Researcher, Accounting Department, Università Luigi Bocconi
Giulia Negri, Research Fellow, Accounting, Control, Corporate and Real Estate Finance Department, SDA Bocconi School of Management
Angela Pettinicchio, Assistant Professor, Accounting Department, Università Luigi Bocconi

77 Transforming banking for the next generation
Bill Schlich, Global Banking & Capital Markets Leader, EY U.S.
Ian Baqps, Deputy Banking & Capital Markets Leader, EY U.K.
Steven Lewis, Global Banking & Capital Markets Lead Analyst, EY U.K.
Karl Meekings, Global Banking & Capital Markets Strategic Analyst, EY U.K.
Part 2: Tactical

93 Effectiveness of loan-to-value ratio policy and its transmission mechanism: empirical evidence from Hong Kong
Eric Wong, Market Research Division, Research Department, Hong Kong Monetary Authority
Kelvin Ho, Market Research Division, Research Department, Hong Kong Monetary Authority
Andrew Tsang, Market Research Division, Research Department, Hong Kong Monetary Authority

Pieter Klaassen, Group Risk Control, UBS
Idzard van Eeghen, Chief Risk Officer, Royal Bank of Scotland N.V.

113 Recursive collective action problems: the structure of procyclicality in financial and monetary markets, macroeconomies and formally similar contexts
Robert Hockett, Edward Cornell Professor of Law, Cornell Law School

129 Do “too-big-to-fail” banks take on more risk?
Gara Afonso, Senior Economist, Federal Reserve Bank of New York
João A.C. Santos, Vice President, Federal Reserve Bank of New York
James Traina, Senior Research Analyst, Federal Reserve Bank of New York

145 An overview of the risk-neutral valuation of bank loans
Danilo Tilloca, Vice President, Group Special Entities Risk and Evaluation, Unicredit Holding SpA
Luciano Tuzzi, Senior Vice President, Head of Group Special Entities Risk and Evaluation, Unicredit Holding SpA

155 Smart beta: too good to be true?
Bruce I. Jacobs, Principal, Jacobs Levy Equity Management
Kenneth N. Levy, Principal, Jacobs Levy Equity Management

161 Regulatory herding versus democratic diversity: history and prospects
Nicholas Dorn, Associate Research Fellow, Institute of Advanced Legal Studies, University of London

175 Squandering home field advantage? Financial institutions’ investing in their own industries
Aneel Keswani, Reader in Finance and Director of the Centre for Asset Management Research, Cass Business School, City University
David Stolin, Professor of Finance, Toulouse Business School, University of Toulouse

189 The new Banking Union landscape in Europe: consolidation ahead?
Dirk Schoenmaker, Dean, Duisenberg School of Finance, and Professor of Finance, Banking and Insurance, VU University Amsterdam
Executive summaries

The science of winning in financial services — competing on analytics: opportunities to unlock the power of data
By Hyong Kim, FSO Global and Americas Data and Analytics Leader, EY U.S. and Errol Gardner, EMEIA FSO Data and Analytics Leader, EY U.K.
This article presents the results of confidential survey data across banking, capital markets, insurance, and wealth and asset management into the role that data and analytics are now playing in their businesses. We find that (1) data is becoming a fourth “pillar” of the business for leading financial services firms, (2) a holistic approach to data is crucial for success (it is not just a technology problem), (3) leading companies are increasingly investing in data for the upside potential, rather than purely to mitigate downside risk, (4) financial services firms are bullish about their capabilities on analytics, but many admit to needing to do much more and (5) chief data officers (CDOs) are increasingly common, but are not a solution by themselves. These findings paint a mixed picture of the level of maturity in data and analytics at financial services companies. Some areas of the industry, and certain individual companies, have made significantly more progress than others. Insurers, for example, appear to be ahead of banks. Nevertheless, all financial services businesses are on the same journey, working toward a greater emphasis on decision-making that is based on the actionable insights generated from their data. They are thinking about how to overcome the obstacles encountered along the way — from fears about regulation and data privacy to the technical challenges presented by legacy systems and skill shortages — and organizing and investing in order to give themselves the best chance of doing so. Part of the challenge will be honest self-appraisal: a willingness to assess current levels of maturity in order to decide where to focus the business’s energies. New challenges, meanwhile, will emerge as companies seek to develop their capabilities. The size of the prize, however, makes these difficulties worth confronting.

Cross-border bank resolution
By José Viñals, Financial Counsellor and Director of the Monetary and Capital Markets Department, International Monetary Fund and Sean Hagan, General Counsel and Director of the Legal Department, International Monetary Fund
Cross-border banking has grown rapidly during the past couple of decades and has brought with it both benefits and costs. The recent financial crisis made it clear that, in the event of a cross-border bank failure, authorities have strong incentives to protect the stability of their own financial systems, potentially at the expense of global financial stability. In response, the international regulatory community has sought to develop a framework to facilitate globally efficient outcomes of cross-border bank resolution. But orderly cross-border resolution is still far from assured. Further alignment of interests and incentives among jurisdictions is needed, as are greater efforts to develop cooperative cross-border resolution strategies. Key priorities are expediting legal reforms at the national level to strengthen resolution frameworks and the powers of resolution authorities, enhancing banks’ loss-absorbing capacity in resolution and strengthening cross-border cooperation in recovery and resolution planning.
Shadow banking: policy challenges for central banks
By Thorvald Grung Moe, Special Adviser, Norges Bank
Central banks responded with exceptional liquidity support during the financial crisis to prevent a systemic meltdown. They broadened their tool kit and extended liquidity support to nonbanks and key financial markets. Many want central banks to embrace this expanded role as “market maker of last resort” going forward. This would provide a liquidity backstop for systemically important markets and the shadow banking system that is deeply integrated with these markets. But how much liquidity support should central banks provide to the shadow banking system without risking their balance sheets? And would not an accommodative market-making role send the wrong signals to market participants? I discuss the expanding role of the shadow banking sector and the key drivers behind its growing importance. There are close parallels between the growth of shadow-banking before the recent financial crisis and earlier financial crises, with rapid growth in near monies as a common feature. This endogenous ebb and flow of shadow-banking-type liabilities is indeed an ingrained part of our advanced financial system. We should think twice before we let central banks backstop the liquidity needs of private shadow banking markets, at least not before there has been substantial market reform. It would indeed be ironic if central banks were to declare victory in the fight against too-big-to-fail institutions, just to end up bankrolling too-big-to-fail financial markets.

The consequences of exit from non-conventional monetary policy
By Philip Turner, Deputy Head, Monetary and Economic Department, Bank for International Settlements
This paper examines the financial and macroeconomic consequences of changes in central bank balance sheets. Large-scale purchases of bonds tend to drive down long-term interest rates. But developments in global markets, which shape the world long-term rate, exert some powerful constraints. Exchange rate effects can also be significant. The prospect of divergent balance sheet policies (balance sheet normalization by the Federal Reserve but further large purchases by the European Central Bank (ECB) and the Bank of Japan) is pushing the dollar up. Moving balance sheets toward more normal levels is important in order to preserve policy flexibility for the future but will present central banks with formidable challenges. This task will require cooperation with treasuries without surrendering monetary policy independence. As central banks pragmatically monitor market resilience, the financial dominance trap is to be avoided.
The audit mandatory rotation rule: the state of the art
By Mara Cameran, Researcher, Accounting Department, Università Luigi Bocconi,
Giulia Negri, Research Fellow, Accounting, Control, Corporate and Real Estate Finance Department, SDA Bocconi School of Management and Angela Pettinicchio, Assistant Professor, Accounting Department, Università Luigi Bocconi

Mandatory audit rotation imposes periodical breaks to audit engagements and is intended to avoid excessively long relationships between the auditor and the client. The E.U. has finally introduced mandatory rotation for the audit firm in addition to the already existing audit partner rotation rules. The U.S., however, has for now decided to retain the partner rotation rule without introducing mandatory audit firm rotations. After an overview of the experience of a number of countries, we summarize the pros and cons of a compulsory change in the audit firm. Moreover, we focus on the empirical evidence collected on the benefits and costs of the rule. So far, investigations into the impact of the rule at corporate and market level have not been able to prove that the benefits outweigh the costs.

Transforming banking for the next generation

The next decade in banking will see both evolution and revolution. Banks must reinvent themselves, not just to respond to the pressures of today, but to be flexible enough to adapt to the world of tomorrow. The most successful banks will be those that have transformed their business models. Banks will increasingly focus on profitability rather than revenues. As a result, they will be defined by narrower scope and simpler structures, but greater reach. They will serve fewer customer segments, but some will operate across more markets. They will deconstruct products, stripping them back to their component parts so that customers can rebuild them tailor-made. At the same time, they will introduce new products that are more aligned to emerging client needs. Some will move to outsource back-office functions that no longer provide a competitive advantage, and some will operate key banking services for new competitors. Transformation is necessary because banks face an array of stakeholder pressures. They must find a way to deliver improved performance for investors who have tired of high volatility but low returns on equity. In doing so, banks will have to grapple with a low-growth environment across much of the developed world and slowing growth in the emerging world. To fund their transformation, they will need to become simpler and more efficient. As banks position themselves to deal with a “new mediocre” era of low growth, they will have to continue to adapt to a post-financial crisis environment, where an often divergent global regulatory reform agenda shows no signs of abating and customer trust must be regained. With governments and regulators becoming increasingly assertive, and customers ever more demanding, this will not be easy. And it will be made harder in a world where the best and brightest talent is looking for a career beyond financial services.
Effectiveness of loan-to-value ratio policy and its transmission mechanism: empirical evidence from Hong Kong

By Eric Wong, Market Research Division, Research Department, Hong Kong Monetary Authority, Kelvin Ho, Market Research Division, Research Department, Hong Kong Monetary Authority and Andrew Tsang, Market Research Division, Research Department, Hong Kong Monetary Authority

This paper provides a non-technical summary of two recent empirical studies to shed light on key important issues regarding the implementation of loan-to-value (LTV) policy as a macroprudential tool, including its effectiveness, potential drawbacks and its transmission mechanism to improve financial stability. Empirical evidence suggests that LTV policy is effective in reducing systemic risk associated with boom-and-bust cycles in property markets. Although the LTV policy may be associated with higher liquidity constraints on homebuyers, we show that the mortgage insurance program (MIP) can mitigate this drawback without undermining the effectiveness of LTV policy. Thus, MIPs play an important role in enhancing the net benefits of LTV policy. Concerning the transmission mechanism, empirical evidence suggests that the policy pass-through to property market activities may be weak. By contrast, there is clear evidence that tightening the LTV cap would reduce household leverage and credit growth, and that lower leverage plays a major role in strengthening banks’ resilience to property price shocks. This finding supports the view that household leverage would be an optimal target of LTV policy.


By Pieter Klaassen, Group Risk Control, UBS and Idzard van Eeghen, Chief Risk Officer, Royal Bank of Scotland N.V.

We introduce a new performance scheme for banks, inspired by the Du Pont scheme for corporates, which clarifies the relationship between return on equity (RoE), risk-adjusted return on capital (RAROC) and return on assets (RoA). The scheme highlights how common financial ratios, together with risk factors, influence the development of RoA, RAROC and RoE. The scheme can be applied by managers, analysts and regulators to analyze the performance of an individual bank as well as the performance of the banking sector as a whole. In addition, it can be used by bank managers to set coherent targets for various key financial ratios that tend to be managed separately within a bank, to achieve a target RoE, RAROC and RoA. We illustrate our performance scheme by applying it to analyze the main drivers behind the development of the performance of the U.S. commercial banking sector during the past 23 years.
Recursive collective action problems: the structure of procyclicality in financial and monetary markets, macroeconomies and formally similar contexts

By Robert Hockett, Edward Cornell Professor of Law, Cornell Law School

The hallmark of a collective action problem is its aggregating multiple individually rational decisions into a collectively irrational outcome. Arms races, “commons tragedies” and “prisoners’ dilemmas” are well-known, indeed well-worn, examples. What seem to be less widely appreciated are two complementary propositions: first, that some collective action problems bear iterative, self-exacerbating structures that render them particularly destructive, and second, that some of the most formidable challenges faced by economies, societies and polities are iteratively self-worsening problems of precisely this sort. Financial markets, monetary systems and macroeconomies in particular are rife with them – as are other complex systems subject to group-mediated procyclicalities or “feedback” effects. I call the mentioned challenges “recursive collective action problems,” and show that a great many familiar regulatory and policy challenges — including asset price bubbles and busts, consumer price hyperinflations and debt deflations, “paradoxes of thrift” and “recessionary spirals” — constitute instances of this general phenomenon. I also hazard suggestions as to how best to address such challenges. Key to the effort is first to recognize their shared structure, second to recognize that collective action problems require coherent collective agency for their solution, and third to recognize that the collective agents in question must act to render no longer individually rational such decisions as aggregate into collectively irrational outcomes. I close with specific examples of what problem-solving strategies informed by the “three recognitions” will tend to look like. The implications for macroeconomic and “macroprudential” finance-regulatory policy in particular are manifold. If we but attend to the shared nature, structure and pervasiveness of recursive collective action problems, I conclude, we can recoup much in the way of wealth and well-being that is now needlessly lost.

Do “too-big-to-fail” banks take on more risk?

By Gara Afonso, Senior Economist, Federal Reserve Bank of New York, João A.C. Santos, Vice President, Federal Reserve Bank of New York and James Traina, Senior Research Analyst, Federal Reserve Bank of New York

The notion that some banks are “too big to fail” builds on the premise that governments will offer support to avoid the adverse consequences of disorderly bank failures. However, this promise of support comes at a cost: large, complex or interconnected banks might take on more risk if they expect future rescues. This article studies the effect of potential government support on banks’ appetite for risk. Using balance sheet data for 224 banks in 45 countries starting in March 2007, the authors find higher levels of impaired loans after an increase in government support. To measure support, they rely on Fitch Ratings’ support rating floors (SRFs), a new rating that isolates potential sovereign support from other sources of external support. A one-notch rise in the SRF is found to increase the impaired loan ratio by roughly 0.2 – an 8% increase for the average bank. The authors obtain similar results when they assess the effect of increased support on net charge-offs and when they narrow their sample to U.S. banks only.
An overview of the risk-neutral valuation of bank loans
By Danilo Tillocca, Vice President, Group Special Entities Risk and Evaluation, Unicredit Holding SpA and Luciano Tuzzi, Senior Vice President, Head of Group Special Entities Risk and Evaluation, Unicredit Holding SpA
This paper provides an overview of a new methodology that allows banks to evaluate loans using the risk-neutral approach. Specifically, it illustrates the methodological framework behind the definition of the risk-neutral default probabilities used to estimate the loans credit spreads. These risk-neutral probabilities are calculated using a contingent-claims approach conceptually similar to the Black-Scholes and Merton framework for modeling corporate liabilities. The proposed risk-neutral approach is suitable for producing estimates, in a fair value computation context, that are as close as possible to the “exit price,” as mandated by International Financial Reporting Standards (IFRS 13), with a lower dependency on internal parameters.

Smart beta: too good to be true?
By Bruce I. Jacobs, Principal, Jacobs Levy Equity Management and Kenneth N. Levy, Principal, Jacobs Levy Equity Management
Smart beta strategies promise to deliver market-beating returns with simplicity and low cost, but the reality is more complicated. Contrary to popular perception, smart beta strategies are neither passive nor well diversified. Nor can they be expected to perform consistently in all market environments. Perhaps most importantly, because of their focus on only a limited number of factors, smart beta strategies fail to exploit numerous potential profit opportunities.

Squandering home field advantage? Financial institutions’ investing in their own industries
By Aneel Keswani, Reader in Finance and Director of the Centre for Asset Management Research, Cass Business School, City University and David Stolin, Professor of Finance, Toulouse Business School, University of Toulouse
In the extensive debate about investment professionals’ ability to add value, there has been scant evidence on the role played by industry expertise. To shed light on this issue, we study own-industry investing. Specifically, we analyze how well individual mutual funds as well as mutual fund companies, banks and insurance companies invest in the shares of listed companies in the mutual fund, banking and insurance industries, respectively. We find little evidence that such inherent industry knowledge enhances these institutions’ ability either to time the investment in their industry as a whole or to select individual stocks within it.
The new Banking Union landscape in Europe: consolidation ahead?

By Dirk Schoenmaker, Dean, Duisenberg School of Finance, and Professor of Finance, Banking and Insurance, VU University Amsterdam

The establishment of the Banking Union (B.U.) creates a large banking market comparable to that of the U.S. This paper calculates the market share of the top 20 banks in the new B.U. France appears to take a prominent place with 5 banks in the top 10, followed by Germany, the Netherlands and Italy. Earlier integration episodes did not lead to cross-border consolidation in Europe. In contrast, the lifting of interstate banking restrictions caused a cross-state merger wave cumulating in large U.S.-wide banks. This paper investigates whether cross-border consolidation can be expected within the B.U. The answer is yes over time, but not yet as subdued growth, lingering influence of national supervisors and cultural differences may hamper cross-border mergers in the short run. Over time, the B.U. will become an integrated market, where banks can manage their balance sheet at the aggregate B.U. level and the European Central Bank (ECB) conducts supervision with a European perspective.

Regulatory herding versus democratic diversity: history and prospects

By Nicholas Dorn, Associate Research Fellow, Institute of Advanced Legal Studies, University of London

Regulatory convergence – within the E.U., across the Atlantic and internationally – is conventionally represented as not only benign but also as essential in crisis prevention. This paper articulates a different frame of reference: one in which regulators “crowd,” “herd” and sometimes merge, so mimicking and exacerbating financial market tendencies toward similarity and contagion, and drawing regulators and markets into the same vortex. The paper looks at some of the historical and contemporary circumstances in the U.K., wider E.U. and the U.S. that have given reign to these tendencies and also at some aspects of regulatory architecture and governance that reduce such tendencies. It mentions pre-crisis tendencies to regulatory subservience to financial markets, with such subservience having a deep history in the U.K. and a shorter one in the U.S., so-called command regulation, which has the potential to either deepen subservience or transcend it; and the institutional preconditions for permanent regulatory vigilance, such as democratic appointment of heads of agencies. The paper concludes by pondering the prospects for the democratic direction of financial market regulation, in terms of its distributional logics and extraterritoriality.
Part 1: Strategic

The science of winning in financial services – competing on analytics: opportunities to unlock the power of data

Cross-border bank resolution

Shadow banking: policy challenges for central banks

The consequences of exit from non-conventional monetary policy

The audit mandatory rotation rule: the state of the art

Transforming banking for the next generation
Part 1: Strategic

The science of winning in financial services – competing on analytics: opportunities to unlock the power of data

Hyong Kim
FSO Global and Americas Data and Analytics Leader, EY U.S.

Errol Gardner
EMEIA FSO Data and Analytics Leader, EY U.K.

Abstract
This article presents the results of confidential survey\(^1\) data across banking, capital markets, insurance, and wealth and asset management into the role that data and analytics are now playing in their businesses. We find that (1) data is becoming a fourth “pillar” of the business for leading financial services firms, (2) a holistic approach to data is crucial for success (it is not just a technology problem), (3) leading companies are increasingly investing in data for the upside potential, rather than purely to mitigate downside risk, (4) financial services firms are bullish about their capabilities on analytics, but many admit to needing to do much more and (5) chief data officers (CDOs) are increasingly common, but are not a solution by themselves. These findings paint a mixed picture of the level of maturity in data and analytics at financial services companies. Some areas of the industry, and certain individual companies, have made significantly more progress than others. Insurers, for example, appear to be ahead of banks. Nevertheless, all financial services businesses are on the same journey, working toward a greater emphasis on decision-making that is based on the actionable insights generated from their data. They are thinking about how to overcome the obstacles encountered along the way – from fears about regulation and data privacy to the technical challenges presented by legacy systems and skill shortages – and organizing and investing in order to give themselves the best chance of doing so. Part of the challenge will be honest self-appraisal: a willingness to assess current levels of maturity in order to decide where to focus the business’s energies. New challenges, meanwhile, will emerge as companies seek to develop their capabilities. The size of the prize, however, makes these difficulties worth confronting.

---

\(^1\) This article is based on a quantitative survey of 150 financial services firms in the banking, insurance and asset management sectors, undertaken between August and September 2014. The respondents were split equally between North America, Europe and Asia. In addition, qualitative interviews were conducted with senior executives at a number of leading global financial services groups in order to gain further insights. We would like to thank the following individuals for their assistance with this report (listed alphabetically by surname): Mathias Born of Zurich Insurance Group, Andy Brown of Lloyds Banking Group, Brady Cole of Wells Fargo, Heiko Fischer of ING DiBa, Ashley Hirst of AIG, Alastair Kellock of Lloyds Banking Group, Holger Kumm of DZ Bank, Philipp Löffler of Deutsche Bank, Bruno Melo of Vontobel Asset Management, Alex Nakhapetian of Zurich Insurance Group, Wendy Seago of Aviva and Lorraine Waters of HSBC.
1. Introduction: reassessing the value of data

Data is fast becoming the “fourth organizational pillar” for modern financial institutions, alongside people, processes and technology. In today’s uncertain environment, characterized by regulatory and economic risk, as well as intensifying competition for customers, financial services leaders have an urgent need to make well-informed strategic decisions. As a result, many financial services firms are now working to create a data-centric business – and awakening to the enormous untapped value of their data.

“For the first time in the history of our company, data and the data strategy have been included in the business strategy that we publish externally – that says a great deal about our commitment and intent,” says Mathias Born, Head of Group Data Management at Zurich Insurance.

This recognition is widespread. In our survey of 150 financial services companies from around the world, more than four in five executives polled (83%) said they now see data as their most strategic asset. An almost identical number (84%) believe that data will be a source of competitive advantage.

“We’re turning ourselves into a data-centric organization,” says Lorraine Waters, Deputy Chief Data Officer at HSBC. “We’re increasing the awareness that, in this digital age, the quality of our data and the way in which we put it to work is absolutely key.”

At the same time, many financial services companies are a long way from extracting the value locked up in their data or exploiting it to its full potential. Almost half (47%) say they do not have a full grasp of that value. Nor are companies as advanced in their implementation of analytics technologies and processes as they are with data collection and management. Almost a third of respondents (31%) describe their ability to realize value from data assets today as immature.

This should not be surprising, as the enormous regulatory burden under which the financial services sector has labored since the financial crisis has been all consuming. The pressure to focus on compliance and risk management, rather than on growth opportunities, has undoubtedly contributed to the delayed realization of data’s value in the industry.

Case study: HSBC’s data-centric vision

“When you’re thinking about strategy and your target operating model, you think of people, process, technology, and we’ve added data, so now there are four pillars rather than three,” says Lorraine Waters, Deputy CDO at HSBC, the global banking group, where she also serves as Global Head of Data Governance.

Two years ago, HSBC appointed a CDO at group level, tasking him to build a data organization with teams that are embedded across. The effect, says Waters, has been to dramatically raise the profile of data as a distinct function throughout the group and to begin “turning HSBC into a data-centric organization.”

Part of the challenge is to confront the fragmented nature of any large financial services group, where customer data, for example, is held separately in a string of product silos. At the same time, HSBC is seeking to scale the analytics capabilities it has already built in parts of its business.

“We’re making data-specific decisions and creating data-specific capabilities,” Waters explains. “Historically, in most banks, data has played second fiddle. Banking tended to be very product aligned and all of a firm’s infrastructure tended to be set up along product or business lines – so, for customer data, for example, banks would persist with data in multiple retail banking systems, even if it’s the same customer in each case. What we’re trying to do is make the product system come to the central data system to ensure consistency across the product lines.”

Despite the huge strides the bank has already made, Waters says that getting an organization of HSBC’s size to a position where it is fully exploiting data and analytics capabilities is a long-term project. “We have incremental business benefits that we have to deliver over the next 1, 2, 3, 4 and 5 years, but, ultimately, this is a 10-year program.”
1.1 Turning the corner

Happily though, there is evidence that the financial services sector is now working to catch up. It is notable in our survey that higher-growth firms – defined as those businesses that have seen earnings before interest, tax, depreciation and amortization (EBITDA) growth of at least 15% per year over the past two years – are noticeably more motivated by the potential of new opportunities.

Meanwhile, many firms are racing to get ahead in this area, not least due to the potential that data holds. Australian bank Westpac attributes an AU$22m (approximately U.S.$17.3m) increase in revenues to big data techniques used to provide targeted offers to customers when they interact with the bank, whether in a branch, via a call center or online. The bank is taking advantage of a rapidly growing number of online customer interactions to develop its “know-me” approach, through gathering a huge volume of customer data and applying sophisticated analytics techniques.

However, while some in the sector are well on their way to capitalizing on data to gain a competitive advantage, other financial services companies are only just beginning to set clear priorities for utilizing their data. The challenge for these firms is to bridge the gap between recognizing their data’s value, and improving how data is collected, processed and analyzed, so it can become a competitive tool in future.

This paper is structured as follows: section 2 looks at the characteristics of leaders in data; section 3 considers how data and analytics tools might be used to enhance revenues and boost growth; section 4 provides a roadmap for overcoming the many hurdles business faces; and section 5 concludes.

2. The race to maturity: data leaders versus laggards

At first sight, the financial services sector regards itself as relatively mature in terms of how successfully it is able to generate value from its data assets. Almost two-thirds of respondents (65%) say their overall ability to realize value from data is somewhat or highly mature. The challenge for these firms is to be at the top of the curve, not just catching up. The survey results suggest that many companies are only just beginning to identify how data can be used to drive growth:

- Overall leadership on importance of data: 23% (Highly mature), 21% (Mature), 23% (Somewhat mature), 24% (Highly immature)
- Relevant data skills and human capital: 22% (Highly mature), 44% (Mature), 19% (Somewhat mature), 12% (Highly immature)
- Relevant data processes: 30% (Highly mature), 34% (Mature), 20% (Somewhat mature), 9% (Highly immature)
- Relevant technologies and systems: 29% (Highly mature), 36% (Mature), 24% (Somewhat mature), 19% (Highly immature)
- Setup of organizational structure for data: 21% (Highly mature), 31% (Mature), 21% (Somewhat mature), 11% (Highly immature)
- Data governance: 24% (Highly mature), 41% (Mature), 16% (Somewhat mature), 15% (Highly immature)

Many organizations are concerned about significant gaps in their capabilities, and there are clear leaders and laggards across the industry. For example, substantial numbers of companies are concerned about their maturity in each of the data-related competencies they were asked about. More than a third of companies described themselves as of only average maturity or worse when it came to possessing relevant data skills and human capital, implementing relevant data processes, the importance attached to data by management, adopting relevant technologies and systems and the setup of the organizational structure for data.

These shortcomings may explain other concerns expressed by many financial services businesses. Just one in four companies (26%) believes it has a full grasp of the potential value of the data their organizations hold, while over half describe themselves as being acutely challenged in sourcing the skills necessary to exploit their data.

Fragmentation is a common challenge. “It is not easy to join up data across a composite business,” says Wendy Seago, Technical Pricing and Performance Manager at Aviva. “With different systems in our life, healthcare and general insurance businesses, it can be difficult to gain a joined-up view of customers.”

---

2 “Westpac using big data to woo customers with offers made to measure,” The Sydney Morning Herald, 4 March 2014.
Moreover, not all financial services organizations are putting their data to good use. Brady Cole, Head of Wholesale Information Services at Wells Fargo, says: “I have a colleague at another firm who talks about ‘data Darwinism,’ that the person with the best data wins – my point about that is that unless you get a business benefit from working on data, the work just dies.” Specifically, benefit means insight that the business can use to make faster and better decisions that produce quantifiably better outcomes.

“You have to know what actionable insight you’d like to draw from the data,” says Alex Nakhapetian, who leads direct marketing efforts for Global Life at Zurich Insurance Group. “This challenge is an opportunity to achieve this clarity of thinking: you can’t have multiple data strategies and you have to be clear on what you want to achieve, but once that is in place, it becomes a beautiful thing to work through.”

In that context, only 16% of financial services companies describe their ability to extract insight from data to improve performance or competitiveness as excellent. And just 7% say they have sufficient numbers of data analysts throughout their businesses – the rest are still at various stages of trying to recruit people with the skills to deliver high-class analytics.

With so many financial services companies still in the process of building their analytics resources, certain areas of the business have been prioritized over others. On average, finance is at least twice as likely to be deriving business insight from data as any other function. The fact that areas such as sales, marketing and strategic planning lag so far behind may help to explain why many companies are not yet able to use analytics to deliver growth-oriented insight.

Nevertheless, some financial services businesses have been able to achieve more than others. The largest companies are much more likely to describe themselves as mature in terms of their ability to generate value from data, given the resources they have to tap into here. From a geographical perspective, North American executives typically rate their organizations as more mature than their European or Asia-Pacific counterparts.

While the finance function has been the focus of the analytics effort in years gone by, marketing is now singled out as the function where the biggest improvements in data-driven business insight will be seen next. At Lloyds Banking Group, for example, Andy Brown, Customer Data Director, says: “We’re working on ways we can drive more value from data for our customers and our business. We’re using data to understand what drives a customer to need a product, at a particular time; to continually develop our operational efficiencies; as well as simplify our processes.”

### Figure 2: How good do you believe your organization to be at extracting useful insights from its data to help improve its overall performance or competitiveness, in comparison with its main rivals?

![Figure 2: How good do you believe your organization to be at extracting useful insights from its data to help improve its overall performance or competitiveness, in comparison with its main rivals?](image)

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>16%</td>
</tr>
<tr>
<td>Above average</td>
<td>59%</td>
</tr>
<tr>
<td>About average</td>
<td>19%</td>
</tr>
<tr>
<td>Below average</td>
<td>5%</td>
</tr>
<tr>
<td>Poor</td>
<td>0%</td>
</tr>
</tbody>
</table>

3. Toward data-driven decision-making: delivering meaningful insights

The lesson to be learned from companies growing quickly and sustainably is that investment in new technology alone will not deliver what financial services companies are looking for – this spending needs to be supplemented with additional resources for the right people, processes and organizational change.

But the survey suggests that, in many areas, the appropriate level of resource has not been forthcoming. The number of companies that say their business functions have begun to recruit more data analysts is just 28%. Similarly, just 13% of companies expect to increase spending on data-related change management programs by more than 20% over the next two years.
High-growth firms: what is their secret?

One of the most notable results from this survey is that high-growth financial services companies — those achieving EBITDA growth of 15% or more in each of the last two years — are significantly more likely to be able to extract value from their data than other businesses in the research.

More than a third of these companies (35%) say they are highly mature in this capability, compared with fewer than one in four (23%) of the rest. At the same time, one in four (27%) high-growth companies describe itself as excellent at extracting useful insights that improve overall performance or competitiveness, more than twice as many as their lower-growth peers (12%).

What is their secret? The answer does not appear to be investment in technology, where increases in high-growth companies’ spending have broadly run in line with their rivals. Instead, high-growth companies are more likely to be significantly increasing spending in areas such as change management (24% plan to spend at least 20% more over the next two years, compared with just 7% of lower-growth companies) and personnel and skills (24% against 7%).

High-growth companies are also much more likely to have a specific individual with ultimate responsibility for analytics and insight (70% have such a role, compared with 58% elsewhere). And almost three times as many have overcome skills shortages in data analysts (14% against 5%), though there is clearly plenty of work still to do.

In other words, it is not the commitment to technology itself that is driving the high-growth companies’ lead in data and analytics, but their emphasis on the people, processes and structures necessary to ensure this technology delivers value throughout the business.

This may, in part, be because organizations are struggling to find the talent they need, rather than simply a failure to commit funds for the necessary investment. It is not that financial services companies do not understand the challenge. The vast majority (85%) agree that exploiting the data they have will require a strong focus on people and processes, not just technology.

One major issue holding the sector back is the ongoing demands of regulation: 21% of companies expect to increase data privacy-related spending by more than a fifth over the next two years. Another practical problem is the lack of skilled data scientists in what is a relatively new profession. “The ability to understand, manage and manipulate data analytically using traditional databases is something that has become ingrained across organizations in recent years,” says Andy Brown of Lloyds Banking Group. “But with the emerging tools and techniques, we will need more data specialists concentrated in key parts of the organization who are able to understand and manage the complexity of the data, the technology and the business implications.”

3.1 From insight to action

Financial services companies also need to address the question of how to disseminate insight so that it may be turned into practical action. “If an analyst can find stuff, that’s great, but how do we make it actionable or meaningful for the segment you’re looking at?” says Brady Cole of Wells Fargo. “It’s not our executives that actually touch the customers for the most part – it’s a relationship manager, and so, to turn insight into something actionable, you need to get to those sorts of people.”
The science of winning in financial services - competing on analytics: opportunities to unlock the power of data

These difficulties are resulting in only patchy success for companies seeking to move from recognizing the value of data to generating and exploiting insight. Some 68% of companies say they are effectively using data to drive more value from existing customers, for example, but only 58% are as confident about initiatives that will identify new customers.

Again, there are clear leaders and laggards. For example, large firms are much more likely than their smaller counterparts to be using data to effectively extract value from customers, identify risks, improve operational efficiency and make better decisions. North American firms also seem to be ahead of their peers elsewhere on most effectiveness measures.

Above all, however, it is the high-growth firms that once again stand out as leaders in the race to generate actionable insight. They are more than twice as likely to be using data very effectively to improve operational efficiency (41% of high-growth firms make this claim, compared with 19% of the lower-growth businesses) and show significant leads in most other areas too.

Lower-growth companies, meanwhile, are especially ineffective at using data to generate insights that will improve the identification of new customers and are also failing to pick up on opportunities to move into new product or service areas. One explanation for this may be that high-growth companies appear to be working hard to generate insight across a much broader range of functions. Asked which business functions are most advanced on analytics, high-growth firms’ answers were split broadly equally between finance (22%), sales (24%) and marketing (14%), while at other firms, the finance function dominated.

In other words, high-growth firms are more likely to be focusing their analytics efforts on the upside — new opportunities for revenue-enhancing growth — while lower-growth businesses have been less ambitious so far.

To see this in action, let’s take an example from Zurich Insurance Group, which is putting data to use in new and exciting ways.

“Not to make better decisions, but to be better decision-makers, and they are very different things.”

Hirst’s unit, which is distinct from a separate function in AIG responsible for collecting and managing data, is focused on working with teams within the business in order to help them make more decisions for themselves, based on analytic insight. “If we’re trying to help them be better decision-makers, then we need to build tools, put them in their hands and teach them how to use those tools, even when we’re not in the room,” he explains.

This approach, Hirst argues, is the key to generating value from data. “We create far more value by working with people who have genuine experience and insight into their business areas and giving them another arrow for their bows than we ever would by simply working on a set of specific questions” he says.

Crucially, Hirst’s team does not operate as a stand-alone technology function but as a business partner with data analysts and scientists who work throughout AIG. The idea is that they’re on hand to respond to particular challenges but also have sufficient understanding of the business to generate their own ideas.

“"We believe that the data toolkit, in its widest sense, combined with great quant talent and great business consultant talent, can add value to every single stage of the value chain in every single part of the company," Hirst adds.

Case study: Making data a science at AIG

“The primary measure of success is, have I helped AIG’s decision-makers be better decision-makers?,” says Ashley Hirst, Head of Science for the Americas and Europe, Middle East and Africa (EMEA) at the global insurer AIG. “Not to make better decisions, but to be better decision-makers, and they are very different things.”

Hirst’s unit, which is distinct from a separate function in AIG responsible for collecting and managing data, is focused on working with teams within the business in order to help them make more decisions for themselves, based on analytic insight. “If we’re trying to help them be better decision-makers, then we need to build tools, put them in their hands and teach them how to use those tools, even when we’re not in the room,” he explains.

This approach, Hirst argues, is the key to generating value from data. “We create far more value by working with people who have genuine experience and insight into their business areas and giving them another arrow for their bows than we ever would by simply working on a set of specific questions” he says.

Crucially, Hirst’s team does not operate as a stand-alone technology function but as a business partner with data analysts and scientists who work throughout AIG. The idea is that they’re on hand to respond to particular challenges but also have sufficient understanding of the business to generate their own ideas.

“"We believe that the data toolkit, in its widest sense, combined with great quant talent and great business consultant talent, can add value to every single stage of the value chain in every single part of the company," Hirst adds.

Again, there are clear leaders and laggards. For example, large firms are much more likely than their smaller counterparts to be using data to effectively extract value from customers, identify risks, improve operational efficiency and make better decisions. North American firms also seem to be ahead of their peers elsewhere on most effectiveness measures.

Above all, however, it is the high-growth firms that once again stand out as leaders in the race to generate actionable insight. They are more than twice as likely to be using data very effectively to improve operational efficiency (41% of high-growth firms make this claim, compared with 19% of the lower-growth businesses) and show significant leads in most other areas too.

Lower-growth companies, meanwhile, are especially ineffective at using data to generate insights that will improve the identification of new customers and are also failing to pick up on opportunities to move into new product or service areas. One explanation for this may be that high-growth companies appear to be working hard to generate insight across a much broader range of functions. Asked which business functions are most advanced on analytics, high-growth firms’ answers were split broadly equally between finance (22%), sales (24%) and marketing (14%), while at other firms, the finance function dominated.

In other words, high-growth firms are more likely to be focusing their analytics efforts on the upside — new opportunities for revenue-enhancing growth — while lower-growth businesses have been less ambitious so far.

To see this in action, let’s take an example from Zurich Insurance Group, which is putting data to use in new and exciting ways.

“"Our ability to detect various triggers that would speak to the propensity of buying certain insurance products is the most pertinent example,” says Alex Nakhapetian. “We must harness the disjointed data elements and convert them into insights.”

4. Clearing the way: how to cope with bumps in the road ahead

The way ahead for financial services companies is becoming clearer as they work toward unlocking the value in their data. However, there will be some very specific obstacles to navigate as the journey continues. Many companies are already conscious of at least some of these difficulties and are beginning to take steps
to confront them. In particular, the sector is primarily grappling with five broad challenges.

1. **Regulation and data privacy concerns still dominate**
   The top two obstacles listed by companies worried about how to maximize their use of data are concerns over regulatory issues (44%) and data privacy restrictions (37%). Financial services firms are right to focus on compliance and risk management, given the potential for such serious financial and reputational damage in the event of a failure. Nevertheless, they must work hard to ensure that this focus does not prevent them from exploiting the upside potential of data. Further, they must also think about privacy and security in the context of customers. “If you use data, you have to create customer value because if it’s not valuable for the customers, they will immediately ask why we are using their data,” argues Heiko Fischer, Head of IT & Project Management at ING-DiBa.

2. **Firms need to develop a clearer vision of what they want to achieve with their data**
   Just one in four firms (26%) disagrees with the suggestion that it does not have a firm grasp of the potential value of its data. Until financial services businesses have accurately mapped what data they currently collect, what might be possible to collect and the potential insight to be extracted from this information, they will not have a full picture of this value – or be able to make rational decisions about which workstreams they should prioritize most highly. Businesses must learn to walk before they try to run, argues Holger Kumm, Head of IT Application Development of DZ Bank. “The data discussion is pointless if you haven't done the basics,” he argues. “The first step is to develop the capability to draw sensible conclusions from the data you already have.”

3. **Legacy systems make it difficult to introduce new technology solutions**
   Many financial services companies operate as groups of businesses that have traditionally operated independently of one another. Many have developed after years of M&A transactions. As a result, these firms typically have a wide array of legacy IT systems that are poorly connected and unable to communicate with one another. Bridging the gaps between these systems or replacing them with new platforms represents a serious challenge. “Most of our legacy systems come from the companies we have acquired,” says Mathias Born of Zurich Insurance. “The problem is that, whenever you need to get a view on any aspect of our business, you have to deal with a multitude of sources – sometimes the end user is not even sure what the source is, and that makes it tough to trust the insights you’re generating.”

4. **Many firms are struggling with skill shortages in key areas**
   Data and analytics are relatively new disciplines and companies in many industries report that it is difficult to find sufficient numbers of people with the right skills. In analytics, in particular, there are simply not enough data scientists to go round. More than one in four financial services companies (27%) sees this issue as a particularly serious obstacle holding its organization back. There are various challenges here. One is that companies need a mix of skills and experience: both those with a pure data science background, and those with more experience of commercial applications. Another is that there is also a need for those capable of working in isolation and for others who can operate within a specific business unit. “There is sometimes a desire for analytics to be driven centrally, but you do need local presence as well to understand local issues and
help to embed the use of analytic solutions,” says Aviva’s Wendy Seago. Meanwhile, despite the competitive salaries the financial services sector can offer, there is significant competition for these kinds of skills from other sectors, notably the hi-tech industry.

5. Unstructured data represents a particular challenge for many firms
Unstructured data – everything from social media inputs to audio and video streams – is a rich resource for those companies with the skills and technologies required to collect this information and to extract value from it. However, many firms lack these attributes: 31% of those polled see this as an obstacle. “We’re still working out how to manage our structured data,” says one executive at a large banking group. “We haven’t even scratched the surface when it comes to unstructured sources.”

5. Making the leap: how to get there
What, then, does the road map look like for financial services companies as they journey toward greater maturity in data and analytics? How do these businesses extract more value from their data, generate greater insight, and boost performance and competitiveness in broader and bolder ways?

The examples set by the leaders in this research provide some of the answers to those questions, particularly for the laggards struggling to catch up. But even those financial services businesses currently leading the way in data and analytics have plenty of work to do before their decision-making processes become entirely data-driven. Three areas in particular will be crucial.

5.1 Investment
It is clear that investment in data and analytics must continue, but it is not only technological capability that should be the priority for spending. People and process will also require their share of resources – and investments will need to be made across the enterprise, rather than in only a handful of more obvious areas.

One important issue for many companies will be how to make the shift to growth-oriented and revenue-enhancing initiatives while still ensuring compliance with the rising regulatory burden covering data.

Currently, one in five financial services companies (21%) expects spending related to data privacy to increase by at least 20% over the next two years, while a further 15% say the same about spending on governance and oversight. By contrast, fewer businesses foresee such dramatic increases in data-related investment in personnel and skills (11%), change management (13%) or technology (13%).

That is understandable, given the threat posed by data legislation. For example, in March 2014, the European Parliament voted the E.U.’s new data protection regime one step closer to reality. If adopted, the proposed Data Protection Directive will impose fines of up to €100m, or up to 5% of a firm’s annual worldwide turnover, for businesses that suffer data breaches. Unsurprisingly, both European and larger firms are spending more on data privacy than other companies in this research.

For many financial services companies, the regulatory burden is especially tough. The Basel Committee on Banking Supervision (BCBS) 239 regulation, which came out of the committee’s work on risk data aggregation and risk reporting, will require systemically important financial institutions to comply with new standards in this area – meeting this challenge is already a major project for many firms.

“What BCBS is saying in essence is that the board of the company needs to have a high degree of confidence that the information
it is using to make decisions is accurate and reliable,” explains Alastair Kellock, Head of Risk Transformation at Lloyds Banking Group. “We have been investing heavily in risk infrastructure since the financial crisis so we’ve done some of the work, but BCBS is still a big change – we need to leverage the work done already and fill in the gaps.”

Nevertheless, businesses that concentrate all their data investments on compliance run the risk of missing out on opportunities that can deliver positive growth. As this report has already noted, high-growth companies, in particular, understand the imperative – they are investing three times more on the skills and human capital needed, as well as three times more on change management.

Similarly, while risk management (20%) and finance (18%) have been the highest priorities for financial services companies making investments in data and analytics over the past two years, resources will need to be shared more broadly between functions. There are encouraging signs here. Though risk management is set to continue to dominate, with 26% of firms expecting it to continue to be the number one priority over the next two years, sales, marketing and strategic planning are all set to rise up the pecking order. High-growth companies, in particular, intend to

invest heavily in data initiatives in sales and marketing. “I’ve held the role of head of direct marketing for global life for a number of years, and data is the latest addition to my responsibilities,” says Alex Nakhapetian of Zurich Insurance Group. “Data is arguably the most important asset for a successful direct marketing methodology – my task is to ensure it is treated as an asset, not a constraint.”

An important target for financial services companies will be to pay more attention to measuring return on investment, so that the right projects are prioritized. At Aviva, Wendy Seago says: “My team is maybe unusual for an analytics team, but we are all about the money.” The practical result, she says, is that it is the most valuable and widespread initiatives that get the green light. “We work very closely with the business areas we support and if we can't see a way to embed something and to get value out of it, we don't pursue it,” she explains.

5.2 Leadership and organizational structure
Financial services companies are still trying to decide what organizational structure is most fit for purpose in order to embed data and analytics capabilities throughout the business. Some issues are more clear-cut than others – for example, 92% of
The science of winning in financial services - competing on analytics: opportunities to unlock the power of data

those polled have already appointed a specific individual with responsibility for data across the organization or intend to, while 81% will create a board-level committee for data.

But while the importance of strong leadership in data is being confronted, potential problems remain. Most worryingly, 39% of companies currently do not have a specific individual with responsibility for analytics or insight, and almost a third of these firms have no plans to appoint anyone to this key role. Data integration and management (34%) suffers from the same difficulty.

One common question is whether to appoint a CDO, someone who can act as a dedicated leader to help create a more data-centric business. The CDO is already a major force in many financial services companies, driven not least by greater pressures, following the financial crisis, to ensure better data quality and transparency. CDOs have been appointed at 38% of the companies surveyed for this report, and are already far more prevalent in some sectors of the financial services industry. In some firms, CDOs are taking over from chief information officers (CIOs) as the individuals with primary responsibility for data, although almost a third (31%) still ask the CIO to lead on this.

Behind the scenes, though, this is often simply a division of focus: the CIO might remain responsible for the underlying IT infrastructure for data, say, while the CDO focuses more on the relevant data policies and governance. “Traditionally, our CIO role focused more on the cost management and service management side of IT,” says Mathias Born of Zurich Insurance. “Now, we have established a CDO role whose mandate is to define the enterprise data management strategy and to maintain and even increase the long-term value of data.”

Leadership is not the only organizational issue to contend with; the wider structure of the data and analytics function is also an ongoing debate. For example, many of the interviewees in this report stress the importance of data initiatives being driven by business leaders with a grasp of the challenges and opportunities facing their companies, rather than by technical experts from IT. Yet almost one in four firms (23%) has no plans to take responsibility for data out of IT and into the business.

In other areas, though, progress is being made. Within the next two years, nearly all companies (94%) will have at least begun creating an enterprise-wide technology platform for data. Almost as many (92%) will be taking steps to identify and address organizational data silos.

Given the diversity of these challenges and the need to ensure that as many as possible are being confronted, it is understandable that the trend has, until now, been for centralization of the data and analytics function, particularly at larger companies, with two-thirds (67%) of firms set up in this way. Over time, however, as businesses mature, there is expected to be a modest shift toward more dispersion of the function throughout the organization.

In certain areas, high-growth firms are particularly likely to be organized in a certain way. For example, they are more likely to have individual leaders in place, especially in analytics, and less likely to intend to leave data and analytics within IT.

At Deutsche Bank, Philipp Löffler, Business Intelligence Platform Management in Group Finance, talks in terms of a “business intelligence value chain.” “We start with the operating systems, in which the contractual data is captured and processed; then we take this data from the operating system and store it in...
data warehouses; out of these warehouses, which are the golden source for everything, we can use purpose-by-purpose applications that we call data marts — these extract certain data from the data warehouse and then focus on a specific area of analysis.”

Nevertheless, Löffler agrees that there is no one-size-fits-all approach to organizing for data and analytics. Depending on a company’s operating model, many differing approaches can work: “These are historically grown landscapes that are complex,” he says.

5.3 Governance
By introducing formalized governance processes that are disseminated, understood and followed throughout the business, financial services companies may, in time, be able to effectively automate some of the compliance work that currently makes such extensive demands on resources.

Structure can be a big help, says Alastair Kellock of Lloyds Bank. “If we have a structure and governance in place, embedding policy changes and achieving compliance is easier;” he argues. “If you don’t have that structure, it is difficult to implement a change like BCBS across a number of areas in a consistent way.”

Case study: Building the platform for analytics at Vontobel Asset Management
“Our management rightly recognizes the value of data and analytics, and that is driving a reconfiguration of our business,” says Bruno Melo, Director and Head of Portfolio Risk at Vontobel Bank and Vontobel Asset Management. “We are introducing a new data system in our asset management business that will replace the three or four separate systems with which we currently operate.”

Until now, the firm had focused on ensuring that its asset management data, both internal and external, including peer analysis, was clean and robust, with relatively little analytics work conducted. “That feels a little limited, given the amount of data we receive,” Melo says. “The ambition now is to go much further – with all our systems coming together, we will be able to use our data far more intelligently to create value.”

Part of the challenge for Vontobel will be to build a more robust data warehouse within the firm’s newly created unitary platform, though the business is still thinking through reporting lines and other structures. “The point is for our data team and its analytics team to collaborate much more closely than they do today,” Melo adds.

Crucially, Vontobel sees this restructuring as a major strategic move. “Replacing these systems with one platform isn’t about saving money, though it will do so in the longer term;” he says. “It’s strategic positioning in an industry where we are competing with peers who are also investing in these capabilities, including some very big players, in order to create a solution that will give us so much more chance of generating insights and value.”

That work has already begun. Within two years, 98% of companies will have a formal, enterprise-wide policy for data governance in place. Many organizations already have an extensive set of governance policies and frameworks in place – more than two-thirds have adopted policies for data privacy, data security and social media usage, for example.
That said, there is still more to do. Just 43% of companies have policies in place relating to employers who use their own devices for work, while only 41% have a data inventory policy. Moreover, not all companies are confident about their ability to enforce those policies they do have in place – just a third (37%) say their organization is highly capable of doing so.

There are also other potential governance issues to address. Only half of the companies surveyed (49%) say they regularly conduct data quality audits, just 4 in 10 (39%) regularly train staff on governance policies and procedures, and fewer than a third (29%) conduct regular risk assessments aimed at identifying potential data vulnerabilities.

As in other areas, the high-growth companies stand out as having fewer weaknesses on data governance. On enforcement of policy, for instance, 44% say they are highly capable of policing governance across the organization. They are also more likely to regularly conduct policy reviews, risk audits and other checks.

While the high-growth companies are not perfect, the progress they have already made on governance has left them with more capacity to focus on extracting greater value and insight from their data.

6. Conclusion: competing on analytics
Making the transition from data and analytics immaturity to a business enjoying significant competitive advantage over its rivals will require a systematic approach to improving capabilities. For now, though, too many financial services companies are still dominated by data management solutions. They have yet to embed analytics into operational decisions on an enterprise-wide basis. And their capabilities are patchy, with no consistency of approach or maturity throughout the business.

For many financial services firms, data and analytics initiatives have, until now, been driven by the risk and regulatory agendas, rather than by the potential for growth. Now is the time to switch from defense to offense.

Making that switch will require greater maturity in both data and analytics. Those that have already moved furthest on the former will find the latter easier – they will be able to depend on good quality data and to integrate internal and external data, for example. But analytics maturity spans a spectrum too, starting with descriptive and backward-looking analysis to techniques such as predictive analytics.
Cross-border bank resolution

José Viñals
Financial Counsellor and Director of the Monetary and Capital Markets Department, International Monetary Fund

Sean Hagan
General Counsel and Director of the Legal Department, International Monetary Fund

Abstract
Cross-border banking has grown rapidly during the past couple of decades and has brought with it both benefits and costs. The recent financial crisis made it clear that, in the event of a cross-border bank failure, authorities have strong incentives to protect the stability of their own financial systems, potentially at the expense of global financial stability. In response, the international regulatory community has sought to develop a framework to facilitate globally efficient outcomes of cross-border bank resolution. But orderly cross-border resolution is still far from assured. Further alignment of interests and incentives among jurisdictions is needed, as are greater efforts to develop cooperative cross-border resolution strategies. Key priorities are expediting legal reforms at the national level to strengthen resolution frameworks and the powers of resolution authorities, enhancing banks’ loss-absorbing capacity in resolution and strengthening cross-border cooperation in recovery and resolution planning.

1 This article draws from the IMF staff paper “Cross-Border Bank Resolution: Recent Developments,” June 2014. The authors wish to thank Johannes Forss Sandahl for assistance in the preparation of this article. The views expressed in this article are those of the authors and do not necessarily represent those of the IMF or IMF policy.
1. Introduction
Between the early 1990s and the start of the global financial crisis, cross-border banking linkages more than doubled to more than half of global GDP International Monetary Fund (IMF) (2014b). This generated benefits for the global economy as banks pursued growth opportunities beyond their borders. Not only did they export capital and banking technologies to other countries, but they also realized economies of scale in managing large, diversified operations, while pooling risk across economies.

However, the global financial crisis underscored the costs and risks of increased financial interconnectedness. In the face of this systemic shock, large cross-border banking groups became instant transmitters of financial stress across borders. Banks retrenched to favor their home markets, and cross-border banking claims dropped from more than half of global GDP in 2007 to about one-third in 2012 [IMF (2014b)]. The high concentration of cross-border activities in a relatively small number of banks amplified the impact of these institutions on the global financial system. In 2008, 145 banks globally accounted for 85% of assets held by the world’s top 1,000 banks [IMF (2014a)].

The crisis exposed several gaps in the domestic and international frameworks for bank resolution. Faced with inadequate information and little incentive to cooperate with their foreign counterparts, national authorities focused on protecting domestic financial stability. This often led to suboptimal outcomes in the resolution of troubled financial institutions. Some authorities took unilateral actions such as the “ring-fencing” of assets and liabilities in domestic jurisdictions, even though such actions raised overall costs. Nor did resolution authorities have the necessary powers and tools to resolve insolvent financial institutions, which led to slow and inefficient resolution processes. Jurisdictions also lacked effective mechanisms for loss allocation, including the ability to credibly subject bank creditors to loss in a bank failure without endangering financial stability. As a result, sovereign financial strength emerged as a key factor in determining national strategies.

This article discusses challenges associated with the resolution of large cross-border financial institutions, recent regulatory initiatives in this area and the gaps that remain in the framework for resolution of cross-border banks. It concludes that, in order to prepare for future crises, further action is needed to put in place an effective cross-border bank resolution regime. These include the more widespread implementation of reforms and policies already agreed at the international level, and efforts to fill remaining gaps in the framework. In addition, efforts to align incentives for cooperation among national authorities are needed, including measures that may affect the structure of cross-border institutions.

2. Problems related to cross-border resolution
The financial trilemma is a conceptual framework that can be used to describe the problems related to cross-border resolution (see, for example, Claessens et al. (2010)). Its essence is that it is difficult for governments to simultaneously achieve the objectives of a stable global financial system, the freedom for banks to engage in cross-border activities and the preservation of sovereignty (Figure 1).

One possible combination of the trilemma objectives is to allow for cross-border banking and to achieve global financial stability by sharing supervisory and resolution resources between countries. This would require authorities to give up some aspects of their national sovereignty. Another option is to prohibit cross-border banking, and thereby facilitate global financial stability and maintain national sovereignty. This could, however, reduce...
the efficiency of the global financial system. The most common approach has been for national authorities to allow for cross-border banking while, at the same time, retaining control over decisions relating to the resolution of banking operations in their jurisdictions. The global financial crisis showed the limitations of this approach, because difficulties in achieving effective cooperative solutions between national authorities exacerbated the initial shock.

A key obstacle to an effective international solution is the mandate of national governments and resolution authorities to preserve financial stability and protect taxpayer interest in their home jurisdictions. Preserving international business lines may minimize the overall economic costs of a cross-border bank failure, but individual country perspectives may differ from the globally optimal approach. Consider a cross-border bank failure that affects financial stability in both home and host countries. If the bank is systemically important in the host country, but not in the home country, or vice versa, it may be difficult for authorities to cooperate because their incentives to commit public resources to bank resolution will differ greatly.

In practice, problems in cross-border cooperation may emerge for several reasons, as became evident during the global financial crisis. The lack of effective resolution powers and tools to enable authorities to manage resolution efficiently at the national level was a key factor, as was the fact that existing mechanisms for cooperation were weak. Disparities in the level of expertise and financial resources between the authorities in different jurisdictions undermined effective cooperation. Problems also arose from differences in accounting rules, and legal and institutional frameworks for resolution.

3. International regulatory reform
Against this background, the international community has done a great deal of work since the financial crisis to strengthen the legal and regulatory framework for the resolution of large, cross-border financial institutions. Strengthening bank capital and liquidity requirements has been a major element of the reform effort. These initiatives aim to create incentives to curb, or provide adequate buffers against, risk-taking in the financial sector, thus reducing the probability of crises.

But future financial crises cannot be entirely ruled out, and initiatives to establish a clear and credible resolution framework for systemically important banks remain critical. It is for this reason that the Financial Stability Board (FSB) and its members have put in place the Key Attributes of Effective Resolution Regimes for Financial Institutions (Key Attributes) – a new international non-binding standard for the design of an international legal and regulatory framework for the resolution of large cross-border financial institutions. The Key Attributes were endorsed by the G20 leaders in 2011 and are a major step forward in ending the perception that some financial institutions are “too big to fail,” including globally systemically important financial institutions (G-SIFIs). The IMF has played an active role in the development of the Key Attributes, in particular through its 2010 board paper on resolution of cross-border banks – a proposed framework for enhanced coordination (IMF (2010)).

The Key Attributes specify principles for the design of cross-border resolution regimes at the national and international levels. In particular, they seek to accomplish three broad objectives:

- Ensuring that national legal frameworks give resolution authorities a comprehensive toolkit of powers to resolve troubled financial institutions, including by taking control of failing financial institutions; merging ailing banks with stronger ones; transferring assets and liabilities to a healthy institution or a so-called “bridge bank” and engaging in a mandatory restructuring, or “bail-in,” of the creditors of the institutions
- Facilitating cross-border cooperation by ensuring that national legal frameworks include effective mechanisms for the recognition and enforcement of foreign resolution measures and for the exchange of information with foreign resolution and supervisory authorities
- Ensuring the establishment of “crisis management groups,” effective recovery and resolution planning, and the conclusion of cooperative agreements between national resolution authorities in relevant jurisdictions for the resolution of G-SIFIs

While the Key Attributes represent an important step forward, the agenda for the establishment of an effective international resolution framework is far from complete. At the national level, the Key Attributes are yet to be fully implemented in many jurisdictions.

---

2 IMF (2014a) reviews several cross-border resolutions from the global financial crisis.
Cross-border bank resolution

jurisdictions. A peer review conducted by the FSB in 2013 found that implementation within the FSB membership was at an early stage [FSB (2013)]. A number of countries have since reformed their resolution regimes, but progress remains mixed overall (see Table 1).

There are also questions about the degree of implementation that might be appropriate for smaller jurisdictions and financial entities. Parts of the Key Attributes were designed primarily for G-SIFIs that operate in deep, sophisticated financial systems. While many features of the Key Attributes have broad application beyond G-SIFIs, not all elements may be relevant to all countries. Reforms to enhance resolution regimes will need to be tailored to the complexity of financial systems in different jurisdictions.

And finally, additional work remains to be done at the international level to fully address the problems of misaligned incentives for cooperation in cross-border resolution. The following sections cover key areas in need of attention.

4. Enhancing mechanisms for loss allocation

Remaining gaps in the agenda largely pertain to how losses will be allocated to private sector stakeholders and how to manage the residual risk that public funds will be needed in resolution.

4.1 Making bail-in a credible option

The “bail-in” regime for the mandatory restructuring of the debt of a failed bank is conceptually attractive but will need to be calibrated carefully to be reliable in practice. Under a “bail-in,” the resolution authority restructures the liabilities of a distressed financial institution by writing down its unsecured debt and or converting it into equity while its business remains ongoing. This reduces the risk that public funds will be required in resolution. Market funding costs for large institutions would adjust as a result, reducing “too-big-to-fail” distortions. Key to the credibility of a bail-in strategy is that there exists sufficient “bail-inable” debt within the institution and that holders of claims targeted for bail-in are able to absorb losses without generating systemic risk themselves as a result of their financial losses.

The FSB has made substantial progress in developing a proposal on total loss absorbing capacity (TLAC) that aims to ensure a buffer of liabilities beyond equity that can be subject to mandatory restructuring in resolution. Key elements of the proposal concern the definition of the claims that would be regarded as credibly loss-absorbing while protecting financial stability, their amount and their location within a banking group. The consultative document for TLAC was submitted to the G20 leaders in Brisbane and will be finalized after incorporation of public comments and a quantitative impact assessment [FSB (2014)].

An appropriate investor base and transparency will be critical to the success of this initiative. For example, cross-holdings of bail-inable liabilities among financial firms may need to be constrained to mitigate contagion risk in the event of a bail-in. More generally, there should be adequate disclosure of TLAC such that buyers of these instruments can assess the riskiness of their investments and price them accordingly.

The location of TLAC within a banking group will also be crucial, as it will determine ex ante (before the event) how losses accrue in resolution. Losses may arise in entities within a financial group that are in jurisdictions different from where TLAC is located. Mechanisms will need to be credibly in place to move losses to where TLAC is, or vice versa. These could include, for example, loans from the parent company to the subsidiary, which could be converted into equity, or collateralized guarantees by the parent on borrowing by the subsidiary.

There remain a number of challenges. Issuing TLAC on a significant scale may prove difficult because of the limited size of markets for unsecured, subordinated debt instruments. The yield of instruments perceived as bail-inable can be expected to rise as a result of the new regime, which would increase banks’ funding costs. Policymakers also need to consider which TLAC requirements are reasonable in jurisdictions where capital markets are less developed and where banks are largely deposit-funded. Another challenge will be the need to carefully balance resolution objectives with those of effective bank supervision and, in the process, to ensure that bank governance after resolution is not made weaker. In particular, the conversion of bank debt into equity in a bail-in should ensure that new controlling shareholders that emerge meet regulatory suitability requirements.

3 Notable among these is the European Union’s Bank Recovery and Resolution Directive adopted in May 2014 by the European Parliament, which would help bring national resolution frameworks in closer alignment with the Key Attributes.
4.2 Aligning creditor hierarchies
Creditor hierarchies determine the order in which losses will be allocated against the claims of various stakeholders in the liquidation of a firm, as well as in a restructuring of a bank in resolution. In the latter case, the restructuring of the bank's claims will generally take into account the order of priorities applicable in liquidation, and treat all of the claims that fall into a particular class in the same manner, although deviations may sometimes be necessary. Significant differences in creditor hierarchies between jurisdictions can be an obstacle to the cooperative resolution of branches of an international bank. For instance, if some senior creditors, such as uninsured depositors, receive preferential treatment in the host country but not under home country law, the host authorities may have an incentive to ring-fence instead of contributing all local assets and liabilities to a resolution per home country law. Incentives for cross-border cooperation will also be undermined if the home country's creditor hierarchy discriminates against foreign or nonresident creditors, or creditors of foreign branches.

The Key Attributes provide useful guidance on the hierarchy to be employed in the context of a bank resolution. For example, they provide that resolution powers should be exercised in a way that respects the hierarchy of claims, while providing flexibility to depart from the general principle of equal (pari passu) treatment of creditors of the same class, with transparency about the reasons for such departures to contain the potential systemic impact of a firm’s failure or to maximize the value for the benefit of all creditors as a whole. They also call on authorities to avoid discrimination among creditors based on nationality, the location of their claims or the jurisdiction in which it is payable. But they also leave important questions unanswered – in particular, on the treatment of insured and uninsured deposits in the creditor hierarchy. Giving preference to depositors (depositor preference), including possibly uninsured depositors, can be useful in mitigating the likelihood of contagion, or effects of bank failures on the payment system and economic activity. Where a deposit insurance scheme is in place, a framework that gives preference to insured deposits mitigates the risk that the deposit insurance scheme will suffer losses. As noted by IMF staff in a 2014 board paper (IMF (2014a)), further guidance at the international level on the treatment of insured and uninsured deposits in the creditor hierarchy would promote cooperation among jurisdictions in the resolution of cross-border banks.

4.3 Recovery and resolution planning
Regulatory initiatives aim to reduce the probability of bank failures and their expected fiscal cost. It cannot, however, be ruled out that state funds will have to be used to inject liquidity or capital into a distressed bank. For this reason, jurisdictions need to jointly prepare for the resolution of cross-border institutions. However, it has proven difficult to establish ex ante agreements between jurisdictions, defining how public support would be committed to resolve a weak or failing institution (Schoenmaker (2010)).

Shifting decisions and actions to the pre-crisis phase can help implement a resolution strategy successfully and reduce the risk to potential public funding for resolution. Recovery and resolution planning can help develop a picture of structural imbalances in the distribution of assets and liabilities in a financial group from a national perspective. The Key Attributes require jurisdictions to put in place recovery and resolution plans (RRPs), covering, at a minimum, domestically incorporated firms that could be systemically significant or critical if they fail. Through RRPs, an assessment can be made of structural financial imbalances between countries embedded in cross-border financial groups' balance sheets, as well as assessments of the potential for sovereign risk transfer. This will help the authorities gauge potential public sector costs in the event of failure. As a result of these assessments, financial groups can, for example, be required to adjust their structure, regulate the placement of some assets and liabilities or put in place prudential limits on specific exposures. In this manner, incentives can be better aligned beforehand to ensure resolution plans will be executed in practice as agreed. More widespread adoption of the RRP requirement at the national level is necessary to help promote effective cross-border resolution for all systemically important banks.

5. Conclusions
There has been significant progress in developing an effective global cross-border resolution framework. Widespread implementation of the Key Attributes will provide resolution authorities involved in cross-border bank failures with a broad set of tools for executing resolution strategies. This will minimize risks to financial stability and force shareholders and private creditors to bear most of the costs of such bank failures.
Cross-border bank resolution

But orderly cross-border resolution is still far from assured. Should a large cross-border bank fail today, policymakers would likely be faced with the same misaligned incentives that undermined cooperation in the global financial crisis. To further align the interests and incentives of jurisdictions and to facilitate cooperative cross-border resolution strategies, additional work is required to:

- Implement the Key Attributes across a wide range of jurisdictions
- Make bail-in a credible option by finalizing and implementing a consistent standard on loss-absorbing capacity in terms of eligible claims, amount and location
- Better align creditor hierarchies and depositor preference between countries

In addition, progress on minimizing residual risks that public funds will be needed to preserve financial stability is necessary to ensure that resolution strategies are credible. International cooperation in recovery and resolution planning can move decisions and actions to the pre-crisis phase, ensure incentives are aligned and minimize the risk that public funds will need to be committed in the resolution of a cross-border financial institution.

While a great deal has been accomplished since 2008 in ending “too big to fail,” much remains to be done (see Table 1). Through its work with the FSB and with its own members, the IMF is committed to making an effective international framework for the resolution of cross-border financial institutions a reality.

References

FSB, 2011, “Key attributes of effective resolution regimes,” Financial Stability Board

<table>
<thead>
<tr>
<th>Gap identified in the crisis</th>
<th>Measures taken</th>
<th>Implementation status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Inadequate resolution powers and tools</td>
<td>Key Attributes 2, 3, 10, 11</td>
<td>Partial: legal changes and recovery and resolution planning under way in several jurisdictions</td>
</tr>
<tr>
<td>2. Inadequate frameworks for cross-border coordination</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inadequate information</td>
<td>Key Attributes 7, 12</td>
<td>Partial: crisis management groups established; guidance on information sharing agreed in 2014</td>
</tr>
<tr>
<td>Insufficient cooperative arrangements</td>
<td>Key Attributes 7, 8, 9</td>
<td>Partial: recovery and resolution planning under way</td>
</tr>
<tr>
<td>Absence of legal recognition</td>
<td>Key Attributes 3, 7</td>
<td>Partial: country practice varies on statutory recognition; FSB work ongoing; ISDA protocol on contractual solutions agreed</td>
</tr>
<tr>
<td>Differences in creditor hierarchies</td>
<td>No measures proposed</td>
<td>Uniform depositor preference in the E.U. under Bank Recovery and Resolution Directive</td>
</tr>
<tr>
<td>3. Inadequate measures for loss allocation</td>
<td>Key Attribute 3 (bail-in power), TLAC agreement</td>
<td>Partial: bail-in powers not universally implemented; TLAC standard being finalized</td>
</tr>
<tr>
<td>Cross-border loss allocation</td>
<td>Key Attributes 7-12</td>
<td>Partial: recovery and resolution planning under way</td>
</tr>
</tbody>
</table>

Table 1: Progress toward effective cross-border bank resolution
Shadow banking: policy challenges for central banks

Thorvald Grung Moe
Special Adviser, Norges Bank

Abstract
Central banks responded with exceptional liquidity support during the financial crisis to prevent a systemic meltdown. They broadened their tool kit and extended liquidity support to nonbanks and key financial markets. Many want central banks to embrace this expanded role as “market maker of last resort” going forward. This would provide a liquidity backstop for systemically important markets and the shadow banking system that is deeply integrated with these markets. But how much liquidity support should central banks provide to the shadow banking system without risking their balance sheets? And would not an accommodative market-making role send the wrong signals to market participants? I discuss the expanding role of the shadow banking sector and the key drivers behind its growing importance. There are close parallels between the growth of shadow banking before the recent financial crisis and earlier financial crises, with rapid growth in near monies as a common feature. This endogenous ebb and flow of shadow banking-type liabilities is indeed an ingrained part of our advanced financial system. We should think twice before we let central banks backstop the liquidity needs of private shadow banking markets, at least not before there has been substantial market reform. It would indeed be ironic if central banks were to declare victory in the fight against too-big-to-fail institutions, just to end up bankrolling too-big-to-fail financial markets.

1 The views expressed in this article are those of the authors and not necessarily of their employers.
1. Introduction
Untraditional central bank liquidity operations prevented a systemic meltdown after Lehman Brothers filed for bankruptcy in September 2008. Central banks broadened their tool kit and extended liquidity support from systemic institutions to core funding markets. Subsequently, they provided additional liquidity support through asset purchases, large-scale market interventions and other creative ways of easing credit conditions. As a result, central banks’ balance sheets have grown dramatically compared with GDP.

Scholars broadly agree that these untraditional polices saved the global financial system from a systemic meltdown in 2008. Extending liquidity support beyond the traditional banking perimeter to shadow banking markets (e.g., asset-backed securities, money market instruments and commercial paper) prevented a wholesale financial panic. Backstopping liquidity in core funding markets is increasingly seen as a natural extension of the traditional “lender-of-last-resort” function of central banks [Carney (2008), Mehrling (2014)] and this market-making function could well become a standard part of the central bank toolkit [Carlson et al. (2015)]. Some are, however, concerned about extending the government safety net too far and would prefer to rein in the expansion of the shadow financial system with stricter regulation [Tarullo (2013, 2015), Turner (2013)].

The shadow banking system represents a special policy challenge for central banks, since its growth is closely linked to the regulation of the banking system. The transmission of monetary policy is also affected by the size and behavior of the shadow banking system [Nelson et al. (2015), Stein (2014)].

The shadow banking system represents a special policy challenge for central banks, since its growth is closely linked to the regulation of the banking system. The transmission of monetary policy is also affected by the size and behavior of the shadow banking system [Nelson et al. (2015), Stein (2014)].

Strengthening oversight and regulation of shadow banking is high on the G20 agenda for 2015. The updated FSB road map includes significant policy proposals related to margins and haircuts, securities financing, rehypothecation, money market funds and the asset management industry (FSB (2014b)). Getting shadow banking policies right is important for the market-based credit system and the stability of the broader financial system. This paper reviews some of the recent developments in the shadow banking system, describes its endogenous process of credit creation, points to some similarities with the past, discusses the increasing use of collateral and how this web of securities financing transactions increases systemic risk and contagion, and then finally lays out the policy challenges facing central banks trying to safeguard core funding market liquidity. I conclude that central banks should not embrace an expanded market-maker role without meaningful shadow banking reform.

2. Shadow banking redefined
There is a growing awareness that the shadow banking system is not a financial system distinctly different from regulated banking. Banks are big players in the shadow banking system, both as collateral providers and as repo participants. Money market funds are major funding sources for the big banks, and the over-the-counter (OTC) derivatives market is an integral part of the shadow banking system through its extensive reliance on pledged collateral. Gabor (2013) shows that big banks are dominant in the shadow banking system in Europe, and reports from the Bank for International Settlement (BIS) note that a few global banks dominate the global OTC market [BIS (2014)]. By recasting the shadow banking debate in this light, we can appreciate that many of the ongoing regulatory debates on collateral policies, haircuts, liquidity rules, high-quality liquid assets, risk weights for sovereign debt and the central banks’ role as market maker of last resort are indeed tightly connected.

When G20 leaders in 2009 asked the FSB “to identify the main risks related to shadow banking and eliminate all dark corners in the financial sector;” the initial approach focused on institutions outside the regulated banking system that could represent a threat to financial stability. Based on this approach, the size of the global shadow banking system was estimated to be almost U.S.$70 trillion, or 25% of global financial intermediation [Adrian et al. (2013)].

Strengthening oversight and regulation of shadow banking is high on the G20 agenda for 2015. The updated FSB road map includes significant policy proposals related to margins and haircuts, securities financing, rehypothecation, money market funds and the asset management industry (FSB (2014b)). Getting shadow banking policies right is important for the market-based credit system and the stability of the broader financial system. This paper reviews some of the recent developments in the shadow banking system, describes its endogenous process of credit creation, points to some similarities with the past, discusses the increasing use of collateral and how this web of securities financing transactions increases systemic risk and contagion, and then finally lays out the policy challenges facing central banks trying to safeguard core funding market liquidity. I conclude that central banks should not embrace an expanded market-maker role without meaningful shadow banking reform.

2 Based on the total assets of “other financial intermediaries” of the national accounts data, but not including insurance and pension funds.
The FSB responded by establishing five separate work streams to deal with the challenges of shadow banking and the European Union. Commission launched its own consultation on how best to tackle systemic risk stemming from shadow banking entities and activities (FSB (2011), E.U. (2012)). Through this work, a more nuanced view of shadow banking has emerged, focusing in particular on fragile repo funding, securities lending, derivatives trading, global liquidity creation and money market financing (E.U. (2013), Gabor (2013)). As the former FSA chairman Adair Turner observed (Turner (2012a)): “We need to understand shadow banking not as something parallel or separate from the core banking system, but deeply intertwined with it.”

Key to this “new” understanding of the shadow banking complex is the collateral intermediation function that underpins the financial plumbing of our market-based financial system (Singh (2014)). The procyclical nature of this collateral-based financial system, through funding and asset price fluctuations, is now seen by many as the essential feature of the shadow banking system. This “new view” of shadow banking is reflected in the updated FSB roadmap towards strengthening oversight and regulation of shadow banking (FSB (2014b)) and the E.U. regulation on transparency and reporting of securities financing transactions (E.U. (2014)), where the focus is now squarely on financial activities, such as securitization, securities lending, repo markets and rehypothecation, rather than on institutions.

The FSB Global Shadow Banking Monitoring Report 2014 (FSB (2014a)) notes that “further improvements in data availability and granularity will be essential for authorities to be able to adequately capture the magnitude and risks in the shadow banking system.” Such data will be essential to judge risks and the potential systemic impact of the shadow banking system. They add: “In the future, improvements in data availability should allow for the mostly entity-based focus of the ‘macro-mapping’ to be complemented with an activity-based monitoring to cover developments in relevant markets where shadow banking activity may occur, such as repo markets, securities lending and securitization (ibid.).”

The capacity of the shadow banking system to operate on a large scale in a way that creates bank-like liabilities through a complex chain on collateral transactions has created multiple forms of feedback into the regulated banking system. The use and reuse of collateral exacerbates procyclical dynamics and makes the whole financial system more fragile. When times are good, market participants tend to be more willing to let counterparties reuse collateral, increase market liquidity and thereby lower the cost of capital. But in more stressed market conditions, market participants become more sensitive to counterparty risk and more reluctant to let their collateral be reused. This puts additional strains on already tight liquidity conditions and tends to amplify the procyclical of the shadow banking system. These recent initiatives to collect more information should give us valuable insights into the “mechanics” of shadow banking, but will not, by itself, address the procyclical nature of this nonbank credit system.

3. Shadow banking in the past

That private money is not cash and that all IOUs are not equal should not come as a surprise. The collapse of the shadow banking system during the recent global financial crisis is not unprecedented if we look closer at earlier crises. Henry Thornton (1802, Chapter III, p. 37) made similar observations in his 1802 book *An enquiry into the nature and effects of the paper credit of Great Britain*: “When confidence rises to a certain height in a country, it occurs to some persons that profit may be obtained by issuing notes, which purport to be exchangeable for money; and which, through the known facility of thus exchanging them, may circulate in its stead.”

Hyman Minsky (1982) noted that this desire for more cash than is available from its usual source sows the seeds for next financial crisis. During a boom, the margin of safety decreases, and economic units take on more and more leverage. Money markets have a tendency to expand during boom periods, providing an elastic source of private credit. As money markets expand, a general decline in the liquidity of households and firms follows. This makes them vulnerable to fall in asset prices. There will be a general expectation about liquidity in key asset markets that cannot be sustained unless the central bank moves in and supports the price, i.e., monetization by the central bank. But this is surely “fair-weather” liquidity, since “no one would seriously defend the proposition that all things should be made liquid” (Simmons (1947)).

Andrew Haldane (2012) adds that “cycles in money and banking credit were indeed familiar from centuries past” and yet, for
some odd reasons, these insights were ignored for perhaps a generation, with near-fatal consequences for us all. “Investors and firms did not expect asset market liquidity to be impaired or funding disruptions to last for so long” [quote from Senior Supervisor Group Report (2008)]. The sudden collapse in liquidity conditions when the Reserve Primary Fund “broke the buck” in September 2008 came as a big surprise, and market stability was only restored after central banks intervened with unprecedented liquidity support.

Former Federal Reserve Chair Ben Bernanke observed that the financial crisis can best be understood as “a classic financial panic transposed into the novel institutional context of the 21st century financial system” [Bernanke (2013)]. He draws our attention in particular to the Panic of 1907, when financial innovations gradually undermined the coordinating role of the clearinghouse and lightly regulated trust companies were used to circumvent reserve requirements. When investors realized that the market was overextended, there was a sudden rush to realize positions, leading to fires sales and further losses. A steep decline in interbank lending was important in both episodes. And, much of the panic occurred outside the traditional banking system, in the shadow banking sector.

The perception that claims on trust companies (or shadow banks) was as good as cash was based on explicit or implicit promises by their sponsors to provide liquidity and credit support. Or the perception was based on the high ratings of the securitized assets on their balance sheets [Tarullo (2013)]. But as a BIS report noted 25 years ago [BIS (1986)]: “The presumed superior liquidity of securitized assets over conventional bank loans may turn out to be a mirage if a substantial number of the creditors attempt to liquidate their holdings simultaneously.”

The fire sales in 2008 resembled the panic liquidation by trust companies in 1907. The sudden withdrawal of funding led to rapid deleveraging and “repo runs.” Fire sales of securities into falling markets created adverse feedback loops of mark-to-market losses, margin calls and further liquidations. This “unwinding of the risk illusion, that is, the assumption that lending to shadow banks was essentially risk-free, helped transform a dramatic correction in real estate valuations into a crisis that engulfed the entire economy” [Tarullo (2013)].

This endogenous nature of private credit (and liquidity) was not sufficiently appreciated before the crisis. Inside money expands like ripples in the pond during the upswing on the back of private promises to pay (back). 3 As Hayek observed in 1931, “the characteristic peculiarity of these circulating forms of credit is that they spring up without being subject to any central control, but once they have come into existence their convertibility into other forms of money must be possible if a collapse of credit is to be avoided” [Hayek (1931)].

This convertibility of inside money (shadow bank money) into outside money (cash) is achieved when central banks intervene in a crisis to support vanishing market liquidity. But how far should central banks stretch their balance sheets to support liquidity in these private, endogenous markets? This becomes a pressing question when markets have grown at an exponential pace, such as the repo and OTC derivatives markets. Should taxpayers’ monies be put at risk to support a financial system with such “excess credit elasticity”? 4

4. The challenge of endogenous finance

The rapid growth of shadow banking has challenged the traditional view of banking, where banks would receive savings and then intermediate them toward the most productive uses. Banks were supposed to receive a tangible “good” – savings – and pass it on to the investor; nothing was lost in the process. The alternative, and more realistic, view of banking now recognizes that “banks can create money out of nothing” [McLeay et al. (2014)]. So can shadow banks – where demand for leverage meets demand for safety [Pozsar (2015)]. It then follows logically that privately created money can disappear as well – in a liquidity spiral. As Adrian and Shin (2009a) note, “... when liquidity dries up, it disappears altogether rather than being re-allocated elsewhere.”

Global liquidity is today highly influenced by this interplay between banks and nonbank financial institutions and the ebbs and flows of risk perceptions in global financial markets [BIS (2011)]. International bank credit exhibits strong boom-bust cycles that appear to correspond closely to episodes of financial distress, and periods of strong growth in cross-border credit

---

3 See Gurley and Shaw (1960) for the distinction between inside and outside money.
4 Borio and Disyatat (2011) introduced the term “excess elasticity” of the financial system.
are often characterized by elevated risk appetite. This “self-
reinforcing interaction between risk appetite and liquidity is not
yet sufficiently appreciated” (Cœuré (2012)), even though it is
quite clear that private liquidity has become highly endogenous to
the conditions in the global financial system.

Adrian and Shin (2009b) explore the hypothesis that “the
financial intermediary sector, far from being passive, is
instead the engine that drives the boom-bust cycle.” Rather
than looking at how financial frictions might affect the real
economy, they go straight to the financial sector and try
to understand how finance became the propagator of the
crisis instead of a conduit for prosperity. They note that
securitization was intended to disperse risks associated with
bank lending so that investors who were better able to absorb
losses would share the risks [Adrian and Shin (2009a)]: “But
in reality, securitization worked to concentrate risks in the
banking sector. There was a simple reason for this. Banks and
other intermediaries wanted to increase their leverage — to
become more indebted — so as to spice up their short-term
profit. So, rather than dispersing risks evenly throughout
the economy, banks and other intermediaries bought each
other’s securities with borrowed money. As a result, far from
dispersing risks, securitization had the perverse effect of
concentrating all the risks in the banking system itself.”

The procyclical process of liquidity creation raises particular
challenges for central banks. First, as Hyman Minsky pointed
out long before the recent financial crisis, “securitization implies
that there is no limit to bank initiative in creating credits, for
there is no recourse to bank capital” (Minsky (1987)). This
makes the supply of credit almost infinitely elastic, as every new
“euphoric era means that an investment boom is combined with
pervasive liquidity-decreasing portfolio transformations” [see
Minsky (1982), and also Borio (2013) on the “excess elasticity”
of the financial system]. Second, the more recent experience
with quantitative easing (QE) has shown that bank credit is quite
autonomous and difficult to influence, as the link between bank
credit and central bank money is weak. Private liquidity tends to
move quite independently of the prevailing stance of monetary
policy, reflecting private sector risk perceptions (the risk channel)
and the ease of arranging nonbank financing (via the shadow
banking infrastructure). These liquidity cycles are then amplified
by the rise and fall in collateral prices, which again propagate
through the collateral chains of the shadow banking system.

Banks and shadow banks are not just allocating pre-existing
savings; collectively, they create both credit and deposits [Turner
(2012b)]. Their cyclical behavior is now at the heart of the more
violent swings in the financial cycles that we have experienced
over the last two decades.

This new financial landscape requires a reorientation in both
theory and policy. Before the crisis, money and credit were seen
as either redundant or at least inessential in the mainstream
New Keynesian paradigm (Borio and Disyatat (2011)).
Standard models are based on one representative, riskless
agent, so anyone’s IOU could and would be immediately and
fully acceptable in payment for goods or services [Goodhart
and Tsmocos (2011)]. There is no need for money! Building
new models that capture the interaction between the financial
and the real sectors and the role of credit are now a key
preoccupation of academics and policymakers. This may
require some novel approaches, as mainstream theory needs
to interact with, and build on, insights from non-traditional
schools of thought. As Borio and Disyatat (2011, p. 31) note,
a deeper understanding of financial crises and the workings
of our modern finance-based global economy will require “a
rediscovery of the essence of monetary analysis.”

5. The steps forward
Fortunately, there is a rich theoretical tradition dealing with the
instability of financial markets that can be tapped to improve
our understanding of modern capitalist economies with banks,
finance and credit. One major contributor is Hyman Minsky, who
built his financial instability theory on the back of J. M. Keynes’s
deep insights into the working of a modern monetary economy.
According to Martin Wolf (2012) of the Financial Times,
“His masterpiece Stabilizing an unstable economy, provides
incomparably the best account of why the mainstream theory
is wrong,” i.e., that the modern capitalist economy is inherently
stable. “Periods of stability and prosperity sow the seeds of their
downfall. The leveraging of returns, principally by borrowing,
is viewed as a certain route to wealth. Those engaged in
the financial system create — and profit greatly from — such
leverage. When people underestimate perils, as they do in good
times, leverage explodes.”

No wonder that former Fed Chairman Alan Greenspan admitted
“shocked disbelief” while watching his “whole intellectual edifice collapse in the summer of 2007” and that he confessed that he had “put too much faith in the self-correcting power of free markets.” And he added: “... the immense and largely unregulated business of spreading financial risk widely, through the use of exotic financial instruments called derivatives, had gotten out of control and had added to the havoc of the crisis.”

The increased procyclicality of the financial system has led to a reorientation in policy. In addition to policy measures aimed at strengthening the robustness of financial institutions, there is now a greater willingness to address the endogenous credit cycles more directly. Macropudrinal instruments will be targeted at excessive credit growth, and central banks and supervisory authorities will work together to improve underwriting standards [IMF (2013)]. In addition, there is a greater willingness among policymakers to intervene in the free workings of financial markets, as “markets are no longer viewed as self-stabilizing” [Tett (2013)].

It remains to be seen if the proposed reforms will be enough to dampen the endogenous cycles of finance. The extraordinary expansion of shadow banking credit is still supported by the preferential treatment of repo and derivative transactions under bankruptcy law [Perotti (2012, 2013)]. And lax rehypothecation rules still encourage the buildup of collateral chains that propagate failure between key actors in core funding markets. As noted, such breakdowns in market liquidity could again lead to pressure for central bank interventions. Central banks’ liquidity policies are thus closely related to the developments in the shadow banking sector and the “changing collateral space” [Singh (2013a)].

6. Shadow banking and collateral pressures
The shadow banking sector is both a user of collateral and a collateral provider. According to Mannmohan Singh of the IMF, shadow banking is really a network of collateral transactions that today constitutes our modern financial system [Singh (2014)]. This “collateral landscape” is now changing, due to various regulatory initiatives and the general move toward more secured financial transactions. The result is a scramble for safe assets and increasing concerns about collateral shortages in the future [IMF (2012)].

Several reports have analyzed the potential shortages of highly liquid collateral [IMF (2012), BIS (2013b)]. Many argue that there will not be a shortage of high-quality liquid assets (HQLA), since primary issuance is expected to remain fairly high going forward [U.S. Treasury (2013)]. However, there could be a scarcity of HQLA, especially if markets become stressed again [BIS (2013a)].

The effectiveness of netting and the size of net exposures will determine the final demand for HQLA. There will surely be effects on pricing, market structure and the workings of markets more broadly [Heath et al. (2013)]. But the true level of asset encumbrance and the risk it poses for banks is so far unknown, and the financial stability implications of increased collateralization of financial transactions and rising asset encumbrance levels remain poorly understood [Gai et al. (2013)].

One predictable effect of the upcoming scramble for HQLA is “collateral transformation services” that can expand the HQLA universe. As collateral becomes increasingly scarce, a key shadow banking function will be to mobilize and reuse collateral [Singh (2014)]. They will source collateral (from insurance companies and pension funds), increase the collateral velocity (i.e., reuse), pool collateral (among firms in the same company) or recreate securitization (creating what appeared to be high-quality assets, as was customary before the recent crisis) [Hauser (2013)].

The downside of this collateral transformation is more interconnections between key players in the financial market and increased risk of contagion. As Singh (2013b) notes: “Collateral transformation is likely to fill the void, but will increase the nexus between banks and nonbanks.” And these new interconnections between financial institutions will weaken...
the resilience of the financial system in adverse conditions [Heath et al. (2013)]. Policymakers, therefore, need to strike a balance between the desire to ensure the soundness of financial institutions and the costs associated with a potentially too-rapid acquisition of safe assets to meet this goal (IMF (2012)).

This concern with shortages or scarcity of HQLA has led to increased pressure on central banks to relax their liquidity policies; banks want cheaper funding and wider collateral pools. They also want to include central bank liquidity facilities in their pool of liquid assets under the new Basel LCR liquidity regulation. Such “committed liquidity facilities” (CLF) could potentially reduce the banks’ need for mobilizing new HQLA. But it would surely undermine the spirit of the Basel liquidity proposal (Schmitz (2013), except in jurisdictions that are short of sovereign debt [Stein (2013a)].

The new collateral-intensive financial system confronts central banks and governments with a deeply political question: how to manage the potential systemic risk generated by the shadow banking system, especially in times of stress [Gabor (2013)]? As regulators try to instill more safety in the system, transaction costs will increase, prices go up and volumes fall. But scaling back the profitable OTC market may be like putting the genie back in the bottle. The pushback from the financial industry over the proposed OTC reforms shows that this will be a tough battle.8 And the new market equilibrium for highly liquid assets is indeed “hard to fully fathom in advance” (Stein (2013a)).

There is also the risk that pressure to collateralize the huge unsecured repo and OTC positions may expose clearing agents (CCP) to new and unexplored concentration risk. This could put pressure on central banks to provide even more liquidity in a crisis to avoid a new systemic meltdown [Murphy (2013), Tucker (2014)].9 And increased collateral requirements would also expose the financial system to procyclical and self-reinforcing spirals as market participants will repo, swap or sell assets to meet collateral calls in times of stress [ESRB (2012)].

The huge scale of the collateral-based shadow banking system represents a dilemma for central banks [Moe (2012)]. Unless the endogenous growth in shadow banking liabilities is somehow constrained, there will be continued pressure on central banks to stop fire sales and create outside liquidity in periods of stress [Perotti (2012)]. However, before central banks commit to backstopping the liquidity needs of the shadow banking system, more conceptual work is needed to clarify the scope for self-insurance against liquidity risk and how to define the modalities of central bank liquidity support [Tarullo (2013, 2014)].

7. Collateral dilemmas

Central banks’ liquidity policies were transformed during and after the financial crisis. Many central banks initiated new and innovative liquidity facilities with a wider set of counterparties, at much longer maturities and against a gradually much wider set of collateral.10 Without this timely liquidity support, the breakdown in market liquidity would most likely have led to the disorderly failure of a number of major financial institutions.

Carlson et al. (2015) argue that this expanded role for the Federal Reserve in liquidity provision was a natural extension of the classical lender-of-last-resort policy prescribed by Walter Bagehot.11 “The Fed lent not only to banks, but, seeking to stem the panic in wholesale funding markets, it also extended its lender-of-last-resort facilities to support nonbank institutions, such as investment banks and money market funds, and key financial markets, such as those for commercial paper and asset-backed securities” (Bernanke (2013)). The scale of liquidity support was massive, as “the Fed’s balance sheet was being used to directly replace the decline in balance sheet capacity of the financial intermediary sector” [Adrian and Shin (2009a)].

The massive increase in central bank liquidity support led to changes in their collateral policies. In principle, central bank credit should only be granted to solvent firms against good collateral.12 This is a safeguard against reckless money growth,

---

8 The Financial Times (2013) notes that the last adjustments in the liquidity rules reduced the largest U.S. banks’ need for liquid assets from U.S.$840 billion to U.S.$192 billion; see also Financial Times (2014) on the tweaking of the leverage ratio regulation.

9 Murphy (2013) notes that “without access to a central bank, a CCP could find itself unable to fund itself in the event of a crisis.”

10 Madigan (2009) provides the rationale for the liquidity policies of the Federal Reserve during the crisis.

11 The Bagehot Rule (Bagehot (1873)) states that central banks should lend early and freely to solvent firms against good collateral and at high rates.

12 How to determine if a counterparty is indeed solvent is a challenging task. According to Stein (2013a): “The line between illiquidity and insolvency is far blurrier in real life than it is sometimes assumed to be in theory.”
Shadow banking: policy challenges for central banks

and limits the central bank’s exposure to financial loss and lessens the need for counterparty credit assessment [Cheun et al. (2009)]. A shortage of eligible collateral acts like a brake on central bank credit, as an anchor, much like gold under the gold system of international finance. A strict collateral policy would, in this way, help in preserving the integrity of the fiat money system. Central bank credit should be backed (and constrained) by something of real value.13

But the breakdown in unsecured interbank credit after the crisis put commercial banks in a squeeze. Their liquidity needs increased dramatically, while their counterparties withdrew collateral at the same time. In response, central banks relaxed their traditional strict collateral requirements in order to accommodate the bank’s desperate need for liquidity. And banks became more creative in finding ways to post low-quality, but acceptable, collateral at the central bank, because better quality collateral had alternative uses with better returns. This type of behavior was well known even before the crisis, as observed by a former European Central Bank (ECB) executive board member [quoted in Chailloux et al., 2008, p.5]:14 “Quite understandably, (central bank counterparties) have economized on the use of central government bonds, which has been often almost the only collateral counterparties could still use in interbank repo markets. Instead, they have brought forward less liquid collateral … including ABSs, for which primary and secondary markets have basically dried up.”

By facilitating this type of “collateral manufacturing,” central banks’ collateral policies facilitated the buildup of leverage before the crisis in the banking and the shadow banking systems. Banks could use their high-quality collateral to obtain repo financing, thereby providing pledgeable collateral for the daisy chains of rehypothecation in the shadow banking system. By running an accommodative collateral policy before the crisis, many central banks thus supported the excessive market growth that they eventually had to bail out during the crisis with even more relaxed collateral standards.

The recent changes in collateral policies of the Bank of England can be seen as a natural extension of this accommodative liquidity policy [BoE (2013)]. The new Governor announced the policy with the headline: “We are open for business” [Carney (2013)]. Facilities will, in the future, be on longer terms, the range of assets accepted as collateral will be wider, including raw loans if required, all facilities will be cheaper and there will also be foreign exchange liquidity lines, based on international swap lines (ibid.).

This new liquidity policy is consistent with the Governor’s long-held position that “in times of crisis, central banks must act as a market-maker of last resort, by becoming a counterparty to major market participants” [Carney (2008)]. But embracing this expanded role also carries certain risks. As Martin Wolf (2013) noted after the new BoE policy was announced: “A central bank can, in principle, create domestic money without limit. But if it uses that power more freely, it will encourage banks and markets to generate more maturity transformation, making themselves and the economy more vulnerable to panic.”

Central banks are thus faced with an inherent tension between the market’s need for liquidity in times of crisis and the strictures of their own collateral rules. As Zorn and Garcia (2011) from the Bank of Canada observed: “The benefits of a flexible collateral policy were demonstrated during the crisis, but how flexible should collateral policies be? How much risk can or should a central bank take on? How can operational readiness to accommodate this flexibility be balanced with the costs, particularly when extraordinary events are, by definition, infrequent?”15

If banks perceive the central bank’s collateral policy in a (new) crisis correctly, they may well hold less good collateral and more bad collateral today (a form of Gresham’s law). And the central bank may not be able to stick to their announced strict collateral policy, just as the banks expect. As Paul Tucker (2009) noted: “In other words, a central bank policy of lending against only the best assets is likely to prove time inconsistent when it comes to the crunch.”

If central banks insist on only highly liquid assets as collateral for liquidity support in a crisis, some otherwise solvent banks with liquidity problems may fail. This is obviously a policy

13 See Lehmenbecker (2008) for a statement of the German Property School of Economics’ view on collateralized money.
14 José Manuel González-Práramo, ECB Executive Board Member, June 2008.
15 See BIS (2015) for a recent discussion of some of these issues.
dilemma for central banks. They risk amplifying the financial crisis by tightening their lending standards during it. This is obviously counterintuitive, as they are supposed to rescue the financial system in a crisis. But it illustrates well the tensions between a “risk-based” collateral policy (Alphandary (2015)) and a “macroprudential-based” liquidity policy (Allen (2013)). A countercyclical central bank collateral policy could indeed be useful in dampening the financial cycle “and provide some funding alternatives when conditions in the market become tight and build an illiquidity discount into some asset prices” (Chailloux et al. (2008)). However, such a policy can only be viable if “collateral neutrality” is restored in normal times. “Otherwise, central banks would increasingly ease their collateral requirements and end up undermining public confidence in the soundness of their balance sheet, potentially weakening the trust in money” (ibid.).

Going forward, central bank collateral policy will have to grapple with these conflicting goals. Central bank collateral policy will also have to be integrated with the broader policy shift toward macroprudential policies (Allen (2013)). Their collateral policy will be important, not only for short-term liquidity policy, but also for the longer-term development of core funding markets. Central banks will have to decide which funding markets are systemic and how far they will accommodate the endogenous growth of shadow bank liabilities in these markets.

8. Policy challenges

Despite a temporary slowdown in global shadow banking activities after the financial crisis, they remain large and growing. The sharp growth in investment funds that offer instant redemption, while investing in relatively illiquid assets, has recently led to concerns among regulators of potential market illiquidity (BoE (2014)). Governor Tarullo (2015) from the Federal Reserve Board has urged policymakers to review shadow banking activities “that pose significant risk of rapid investor flight during stress periods.” And Constâncio (2015) from the ECB has confirmed that “we are now witness to the emergence of a shadow banking sector that is also vulnerable to runs.” This new financial landscape raises some important policy challenges for central banks, especially how far they should go in accommodating the potential demand for safe assets from the shadow banking sector.

Johnson and Santor (2013) from the Bank of Canada argue that central bank liquidity support should be permanently available and that the traditional lender-of-last-resort function should be expanded to include support of core funding markets, “with the central bank being a “market maker” of last resort if necessary” (Johnson and Santor (2013)). Carlson et al. (2014) go further and argue that central banks should act aggressively as market makers of first resort in a systemic liquidity crisis, since it would be inefficient for individual entities to self-insure for a systemic liquidity event.

But if market participants are ill informed and under “the illusion of market liquidity” (BoE (2014)), liquidity will be underpriced based on the expectation of the central bank backstopping the system. It would then be fully rational for individual asset managers to operate on the premise that they can “dump their collateral and get out of town before things get ugly” (Tarullo (2013)). But should we not expect more from market participants that invest in relatively illiquid assets and offer instant redemption? Former Undersecretary of Domestic Finance (U.S. Treasury) Mary Miller (Miller (2014)) notes that it’s time to take a closer look at the asset management industry and check whether their broad spectrum of products all meet the test of liquidity.16

We need to find a way to curb private liquidity creation before central banks can become market makers of any resort. Limiting the growth of the shadow banking system is one key element in this new balance. As Borio (2013) notes: “The Achilles heel of the international monetary and financial system is not so much the risk of a structural excess demand for safe assets,” but rather the “excess elasticity” of the same system, i.e., the inability of policy regimes in place – monetary, prudential and fiscal – to prevent successive financial boom and bust cycles. Countercyclical haircuts and margins for securities financing transactions could become part of the new macroprudential toolkit needed to lean against this spontaneous credit creation in the shadow banking system.17

Better reporting of securities financing transactions [E.U. (2014)] will help in transparency, but will not in itself dampen the procyclicality of the shadow banking system. As long as the underlying incentives are strongly supportive of continued growth

---

16 IMF (2015) discusses the potential risks stemming from the asset management industry.
17 Although the FSB proposals for mandatory minimum haircuts have recently been watered down, and will now only be applied to a small subset of the shadow banking universe.
Shadow banking: policy challenges for central banks

in nonbank credit, in large part due to low-risk weights and the preferential status of collateral-based credit transactions, the reporting requirement may well be in vain [Perotti (2013)]. Sheila Bair (2013) is blunter when she notes: “Repos among financial institutions are treated as extremely low risk, even though excessive reliance on repo funding almost brought our system down. How dumb is that?”

Central banks will, therefore, continue to be under pressure “to stop fire sales and create outside liquidity” in a systemic crisis [Perotti (2012)] unless the incentives are changed. The recent FSB roadmap for shadow banking includes important policy proposals that could make a difference and provide the basis for a new form of macroprudential market regulation [FSB (2014b), Tarullo (2015)]. Cash buffer requirements and international minimum standards for fund redemption fees are other policy tools that are currently being considered [Constâncio (2015)].

As Minsky (1985) noted long ago, a flexible central bank liquidity policy should be combined with tough regulatory measures both before and after the crisis: “Clearly, central bank lender-of-last-resort interventions must lead to legislated or administered changes that favor hedge financing and ... the central bank should continuously ‘lean against’ the use of speculative and Ponzi financing” (ibid).

Central bank liquidity support for core funding markets should, therefore, be conditioned on meaningful structural reforms that can bring better balance between the shadow banking sector’s ability to self-insure and central banks’ capacity and willingness to provide backup liquidity. “The idea that a huge expansion even of a reformed financial system would bring great global benefit is doubtful” [Wolf (2013)] and “even right-wing voices now think it makes sense to restrict the size and behaviour of banks” [Turner (2013)].

We need to establish a sound system of credit creation reflecting the real economy’s need for finance. The current shadow banking system, backed by highly volatile collateral values, has made our whole financial system more fragile. “If credit creation left to itself goes beyond optimal levels, constraining it may be beneficial” [Turner (2013)].

Central banks should be especially concerned about providing liquidity to core financial markets without meaningful structural reform. A judicious review of the robustness of core funding markets is at least needed before central banks commit fully to the new role of market maker of last resort. Until it can be shown that these financial markets are reasonably able to stand on their own without central bank support in a crisis, authorities should insist on further reforms.19 It would indeed be ironic if central banks declared victory in the fight against too-big-to-fail institutions, just to end up bankrolling core funding markets.20

---

18 Note that Minsky here anticipated the recent macroprudential policy trend of “leaning against the wind” by some 30 years!

19 For example, the Federal Reserve Bank of New York has long been urging structural changes in the tri-party repo market [Stein (2013b)] and the Federal Reserve System has been equally vocal in its calls for reform of the money-market industry [Federal Reserve (2013)]. See also Tarullo (2013, 2015) for a specific proposal for reform in the shadow banking system, including minimum haircuts and OTC margins. See Fischer (2015) for stock take of progress on financial reforms in the nonbank financial sector.

20 Thomas Baxter, General Counsel of the New York Fed, has observed that broad-based liquidity support, such as the Primary Dealer Credit Facility during the crisis, could be permitted in the future as a form of “macroprudential” policy, while institution-specific liquidity support, such as the support for AIG, would be prohibited according to the new Dodd-Frank law [Baxter (2013)].
industry and financial stability,” International Monetary Fund
Monetary Fund, June 10
Johnson, G., and E. Santor, 2013, “Central bank liquidity provision and core
funding markets,” in Heath, A., M. Lilley and M. Manning (eds.), Liquidity and
funding markets, Reserve Bank of Australia
discussion paper no. 3, Bremen, Germany: Faculty of Business Studies and
Economics, University of Bremen
policies to combat the financial crisis,” Jackson Hole Symposium, Federal Reserve
Bank of Kansas City
introduction,” Bank of England Quarterly Bulletin Q1
Mehrling, P., 2014, “Why central banking should be re-imagined,” in Re-thinking the
lender of last resort, BIS papers no. 79, Bank for International Settlements
Miler, M., 2015, Remarks at Brookings Institute seminar on asset management,
financial stability and economic growth, Washington, D.C., January 9
Minsky, H. P., 1982. Can “it” happen again?” Essays on instability and finance,
Armonk, NY: M. E. Sharpe
Economics Institute of Bard College
Moe, T. G., 2012, “Shadow banking and the limits of central bank liquidity support:
how to achieve a better balance between global and official liquidity,” working paper
no. 712, Annandale-on-Hudson, NY: The Levy Economics Institute of Bard College
Murphy, D., 2013, “OTC derivatives, bilateral trading and central clearing: an
introduction to regulatory policy, market impact and systemic risk,” Palgrave McMillan
Perotti, E., 2013, “The roots of shadow banking,” CEPR policy insight no. 69,
December
Pozsar, Z., 2015, “A macro view of shadow banking: levered bets and wholesale
funding in the context of secular stagnation,” INET WP
during the recent market turbulence,” March 6
Simmons, E. C., 1947, “The relative liquidity of money and other things,” American
Economic Review 37:2, 308-311
Singh, M., 2013a, “The changing collateral space,” IMF working paper WP/13/25,
International Monetary Fund
Singh, M., 2013b, “The economics of shadow banking,” in Heath, A., M. Lilley and
M. Manning (eds.), Liquidity and funding markets, Reserve Bank of Australia
Stein, J. C., 2013a, “Liquidity regulation and central banking,” presented at the
“Finding the right balance” 2013 Credit Markets Symposium sponsored by the
Federal Reserve Bank of Richmond, Charlotte, North Carolina, April 19
Stein, J. C., 2013b, “The fire-sales problem and securities financing transactions,
presented at the Federal Reserve Bank of New York Workshop on “Fire sales as a
driver of systemic risk in tri-party repo and other secured funding markets,” New
York, October 4
Stein, J. C., 2014, Comments on “Market tantrums and monetary policy;” at the
2014 U.S. Monetary Policy Forum, New York, February 28
Tarullo, D. K., 2013, “Shadow banking and systemic risk regulation,” remarks at the
American for Financial Reform and Economic Policy Institute conference, November
22
Tarullo, D. K., 2015, “Advancing macroprudential policy objectives,” Speech at the
4th OFR/FSOC conference on “Evaluating macroprudential tools: complementarities
and conflicts,” Arlington, 30 January
26
Thornton, H., 1802, An inquiry into the nature and effects of the paper credit of
Great Britain, London: George Allen and Unwin
Cass Business School, 14 March
at South African Reserve Bank, 2 November
Turner, A., 2013, “Too much debt, financial system stability and wider economic
impacts,” speech at Chicago Federal Reserve Bank International Banking
Conference, November 7
system: last resort lending, market-making, and capital,” remarks at the Bank of
Japan 2009 International Conference, Tokyo, Japan, May 27-28
Center on Fiscal & Monetary Policy, Brookings
Financial Times, October 29
Zorn, L., and A. Garcia, 2011, “Central bank collateral policy: insights from recent
experience,” Bank of Canada Review, Spring
The consequences of exit from non-conventional monetary policy

Philip Turner
Deputy Head, Monetary and Economic Department, Bank for International Settlements

Abstract
This paper examines the financial and macroeconomic consequences of changes in central bank balance sheets. Large-scale purchases of bonds tend to drive down long-term interest rates. But developments in global markets, which shape the world long-term rate, exert some powerful constraints. Exchange rate effects can also be significant. The prospect of divergent balance sheet policies (balance sheet normalization by the Federal Reserve but further large purchases by the European Central Bank (ECB) and the Bank of Japan) is pushing the dollar up. Moving balance sheets toward more normal levels is important in order to preserve policy flexibility for the future but will present central banks with formidable challenges. This task will require cooperation with treasuries without surrendering monetary policy independence. As central banks pragmatically monitor market resilience, the financial dominance trap is to be avoided.

1 I would like to thank Morten Bech, Claudio Borio, Benjamin Friedman, Blaise Gadanecz, Boris Hofmann, André Icard, Enisse Kharroubi, Paul Mazen, Richhild Moessner, Hyeon Song Shin, Bruno Tissot and Fabrizio Zampolli for their comments and suggestions on earlier drafts. I am very grateful to Sonja Fritz and Jhuvesh Sobrun for their efficient help with this paper. The views expressed in this paper are those of the author and do not necessarily represent those of the BIS.
1. Introduction
One legacy of the monetary policies pursued since the financial crisis is that central banks in most advanced economies now have exceptionally large balance sheets. And commercial bank reserves (which with currency is often termed the “monetary base” or “high-powered money”) have risen by several multiples. Neither development is a cause for panic – there is no mechanical link between balance sheet expansion and “bad” outcomes, such as runaway inflation or dangerous credit expansion.

Nevertheless, these policies have made the exit challenge faced by central banks more complex, with no consensus on the “new normal” for the balance sheet of central banks. This paper argues that the crisis has forced a critical examination of some widely held beliefs about the division of labor between different agencies of government in implementing macroeconomic policies. Central banks have become more dependent on what their government decides – on fiscal policy, on government financing choices and on regulations requiring banks and other financial firms to hold government bonds. The exit will succeed only if central banks remain free of fiscal dominance. Central banks also have to pay greater attention to bond market reactions – but must not become captive to market expectations. They must remain free of what Hervé Hannoun has called financial dominance.

The main focus of this article is on the interlinkages between the balance sheets of a central bank, of banks and of government. The next section of the paper argues that the new classical macroeconomics, which gained increasing sway from the late 1980s, led to an unfortunate neglect of the financial and macroeconomic consequences of changes in central bank balance sheets. The following section explains how any analysis of such effects must take into consideration the fact that government financing choices can shape a central bank’s balance sheet, and thus influence the monetary implications. Section 4 reviews how many central banks have used their balance sheets to influence the long-term interest rate. This can create conflicts with the U.S. treasury’s debt management strategy. The empirical evidence is that the Federal Reserve has been able to drive down the yield on 10-year U.S. treasuries. Section 5 considers the international dimension, arguing that globalization has in effect created a “world” long-term interest rate – under the control of no single central bank. The term premium in all core bond markets has been driven down to levels that are unsustainably low. This, and the substantial accumulation of interest rate risk on the balance sheets of financial firms, will confront policy-makers with many difficult dilemmas once the exit from extraordinary monetary accommodation begins.

2. Three “dogmas” about monetary quantities
The sheer size of central bank balance sheets worldwide will require a major effort by economists to understand the financial and macroeconomic consequences of changes in central bank balance sheets. Quantity variables used to have pride of place in monetary analysis. Before the mid-1980s, the key elements in mainstream analyses of monetary policy were, in addition to the policy rate, central bank market operations in government securities markets, the liquidity of financial markets and market expectations of the future. The liquidity of the balance sheets of the banks was viewed as affecting their lending decisions. How central bank purchases or sales changed the market prices of financial assets depended on the substitutability between money and other assets in investors’ portfolios. Many central banks imposed “liquid” or “reserve asset ratios” on banks. Such ratios were used not only for monetary control, but also for influencing bank lending and for keeping the banking system safe. In short, portfolio rebalancing effects were seen as highly relevant for monetary policy. Several economists in the 1950s and 1960s aimed at providing rigorous theoretical foundations for such effects. Culbertson (1957) and Modigliani and Sutch (1966, 1967), among others, highlighted the existence of market segmentation and imperfect substitution between different maturities, proposing the preferred habitat theory as a possible explanation. Despite extensive econometric research, however, there was no agreement about the size, or about the stability over time, of portfolio rebalancing effects.

But this quantity-focused theory of central banking was progressively undermined by the rise of rational expectations models associated with the new classical macroeconomics. Applied to monetary economics, this developed into the New Keynesian model. This model took account of the macroeconomic consequences of imperfections in goods and labor markets. But it assumed perfect financial markets [Hahn and Solow

---

2 Before 1971, for instance, the U.K. clearing banks were required to hold 28% of their deposits in liquid assets – short-term treasury bills counted as a liquid asset but long-term bonds did not.
The point of departure of these models was the rational intertemporal choices of a single representative agent who has perfect foresight for each future state of nature (or who could trade in complete markets). A central bank had only to set the short-term rate and markets would determine the shape of the yield curve according to expectations of future short rates and of macroeconomic prospects. Various “irrelevance theorems” were devised for government budgets and central bank balance sheets. Households would understand the implications for their future taxes of decisions about government’s or central bank’s balance sheet, and would react in ways that neutralize the putative effect of such official action (Ricardian equivalence).

The beauty of such an intertemporal approach to economic policy is that it focuses attention on the fact that policy action taken today for short-term benefit may have longer-term consequences. One such consequence is that public sector balance sheets change in ways that can constrain future policies. Another is that what the authorities have done today shapes the expectations of private agents, influencing their behavior. This intertemporal insight has an important bearing on current monetary policy debates. But the key drawback with the specific form such models took was that coordination failure among private agents was ruled out by assumption. Liquidity constraints and many other interesting macroeconomic questions were, in effect, sidestepped. Several articles in BIS (2012) explore these issues in more detail.

The New Keynesian perspective incorporating rational expectations and perfect asset substitutability also had a number of convenient implications for policymakers. It shaped what has been called the pre-crisis “doctrine” of monetary policy, and therefore was partly responsible for the severity of the recent crisis (Aglietta (2013)). Note three “dogmas” that are of interest for the purpose of this paper:

- Open market operations in government bond markets (or in foreign exchange markets) do not change relative prices: Ricardian equivalence applies to central banks; any purchase or sale of particular assets would lead only to offsetting changes in private demands, with no impact on prices. One corollary of this is that government debt management (that is, the relative supply of short-dated and long-dated bonds by the Treasury) can be separated from monetary policy.

- The central bank short-term policy rate is the unique instrument of monetary policy aimed at macroeconomic objectives: the impact of policies on other core financial market prices, such as the term premium in the long-term interest rate, was neglected. Developments in monetary quantities (e.g., M2 and bank reserves), seen as reacting endogenously to policy rate decisions, had little or no influence on policy. And the Taylor Rule linked only the short-term rate to macroeconomic developments.

- The “liquidity” of the balance sheets of commercial banks is irrelevant: if adequate capital standards are in place to ensure the viability of a bank, there is no additional need for bank regulators to worry about the liquidity of banks because a sound bank could borrow readily in interbank markets to meet any “temporary” liquidity squeeze. Hence, the failure of international regulators in the 1980s to develop common measures of the overall liquidity of a bank (and the decline in liquid asset ratios) seemed unimportant.

Not everybody believed in these dogmas of course. Quantity measures related to balance sheets continued to guide policy considerations at the Bundesbank long after they had been abandoned by other central banks. Central banks in many emerging markets retained quantity-based policies and enforced liquidity rules on banks. The size and nature of a central bank’s balance sheet was key to Bernanke’s analysis, both of the 1930s depression and of Japan’s stagnation in the early 2000s (Bernanke (2013)). Greenspan puzzled before the crisis about why the long-term rate had proved impervious to rising short-term rates.

In any event, all three dogmas have been shown by recent events to be false. Central bank balance sheets matter. Large-scale central bank purchases of bonds (and other assets) have lowered long-term interest rates, leading economists to re-examine the portfolio rebalancing effects that the new classical school had dismissed. The neat separation between central bank open market operations and government debt management has been blurred. And banks now pay much closer attention to the liquidity of their balance sheets (with bank regulation in this area having been strongly reinforced since the crisis).

Equally, the scale of balance sheet measures taken by central banks actually reinforces the fundamental logic behind the new
classical theories. An intertemporal perspective — a key insight of rational expectations — has become even more necessary. Because of the substantial lengthening in the maturity of central bank assets, the decisions taken during this crisis will have more long-lasting (and therefore more uncertain) effects than if policy action had been limited to short-term interest rates or short-dated paper.

3. The analysis of a central bank’s balance sheet
Many debates about the impact of central bank balance sheet policies on the real economy are confused by the failure to recognize three fundamental elements:

- Any change in the total assets of a central bank must be accompanied by an equal change in its total liabilities. Many analyses of the economic effects of policy-driven changes in central bank balance sheets consider only the asset side (e.g., purchases of mortgage bonds lower financing costs); but the associated changes in central bank liabilities imply changes in the assets of other sectors, which may also affect aggregate demand.

- Even if they have not done so up to now, a government could decide to step in so that the change in central bank liabilities is vis-à-vis the government, not the private sector. To do so, it could borrow in capital markets in order to increase its deposits with the central bank.

- Much of the increase in the liabilities of a central bank — particularly in a crisis — will usually be assets of the commercial banks and could therefore influence bank lending.

Table 1 illustrates these points with simple balance sheets.

During the past decade, the size of central bank balance sheets almost everywhere has grown more than anyone would have thought likely before the crisis. Although central banks have acquired quite different assets, all have had some impact on the yield on government bonds in their jurisdiction. EME (emerging market economy) central banks (and Switzerland) have mainly bought foreign assets, with an impact on the yields of the major government bond benchmarks [Pradhan (2014)]. Before their 2015 bond buying program, the ECB had concentrated on medium-term loans to banks: this has led banks in countries hardest hit by the crisis to increase the purchase of domestic government bonds (see Table 2). In Italy and Spain, banks’ holdings of bonds of their own sovereign show stronger home bias than in France and Germany [ESRB (2015)]. Valla (2014) says central banks thus supported a profitable carry trade, which helped bank recapitalization. The Bank of England, the Bank of Japan and the Federal Reserve have bought government bonds and other bonds directly. Policies aimed at lowering bond yields in the domestic market will usually cause reallocations in international bond portfolios, and so are likely to influence exchange rates (discussed further in section 5).

Equally important, how central banks have altered the size and the nature of their balance sheets in response to macroeconomic and financial conditions can itself shape private sector expectations about what the central bank will do in the future. Sometimes, a central bank will even make an explicit commitment of its future responsiveness. At several junctures of the recent crisis, it has been the commitment of central banks to buy specific assets — even before actual purchases — that has eased...
liquidity constraints of holders of these assets. Investors will seek profits by buying ahead of a central bank. For these reasons, there will be no simple and mechanical link between the central bank balance sheet and financial or macroeconomic variables.

In most cases, central bank asset purchases (or loans) were in effect financed in large part by the increase in commercial bank reserves with the central bank – an expansion of the monetary base. But there is nothing intrinsic or inevitable about the link between central bank asset purchases and monetary expansion. A government could have prevented any increase in liabilities to the banks either by injecting equity capital into the central bank or by increasing its own deposits with the central bank. If a government were to finance this by issuing bonds, however, it would tend to drive up the benchmark long-term rate. Other reasons for government reluctance to increase its own borrowing included electoral sensitivities (“borrowing to help big banks”), the assessment methodologies of credit rating agencies and the difficulty of securing rapid parliamentary approval. In any event, no government sought to finance the bulk of the expansion in central bank balance sheets.

Nor is there any simple link between the monetary base and bank credit to the private sector: the old bank multiplier model of textbooks has long been abandoned. Nevertheless, the fact that the balance sheets of commercial banks have tended to become more liquid as central bank liabilities have increased could at some point stimulate credit expansion. The right-hand panel of Figure 1 shows the case of the U.S. – exceptional central bank asset purchases drove bank reserves (i.e., commercial bank deposits with the Federal Reserve) to over 25% of total bank deposits. But part of this increase in reserves may be permanent, reflecting a stronger bank demand for liquid assets. The crisis did indeed teach banks in the advanced economies that they need to hold more liquid assets even in normal times. New international bank regulation is reinforcing this orientation. It is, however, too early to tell what banks’ new liquidity preference will be in the medium term. Moreover, there is no consensus on the impact of more liquid balance sheets on future bank behavior. One argument is that banks will lend more mainly when the prospective returns are attractive, with liquidity acting only as a constraint (that is, 3 In emerging markets, by contrast, many central banks did not want their commercial banks to become more liquid as a result of central bank purchases of foreign exchange and took positive steps to counteract this [Mohanty and Turner (2006)]. Nevertheless, increased bank holdings of government bonds in EMEs do appear to lead to an expansion in bank credit to the private sector [Gadanecz et al. (2014)].
liquidity has a binary, on/off nature). The counterargument is that liquidity effects may be continuous. After all, the larger the stock of liquid assets, the further tail risk of sudden illiquidity and bank runs is reduced and so the more lending can be increased.

Could this bloated level of bank reserves create difficulties for central banks when they want to normalize monetary policy? There are very divergent views on this issue. Some argue “no.” An independent central bank’s control of short-term rates is enough to influence the whole structure of interest rates. The central bank can control how quickly banks run down their reserves by raising interest rates paid on reserves irrespective of the size of the central bank’s balance sheet. Others argue “yes.” Central banks have little experience of dealing with such massive reserves. Marvin Goodfriend (2014) has argued that quantitative easing (QE) is a “bond market carry trade” that could land the central bank with sizable losses if short-term rates have to rise sharply. He makes the political (not economic) argument that this might jeopardize the operational credibility of monetary policy. Resolution of this disagreement probably partly depends on the nature of the shocks to hit the economy when central bank balance sheets are still large.

A further complication is that there is little consensus about the new optimum level of banks’ liquid assets in the post-crisis world. Because several new regulations require banks and other financial institutions to have more liquid balance sheets than before the recent financial crisis, the central bank may have to leave more “liquidity in the financial system on a permanent basis” [Gagnon and Sack (2014)]. Getting this right when markets are unsettled or when monetary policy frameworks are in flux may not be easy. Some have argued that, in certain circumstances, the authorities in advanced countries may want to support their exit policy by imposing quantitative rules. After such a prolonged period of monetary accommodation, the task of managing what Blinder has called the “veritable mountain” of excess reserves will not be simple. It may require a new operating regime [see Gagnon and Sack (2014)].

4 The Journal of Financial Perspectives

4. The long-term interest rate as key indicator of monetary policy

The long-term interest rate matters greatly for the macroeconomic effects of monetary policy. It affects the decisions (and the balance sheets) of all borrowers and lenders who have long-term loan or debt contracts. If changes in the long-term interest rate could be fully explained by changes in expected future short-term rates, central banks would need to worry about only the current short rate (which it controls) and expected future short rates (which it might hope to influence by forward guidance). In practice, however, the long rate can move because of a central bank buying longer-dated paper and driving down the term premium (see Figure 2). This effect recalls James Tobin’s classic work on portfolio rebalancing mechanisms in the transmission of monetary policy. Recent research has revived the old preferred-habitat models of the 1950s and the 1960s [Vayanos and Vila (2009)]. The impact of central bank balance sheet policies in shaping the term premium can be crucial. Tobin’s work has also found a recent echo in the finding of Gertler and Karadi (2013) that the changes in the term premium have come to play a significant role in monetary policy transmission in the U.S. Economists are once again recognizing not only that aggregate demand depends on the long rate (not just the short rate), which was once standard in macroeconomic models [Reifschneider et al. (1999)], but also that the term premium can be influenced by monetary policy.

Uncertainty about how and when a change in the policy rate would affect the long-term rate means that open-market operations in long-term securities could improve the chances of timing countercyclical monetary policy correctly. Equally, shocks to the long rate could frustrate the monetary policy stance the central bank is seeking to impose through its policy rate. In

---

4 Aglietta (2014) includes required reserve ratios in his macroprudential policy instruments that should accompany monetary policy. Siegel (2013) recently argued that the exit strategy could be better managed if the Federal Reserve were to impose a 15% reserve ratio on banks (the Federal Reserve’s “third policy tool” was the expression he used). In a similar vein, Goodhart (2013) argues that banks could be required to hold a higher proportion of their balance sheet in liquid assets (“financial repression” was the expression he used). See also Feldstein (2014).

5 The Federal Reserve began paying interest on such reserves in October 2008. The decision on the rate of interest to be applied to bank reserves is the responsibility of the Board of Governors, not the FOMC. As Blinder (2010) explains, this is significant because Chairman Bernanke had himself said that this interest rate—not the federal funds rate—could be the more reliable guide to the stance of monetary policy in the early stages of exit. In addition, the interest rate on full allotment reverse repos is also likely to be important.

6 Keynes argued along these lines in both the Treatise on Money and the General Theory, and James Tobin and Milton Friedman echoed this analysis [Turner (2013)]. Brainard’s (1967) uncertainty principle echoes this thinking.
some circumstances, then, sales or purchases of government debt would be preferable to relying only on the policy rate. For example, the size of the adjustment in the policy rate needed to have the desired impact quickly on the long-term rate could be too disruptive for borrowers with short-term debts. Banks, whose funding costs are closely linked to the policy rate, could be especially vulnerable to sudden large movements in the policy rate. And long-term rates might well, with a lag, actually overreact, forcing a reversal of the policy rate increase.

It is an open question whether the size and nature of the central bank’s asset portfolio would become a second instrument of monetary policy – to shape the yield curve and to influence bank credit. Historically – that is, before the 1980s when the New Keynesian model led many to downplay asset portfolio rebalancing effects – the central bank’s balance sheet had been viewed as important for monetary policy. Ben Friedman (2014) recently argued that the central bank’s balance sheet is likely to become part of the standard toolkit of monetary policy in normal times. Echoing Tobin’s portfolio balance theory, he stresses that “the central bank’s ability to choose what quantity of assets to purchase (with consequent increases in its liabilities) is not merely an artifact of the policy interest rate being at the lower bound.” Others have argued that, if inflation around zero were to become the new norm, the zero lower bound (ZLB) problem would resurface, putting balance sheet policy back on the agenda.

The long-term interest rate and the slope of the yield curve are also key for financial stability. In the absence of sovereign default risk, the long-term interest rate on government bonds defines the credit risk-free maturity transformation over time. It provides the basic discount rate, and is thus central to the pricing of all long-term assets. When the long-term rate is “too low,” the prices of long-term assets are “too high.” And a higher market value of the assets that potential borrowers own allows them to pledge more collateral in order to get new loans. A flatter yield curve reduces banks’ earnings from the maturity transformation so that higher short-term rates induce banks to reduce the supply of credit [Adrian and Shin (2011)]. For all these reasons, the management of the yield curve can be a key element of financial stability policy.

4.1 Link with government debt management
The central bank controls the overnight rate but does not have sole jurisdiction over national policies aimed at the long-term rate. This is because government debt management policies affect the maturity of government debt held by outside investors – that is, the private sector and foreign official investors – just as much as central bank market operations. Both the government and the central bank can, through their transactions, alter the portfolios of assets held by the market. Given the imperfect substitutability of bonds with different maturities (contrary to the new classical view), this allows them to influence key interest rates at different points along the yield curve. Bernanke (2002) did advocate, if the federal funds rate were to fall to zero, the Federal Reserve “announcing explicit ceiling for yields on longer-maturity treasury debt (say, bonds maturing within the next two years).” But note that the logic of portfolio rebalancing effects is independent of the level of short-term rates: it is not a special case applying only at the ZLB [Goodhart (2012) and Ellison and Tischbirek (2013)]. While it is true that asset substitutability is lower in crisis times (so portfolio rebalancing effects are stronger), the evidence cited in the next section suggests portfolio rebalancing effects are still significant in normal times.

There is nothing new in the observation that the authorities can influence the yield curve. Government debt management was a key part of monetary theory and practice from at least 1930 right up to the 1980s. Governments and central banks failed in the 1930s to understand their power. Recall that Keynes argued strongly that the British Government in 1930 was mistaken in its “sound money” policy of lengthening the maturity of gilts. Their policy inadvertently weakened the monetary policy expansion intended by the abandonment of the gold standard (which allowed short-term rates to fall) and by foreign exchange intervention designed to depreciate sterling [see Allen’s article in BIS (2012)]. He also urged the Federal Reserve to buy long-term treasuries, but such purchases began in earnest only during World War II [Tily (2010)]. During much of the post-war period, many central banks in Europe were closely involved in government debt management – but most were not independent.

From the mid-1980s, however, the view grew that giving central banks the dual mandate of both setting monetary policy and managing government debt created a conflict of interest. Trying

7 “Monetary policy determines the composition of the government portfolio,” was the first line of Wallace’s 1981 article expounding the logic of a Modigliani-Miller theorem for central bank market operations.
to keep debt service costs down (or even limiting the volatility of such costs) could conflict with the monetary policy need to adjust interest rates in the light of changing economic conditions (“fiscal dominance”). Even market perceptions of such a conflict could unsettle inflation expectations.

Government debt managers were, therefore, given a degree of independence and assigned clear objectives. They were normally expected to minimize anticipated costs over time (and avoid “spikes” in future repayments) subject to some risk tolerance limits. Two “separability principles” governed their interaction with central banks:

1. Central banks should not operate in the markets for long-dated government debt, but should limit their market operations to the bills market.
2. Government debt management should be guided by cost minimization mandates, and not by macroeconomic developments. Issuance of short-dated debt should be minimal.

The large-scale acquisition of long-dated government paper by central banks as a result of the financial crisis has obviously undermined this separation. The left panel of Figure 1, above, shows a dramatic rise in the average maturity of the Federal Reserve’s holdings of public debt, in effect taking duration risk out of the market. But note that the significant rise in the average maturity of U.S. Treasury issuance (dashed line) has worked in the opposite direction. Indeed, the published minutes of the Treasury Borrowing Advisory Committee reveal some interesting discussion on whether the Fed and the U.S. Treasury have been working at cross-purposes [Turner (2011)]. Larry Summers argued in a recent Brookings Panel discussion for closer coordination.

8 Coordination in the U.K. is more explicit. The U.K. Treasury has agreed to indemnify the Bank of England against possible losses from its QE program. The Bank of England in its May 2014 Inflation Report said it would liaise with the Debt Management Office when deciding any program of sales. The thorny issue of coordination between the central bank and the Treasury is reviewed in BIS (2012).

The prospect of much higher government debt to finance makes prudent debt managers want to lengthen the maturity of their issuance. Moreover, a temporary change to the yield curve induced by central bank action may lead the debt manager to alter its issuance policy to take advantage of what it might view as a temporary interest rate “distortion.” Or it may find it can move quickly to attain a pre-existing maturity-extending objective thanks to favorable market conditions created by the central bank. In any event, there is empirical evidence that the U.S. Treasury has lengthened the average maturity of its debt issuance when the fiscal deficit is high and issued shorter term when the spread of the 10-year yield over the federal funds rate is high. 9

5. The international dimension
5.1 The “world” long-term interest rate
The previous sections concentrated on domestic determinants of the long-term interest rate. But long-term interest rates are more and more set in global markets. International investors and debt issuers can shift their transactions from one currency to another. The yield on 10-year U.S. Treasuries is, of course, the global benchmark. But it is a benchmark that is driven by developments worldwide – and not only by U.S. growth or monetary policy. An idea of magnitudes might help to clarify this point. Detailed calculations by BIS economists suggest that aggregate borrowing in U.S. dollar bond markets by nonbanks outside the U.S. now exceeds U.S.$4 trillion – a fourfold rise since the start of 2000 [McCauley et al. (2015)]. Of this, U.S.$1.3 trillion comes from bond investors in the U.S., but U.S.$2.7 trillion comes from bond investors based overseas. It is this huge and growing volume of transactions between non-U.S. residents in dollar bond markets that has made dollar bond markets truly global. These strong global forces mean that the combination of higher U.S. short-term rates but large bond purchases in the euro area and in Japan could well result in low maturity of government debt do matter for monetary policy.


10 Evidence is provided in Table 2 in Blommestein and Turner in BIS (2012).
dollar long-term rates for a prolonged period – a revival of the Greenspan “conundrum!”

There is, of course, no unique way of measuring the “world” interest rate. But the estimate prepared by Mervyn King and David Low (based on advanced economy bond market data) is a good starting point (Figure 2). Movements in the yield on 10-year U.S. Treasuries – shown as a yellow line in the graph – dominate this “world” interest rate. The real-world long-term interest rate has been falling for more than a decade and is now close to zero. Panel B of Figure 2, based on calculations from Hördahl and Tristani (2014), shows that the decline in U.S. 10-year yields has been largely driven by a compression of the term premium – the reward for holding long-dated rather than short-dated bonds. It has, therefore, declined for reasons other than changes in expected future short rates. Note the very similar trend in the term premium in the euro area: see panel C of the figure. (French Treasuries are used in BIS calculations because French inflation-linked bonds have been issued for longer, and enjoy greater liquidity, than those of other euro area countries). If the term premium in global markets remains low once the normalization of U.S. monetary policy gets under way, then, all else equal, it would be appropriate, in the words of the President of the Federal Reserve Bank of New York, “to choose a more aggressive path of monetary policy normalization” (Dudley (2015)).

In any event, the idea from rational expectations theory that the term premium can be viewed as independent of central bank or government balance sheets has been discredited. A number of empirical studies done before the crisis had shown that shifts in the demand of large investors (e.g., foreign official demand for high-quality U.S. dollar debt and maturity arbitrage in U.S. dollar securities by European banks) did depress term premia in U.S. Treasuries [see, for example, Bertaut et al. (2011)].

In addition, new prudential regulations, mark-to-market accounting rules, actuarial conventions, etc., have induced financial intermediaries to hold a higher proportion of their assets in government bonds, driving down benchmark yields. A crisis-induced flight to quality has also been important.

Several event studies have shown that the announcements of large-scale central bank bond purchases in Japan, the

---

Figure 2: Long-term interest rates (%)
(a) Sum of inflation and real yield risk premia in the 10-year Treasury yield. These are calculated using the BIS term structure model. Sources: King and Low (©February 2014); Bloomberg; national data; BIS calculations.

---

11 Shin (2012) provides a comprehensive summary of these important mechanisms. European banks – not U.S. banks – held most of the special purpose vehicles holding U.S. mortgage-based securities. See also Borio and Disyatat (2011).
The consequences of exit from non-conventional monetary policy

U.K. and the U.S. have indeed lowered long-term rates. One general shortcoming of such studies, however, is the neglect of contemporaneous changes in Treasury issuance policy. In addition, such estimates are based on the difficult financial market conditions prevailing in the post-crisis period. Capital constraints on banks and other financial firms, worries about the creditworthiness of wholesale market counterparties and uncertainty about future regulations would all inhibit arbitrage by the private sector. The new classical model would not apply in such circumstances — but may again apply in normal conditions.

To address these issues, Chadha et al. (2013) focused on long-term rates over a pre-crisis period (that is, 1976 to 2008). They examined the determinants of 5-year forward 10-year yields, an interest rate that should be less influenced by the business cycle and monetary policy than the contemporary 10-year yield. The paper used expectations of future variables, not the current readings. Computations of the size and the maturity of U.S. Treasury debt held outside the Federal Reserve are reproduced in Figure 1. The main new finding of this paper was that shortening the average maturity of total outstanding federal debt held outside the Federal Reserve by one month lowers the long-term yield by 12–13 basis points. A one percentage point rise in long-horizon inflation expectations adds about one percentage point to the forward 10-year yield.12 Hence, shortening the maturity of government debt issuance does lower the long-term rate, provided inflation expectations do not change.13

A number of other recent empirical studies have also found that changes in the relative supply of bonds influences interest rates. Greenwood and Vayanos (2010) find that the relative supply of long-dated securities is positively related to the yield spreads and subsequent excess returns over short-term yields. In a similar vein (but looking from the demand side), Hanson and Stein (2012) find that commercial banks and primary dealers change the maturity of their government bond portfolios in response to changes in short-term interest rate expectations, thus affecting term premia. The preferred-habitat models of the 1960s are getting renewed attention.

Further research on how central bank purchases in major reserve currency countries have affected other government bond markets would be very useful. By way of simple illustration, Table 3 shows, for both a pre-crisis and a post-crisis period, that changes in the U.S. 10-year yield (Δ LTR (U.S.) in the table) have a greater correlation with yields in France, Germany and the U.K. than with changes in their local short-term rates (Δ STR). Japan is an exception, perhaps because it is hard to measure interest rate effects when rates have been so low for so long. It may also be a reflection of the very strong home bias of Japanese banks and other financial institutions. The events in the summer of 2013 seem to confirm this dependence. French, German and U.K. yields actually rose in line with movements in U.S. rates despite Bank of England and ECB forward guidance on 4 July 2013 seeking to assure investors that their policy rates would remain low. All yields then fell on the Federal Reserve’s “no tapering” announcement on 18 September. Equally, events in early 2015 — dominated by new policy announcements from the ECB — show that U.S. long-term rates can be pulled down by foreign forces (in this case, the sharp decline in euro area long-term rates following the announcement of large-scale government bond purchases by the ECB).

12 The estimates of the coefficients of other variables were very close to those found by Laubach (2009).
13 Using a very similar specification, Iwata and Fuada-Samikawa (2013) also found that shortening the maturity of government bond issuance in Japan pushed long-term interest rates up significantly.
5.2 The exchange rate
Other things equal, portfolio rebalancing effects mean that central bank purchases of domestic assets on a large scale (that is QE), driving down the yield on such assets, will lead to currency depreciation. Conversely, reversing QE will lead to currency appreciation. Figure 3 shows the balance sheets of the ECB, the Bank of England and the Bank of Japan relative to that of the Federal Reserve. During the past few years, the relative movements have been very large. And those shifts do appear to have influenced exchange rates.

At the time of writing (March 2015), the balance sheet policies of the world’s major central banks have begun to diverge in a new way. The Federal Reserve and the Bank of England, which expanded their balance sheets aggressively in earlier years, have now ended net asset purchases but the Bank of Japan is continuing to purchase Japanese Government Bonds (JGBs) on a massive scale. The ECB has also launched a major program. Expected future movements in the policy rate mirror this divergence. Markets expect the federal funds rate and the U.K.’s bank rate to rise during 2015. By contrast, euro and yen interest rates are expected to remain close to zero until end-2016.

One important consequence of these developments has been a sizable appreciation of the dollar (and to a lesser extent of sterling) and corresponding depreciations in the euro and the yen. The financial implications of any large and sustained dollar appreciation for non-dollar countries will be significant. International debts – bank loans, bonds, etc. – are still denominated predominantly in dollars. Partly for this reason, many non-U.S. central banks (especially in emerging markets) would be inclined to follow changes in the Federal Reserve’s funds rate. They may also raise rates to forestall too sharp a depreciation against the dollar.

6. Asset prices and interest rate risk
Lowering interest rates further out the yield curve reduced the funding costs of banks (which do not depend only on the policy rate) and helped lower bank lending rates (Illes et al. (2015)). Getting long-term rates down brought financial asset prices in the core economies back up to pre-crisis levels, even higher. And, with the arguable exception of the Lehman debacle, central banks recognized their lender-of-last-resort role for the banking system better than in the 1930s. Gambacorta et al. (2012) find that the expansion of central bank balance sheets increased real GDP. For all these reasons, there is little doubt that QE and other exceptional policies have “worked.”
There is, however, an intertemporal rejoinder. Will the exit from this extraordinary balance sheet position be handled well enough to avoid negative consequences in future years? There is a reassuring answer. The massive purchases of central banks have had wealth effects that should, in time, stimulate global demand. In addition, stronger asset prices should raise the value of potential collateral for new loans and, therefore, ease the borrowing constraints facing firms and households. Once stronger aggregate demand is assured, the central bank could readily unload the assets acquired during the crisis.

The problem with this reassuring answer is that the recent recession – now more than five years long – has lasted so long. Financial asset prices did get a considerable boost. Yet the hoped-for growth in real GDP that would have allowed central banks to scale back crisis-related asset purchases did not materialize. This disconnect between the rapid rise in asset prices and the persistent weakness of demand is worrying. Is this a bubble that could suddenly deflate? Or do forecasters underestimate the strength of real demand over the next couple of years?

Another worry is that global net interest rate exposures must have risen substantially since the crisis. Yields in all major bond markets have fallen to very low levels even as outstanding stocks have risen. Bond investors will lose once long-term rates increase. Much of this risk is in the banking system: sovereign exposures accounted for 19% of total banking book exposures of large international banks in mid-2012, compared with 11% at end-2008 (estimates given on page 44 of Turner (2013)).

Lower-rated corporations have also benefited from the negative or zero term premium in government debt markets, so credit risks have probably risen too. Furthermore, the link between U.S. yields and yields on EM bonds has increased substantially over the past decade, and EM bond issuance has risen (Pradhan (2014)).

The vastly increased volume of bonds outstanding, some held in leveraged portfolios, means that bond market volatility will rise much more when market sentiment changes than it did in the past when outstanding stocks of bonds were much lower. The turbulence also illustrates the dominance of U.S. Treasuries. A substantial rise in U.S. long-term rates took place without any change in the policy rate in the U.S.14 Such a strong and global market reaction suggests some sudden unwinding of leveraged positions and powerful contagion across markets.

It is difficult to know what lies ahead. Central banks in the advanced economies are not comfortable with the size and structure of their balance sheets. From September 2009, governors of the major central banks (including Messrs Bernanke and Trichet) expressed the hope that they would soon be able to begin their “exit” from unconventional policies. But such hopes were dashed by the deepening euro crisis from mid-2010. Not only have central bank balance sheets further expanded but – equally important – the maturity of their assets has become much longer.15

Since their liabilities have remained of very short maturity (typically bank reserves), central banks have a growing maturity mismatch. A sizable term spread gives the central bank a positive running yield: this has boosted its profits typically remitted to the Treasury, often creating a favorable impression with parliament. But higher short-term rates could at some point lead to central bank losses. This has no fundamental significance. The central bank does not face the financing constraint in its own currency that a private agent faces: it can print money. Likewise, a government can raise taxes. But losses could have political consequences that might weaken central bank independence. [And in some circumstances – for instance, a disruptive shock to inflation expectations – macroeconomic policy would face more intractable dilemmas: see Sims (2013)].

There will be many years ahead when central banks will have government and other bonds on their balance sheets. The accumulation of such substantial holdings was justified mainly by the crisis that confronted central banks. It is difficult to know at present what the new “normal” size of such holdings will be. How quickly central banks reduce their bond portfolio will depend on (unknown) macroeconomic or financial developments over the next several years.

14 This was quite unlike the 1994 bond market sell-off, when successive rises in U.S. long-term yields were driven by changes in expectations of future policy rates (Adrian and Fleming (2013)).

15 Some use the term “qualitative easing” to cover central bank purchase of longer-dated or higher-risk paper.
Could central bank sales or purchases of government bonds become viewed as a second policy instrument once monetary policy begins to be tightened? Actively selling assets in secondary markets ("quantitative tightening") could well moderate any increase in the policy rate. Both the Federal Reserve and the Bank of England have explicitly recognized this trade-off. The May 2014 Inflation Report of the Bank of England noted that "any reduction in the stock of purchased assets is likely to be associated with a lower path of Bank Rate." The FOMC minutes in April 2011 reveal that participants noted that "for any given degree of policy tightening, more-gradual sales that commenced later in the normalization process would allow for an earlier increase of the federal funds rate target from its effective lower bound than would be the case if asset sales commenced earlier and at a more rapid pace."

But one practical difficulty is that it is impossible to quantify how bond markets would react to central bank sales. Using estimates based on past experience of the policies that change the volume and maturity of government debt to be sold (such as those previously mentioned) fail to take account of signaling effects. News of central bank selling, even on a modest scale, could send markets a signal that is more powerful than the actual sales — "They are testing the water for further, larger sales," Financial markets know not only the size of central bank bond holdings, but also the great strategic power of central banks as non-commercial players.16 The hypersensitivity of markets to guesses about future central bank sales was very well illustrated over the summer of 2013. The mention by Chairman Bernanke of what should have been obvious — that at some point the Fed would reduce the pace of its purchases — wreaked havoc in global bond markets, even with the very clear commitment of the Fed to keep short-term rates close to zero for a considerable time. The size and spread of this market adjustment suggest that many investors had highly leveraged positions.

For these reasons, central banks have underlined their caution about using their balance sheets as readily as they might adjust the policy rate “because of uncertainty about how non-conventional tools will work or because of the potential
costs associated with the use of such tools in terms of market functioning and the risks of future financial instability” [Dudley (2013)]. Both the Federal Reserve and the Bank of England have ceased new asset purchases on a net basis, but continue to reinvest the proceeds of maturing debt. When they ended asset purchases, both central banks underlined their determination to keep the policy rate at near zero for a long time. And both have indicated that the initial normalization steps will take the form of policy rate increases, rather than sales of assets. The Bank of England has explained its logic of sequencing by indicating "it is likely to defer sales of assets at least until the Bank Rate has reached a level from which it could be cut materially, were more stimulus to be required."

There is great uncertainty about when and how fast central banks will reduce their holdings of bonds beyond the next couple of years, once policy rates are well clear of zero. The option of just allowing bonds to mature — apparently the easy option because it avoids contentious decisions about actual sales — would not be a neutral policy choice. It would mean central bank balance sheets remaining large beyond 2020. And it would also mean that the timing of shrinking — which would have effects on financial markets and the macroeconomy — would depend only on the pattern of past purchases and be quite independent of future economic conditions. It could even continue into the next recession.

No central bank would want to rule out active steps to shrink their balance sheets irrespective of circumstances. As the former Governor of the Bank of England has noted [King (2012)], the central bank must keep the ability to sell the government bonds on its balance sheet if needed, to maintain control of inflation, monetary conditions and the supply of credit.

The policy issues faced by a central bank set on deliberately scaling back its balance sheet have many dimensions. Should the exit be discretionary or rules based? Should the rule be quantity-based or price-based?

Even when the central bank seeks to retain complete discretion, it would have to find a way to communicate its near-term plans. It would have to clarify to the fiscal authority its future actions in government debt markets. No central bank would want to be seen as gratuitously making the financing conditions of long-term

---

16 As ElErian (2012) puts it, “in game theoretic terms, central banks are non-commercial players ... [they have] a printing press ... and the structural patience that far exceeds the ability of any other participant to remain in the trade.”
government debt more difficult. Clarifying its future intentions could also help to stabilize market expectations. But any rule would quickly lose credibility if too inflexible. Balancing the need to stabilize market expectations with the need to adapt to an uncertain and changing reality makes it more likely that any announcement or commitment by a central bank will be limited in time and in scope.

In making a commitment, central banks have several general options. A number of quantity-based rules can be considered. For example, the central bank could commit to cease purchasing new debt as current bond holdings mature. Alternatively, it could announce a fixed amount of sales to be executed over a given time frame. Any announced plan of sales could also come with specific conditions. For example, sales could be made contingent on the borrowing needs of the government (e.g., by preventing sales when deficits are above a certain level). This would, however, have to be communicated in ways that do not raise a suspicion of fiscal dominance. The former Governor of the Bank of Japan [see Shirakawa (2013)] has repeatedly warned central banks to avoid the trap of fiscal dominance.

Any announced plan of sales could also come with option. A number of quantity-based rules can be considered. For example, the central bank could commit to cease purchasing new debt as current bond holdings mature. Alternatively, it could announce a fixed amount of sales to be executed over a given time frame. Any announced plan of sales could also come with specific conditions. For example, sales could be made contingent on the borrowing needs of the government (e.g., by preventing sales when deficits are above a certain level). This would, however, have to be communicated in ways that do not raise a suspicion of fiscal dominance. The former Governor of the Bank of Japan [see Shirakawa (2013)] has repeatedly warned central banks to avoid the trap of fiscal dominance.

Any quantity-based rule could be subject to price-based constraints. For instance, sales could also be halted if market volatility were to jump or if yields were already rising strongly. An explicit interest rate ceiling could put a floor under the price of government debt. This might help banks, pension funds and others who are holding large stocks of government bonds. As Hannoun (2012) points out, however, a major risk with such policies is that of “financial dominance” – that is, the central bank fails to tighten a policy when needed because it is frightened of the bond market’s reaction. He also warned of the danger of an asymmetric policy reaction in entering and exiting balance sheet

The consequences of exit from non-conventional monetary policy

There is a simple intertemporal reason why asymmetric reactions in macroeconomic policy can be troubling: they can move balance sheet variables further away from desirable levels, and so constrain future policy. It was because central banks had very small holdings of government bonds at the beginning of the recession that they were in a position to buy bonds in the recession on such a large scale without compromising their credibility. If central banks had already held 30% of outstanding government debt in 2008, would they have got away with purchasing a further 30% during the crisis?

7. Conclusion

The financial crisis, and the monetary policy responses to it, is forcing a pragmatic rethinking of the theory and instruments of central banking. There is greater awareness of the inherent limitations of monetary policy, and no new, simple paradigm has emerged. “Dogmas” of the earlier “doctrine” have been proved false. Central banks have rediscovered old tools, and used them with apparent and immediate success in influencing financial variables. The near-term risks that some in the early stages of the crisis saw (e.g., higher inflation) have not materialized. Time will tell about the longer-term side effects.

The consequences of the massive increase in long-term assets held by the central bank will last much longer, and are more uncertain, than changes in the policy rate. The effect on exchange rates of great divergence in changes in the size of the balance sheets of the major central banks is difficult to quantify. Recent large declines in the dollar value of the yen and the euro suggest such effects could be large.

The wide sweep of post-crisis policies has in effect increased the number of variables in future central bank reaction functions. And many of these variables are more susceptible to government

---

footnotes:

17 The Federal Reserve’s communication about planned asset sales has changed as the stock of bonds on its balance sheet has risen. At first, a policy of gradual, active sales was envisaged. The FOMC meeting in June 2011 stated that the timing and pace of sales would be communicated to the public in advance: once sales begin, the aim would be to eliminate “holdings of agency securities over a period of 3 to 5 years.” No mention was made of Treasury securities. More recently, the sense has been that bond holdings would just run off as they mature. Sagnon and Sack (2014) point out that, at the press conference following the June 2013 FOMC meeting, Chairman Bernanke said that a “strong majority” of the FOMC does not expect to sell assets during the process of normalizing monetary policy.

18 Federal Reserve Governor Stein (2014) explains that much depends on how markets formulate their expectations of future central bank action. In current circumstances, he defends gradualism in policy to avoid unnecessarily destabilizing markets. But, in other circumstances, “one may be able to achieve a better outcome with a central banker who places a lower weight on the intermediate objective of not roiling the bond market.”

19 See Posen (2013).
policies than is the policy rate. Markets have much more to guess about than before the crisis. Don Kohn (2013) believes that “this second guessing that is more intense than normal” could undermine “the perceived political legitimacy of the central bank and support for its independence in the conduct of policy.”

For many years ahead, the balance sheets of central banks will remain much more extended (both in size and in the nature of assets held) than is desirable in any normal equilibrium. Markets will, therefore, have to make judgments about how central banks would seek to adjust the size and composition of their balance sheets – over a period of years – to future macroeconomic or financial shocks. Markets will also have to assess fiscal policy: any prospect of chronic government overindebtedness will make it harder to sell government bonds. Another important influence will be government’s financing choices (what maturity? more floating rate debt?). Furthermore, regulations that affect financial institutions’ holdings of government bonds have an impact that is at present difficult to gauge. During the years ahead, the level (or at least the volatility) of benchmark long-term interest rates will get more attention than in the pre-crisis period. But any policy guideline for the long-term interest rates, inevitably controversial, may remain only implicit. As central banks exit, and a normal term premium is re-established, the macroeconomic focus of government debt managers – and how they react to movements in the long-term rate – will attract more scrutiny.

Moving central bank balance sheets toward more normal levels will require coordination with Treasuries without surrendering monetary policy independence. The warnings of the former Vice Chairman of the Federal Reserve Board [see Kohn (2014)] about the increased threats to central bank independence need to be taken seriously.20 Clerc and Raymond (2014) strike a similar note about the need for institutional independence.

The communication of future balance sheet policies raises delicate issues. Some opacity is inevitable. When a central bank buys an asset in order to drive up its price, it is unlikely at the same time to announce a specific date for selling the asset back to the market. Doing so, and particularly if the announced date is near, would just blunt the impact of the initial purchase. It would also violate the principle that central bank policies should be “data dependent,” responding to economic developments as they actually unfold. Nevertheless, communicating at least near-term plans for their balance sheet in a convincing way can help to lead markets in the desired direction. As central banks pragmatically monitor market resilience – “How much asset sales can the market absorb now?” – the “financial dominance” trap is to be avoided.

20 As he puts it, “Instrument independence is necessary to overcome the short-term perspective of politicians, who tend to be more interested in boosting economic growth before the next election and less focused on the longer-term inflationary consequences of such actions.”
Mohanty, M. S., and P. Turner, 2006, “Foreign exchange reserve accumulation in emerging markets: what are the domestic implications?” in BIS Quarterly Review, September

Posen, A., 2013, “The myth of the omnipotent central banker,” Foreign Affairs, July/August 166-170


Shirakawa, M., 2013, “Is inflation (or deflation) 'always and everywhere' a monetary phenomenon?” Keynote speech at the People's Bank of China/BIS Research Conference, Beijing, September


Stein, J. C., 2014, “Challenges for monetary policy communication,” Speech to the Money Marketeers of New York University, 6 May

Tily, G., 2010, Keynes betrayed: the General Theory, the rate of interest and 'Keynesian' economics, Palgrave Macmillan


The audit mandatory rotation rule: the state of the art

Mara Cameran
Researcher, Accounting Department, Università Luigi Bocconi

Giulia Negri
Research Fellow, Accounting, Control, Corporate and Real Estate Finance Department, SDA Bocconi School of Management

Angela Pettinicchio
Assistant Professor, Accounting Department, Università Luigi Bocconi

Abstract
Mandatory audit rotation imposes periodical breaks to audit engagements and is intended to avoid excessively long relationships between the auditor and the client. The E.U. has finally introduced mandatory rotation for the audit firm in addition to the already existing audit partner rotation rules. The U.S., however, has for now decided to retain the partner rotation rule without introducing mandatory audit firm rotations. After an overview of the experience of a number of countries, we summarize the pros and cons of a compulsory change in the audit firm. Moreover, we focus on the empirical evidence collected on the benefits and costs of the rule. So far, investigations into the impact of the rule at corporate and market level have not been able to prove that the benefits outweigh the costs.

1 We would like to thank François Langlois, Rauf Rashid and Tan Seng Choon from EY for helping us with some of the information included in Table 1, compiled by Giulia Negri.
1. Introduction
In response to a number of international accounting scandals, as well as the recent financial crisis, regulators around the world started considering ways in which audit quality could be enhanced. Auditors were accused of not being sufficiently competent and/or independent to prevent such scandals and to foresee the recent financial crisis.

Michel Barnier, E.U.’s Internal Market and Services Commissioner, stated that “the financial crisis and more recent inspection reports by national supervisors highlighted major shortcomings in the European audit sector. To address these deficiencies, we made ambitious proposals in November 2011 to clarify the role of statutory auditors, to strengthen their independence and to enhance supervision” [Barnier (2014)].

Perhaps the most questionable outcome of these discussions was the mandatory rotation rule. This rule imposes periodical breaks to audit engagements and is intended to avoid excessively long relationships between the auditor and the client, which is believed to damage the quality of the audit. Mandatory rotations can refer to the firm, in which case the entire audit firm has to rotate after a certain number of years, or to the audit partner, in which case only the leading partner who signs the opinion has to change periodically.

Such a rule, however, comes at a cost. Changing the auditor results in, among other things, organizational disruptions, start-up costs, loss of client-specific knowledge and the ability of the client to negotiate on audit fees, and could impact the quality of the service delivered. This is the reason why the rule has not been applied consistently around the world. As an example, the two most important regulators in the world, namely the European Commission (E.C.) in Europe and the PCAOB in the U.S., have taken very different routes. While the E.U., after having implemented the mandatory rotation rule at the partner level in 2006, recently decided, in 2014, to extend periodical changes to audit firms as well, the U.S. implemented partner mandatory rotation back in 1978, but decided, after long discussions, which also involved academics and public hearings, not to introduce the rule at the audit firm level, at least for now (for more details, please refer to section 2).

The objective of this paper is to provide an overview of the status of the application of the rule around the world and, most importantly, to summarize the main conclusions on the effectiveness of the rule that the academic literature has reached in recent decades. The main focus is on the audit firm rotation rule; even if, for the sake of completeness, we have also dedicated some space to the partner rotation rule.

The picture that arises from the literature is that there is not a clear path regarding the application of the mandatory rotation rule. First, from a legislative point of view, countries around the world have taken different decisions. While partner mandatory rotations seem to have been better accepted, and therefore are more commonly applied around the world, audit firm mandatory rotations have been implemented in only a few countries, perhaps because of the higher costs associated with doing so. However, the recent decision of the E.U. might change this trend in the future.

With regard to the empirical evidence on the benefits and costs of the rule, the picture is not clear at all. Different studies have investigated different research settings, with only a few of them analyzing proper mandatory rotation environments (most studies have tried to indirectly infer the potential effects of the rule by looking at cases of voluntary rotations). Unfortunately, in general, the evidence collected is far from being conclusive. It is not clear whether the benefits derived from imposing periodical changes to the auditor are greater than the costs incurred, both at the partner and at the audit firm levels.

A clearer picture can be obtained by looking at the case of Italy, which introduced mandatory audit firm rotation back in 1975. Looking at the few studies that investigated the case of mandatory audit rotation in Italy, we find that while it is quite costly, its benefits, in terms of improved audit quality, cannot be confirmed.

The rest of the paper is structured as follows: in the next section, we provide a brief overview of the application of the rule around the world. We then discuss the arguments for and against the application of the mandatory audit rotation rule in section 3. Section 4 describes the empirical evidence gathered by the academic literature and section 5 concludes the paper.

2. The application of the rule around the world
In this section, we review the decisions made by a number of
countries about the process of audit rotation. Overall, we find that, while only a few countries are keen to adopt audit firm rotation, many support the idea of partner rotation. After providing an overview of the actions taken by a number of countries in this regard, we will focus specifically on the U.S., which has been debating audit rotation for decades and has, as yet, not introduced it, and the E.U., which finally adopted the audit firm rotation rule in 2014, effectively starting from mid-2016.

In Canada, mandatory firm rotation was introduced much earlier than it was in other countries (shortly after the Home Bank failure of 1923) and was in place until 1991 when it was removed with the revision of the Bank Act (Government Accountability Office (GAO) (2003)), which blamed the switching costs borne by the incoming auditor as not offsetting the benefits. In other countries, such as Spain, audit firm rotation was adopted and shortly after abrogated. The firm rotation was mandatory in Spain every nine years from 1988 to 1995 (Ruiz-Barbadillo et al. (2009)). The first nine-year term of the audit firm tenure had not even expired when the rule was abolished. Singapore also introduced the mandatory rotation in 2002 for banks, only to “temporarily suspend” it in 2008. This was justified on the basis that the regulators did not wish to place further cost burdens on banks during the last financial crisis. Finally, the case of Costa Rica is emblematic: audit firm rotation was introduced in 2005, appealed in 2006, ejected in 2007 and re-implemented in 2010.

Table 1 provides an overview of audit rotation status in a number of countries. This rule is defined predominantly for public interest entities (PIEs), but the exact identification of the kind of legal entities that are mandated to rotate either the firm or the partner is beyond the scope of this paper. This brief review takes into consideration the Member States of the E.U., those countries that are similar to the E.U. due to their economic conditions and those countries that do not belong to the E.U. but that have taken a peculiar stance vis-à-vis mandatory rotation, for which we were able to find information. Data was retrieved from the sources indicated in the table. The case of the U.S. will then follow.

2.1. The U.S.
The U.S. has always been somewhat different from other countries when it comes to audit firm rotation. On one hand, the U.S. has had a long experience with partner rotation: seven-year audit partner rotation was introduced back in 1978 (AICPA (1978)). On the other hand, the debate about firm rotation was on the House of Representatives’ agenda until 2014.

Typically, the level of interest in partner and audit firm rotation has been influenced by episodes of fraud or bankruptcy. After the major cases of fraud that took place at the beginning of the 2000s, Section 203 of the Sarbanes-Oxley Act of 2002 expressly stated that: “[…] it shall be unlawful for a registered public accounting firm to provide audit services to an issuer if the lead (or coordinating) audit partner […], or the audit partner responsible for reviewing the audit, has performed audit services for that issuer in each of the five previous fiscal years of that issuer.” This policy emphasized the five-year mandatory partner rotation and even reduced the tenure by two years. At the same time, Section 207 of the Act required that the GAO undertakes the “Study of mandatory rotation of registered public accounting firms.” Under Section 207, the Securities Exchange Commission (SEC) also mandates a study on the potential implications of this kind of rotation. After a year of conducting surveys among the accounting firms and the Fortune 1,000 publicly traded companies in the Required study on the potential effects of mandatory audit firm rotation (GAO (2003)), the GAO stated that it “believes that mandatory audit firm rotation may not be the most efficient way to strengthen auditor independence and improve audit quality considering the additional financial costs and the loss of institutional knowledge of the public company’s previous auditor of record.”

Consequently, the topic was dropped for another decade, until the PCAOB suggested in 2011 that firm rotation would have been the best way to grant a high level of professional skepticism. Two years later, in July 2013, the U.S. Congress ruled against the possibility of audit firm rotation. In the end, after three years of work, in February 2014 the PCAOB Chairman, James Doty, told the SEC that: “We don’t have an active project or work going on within the Board to move forward on a term limit for auditors (but we) continue to think about what impacts independence” (Doty (2014)), setting aside the mandatory firm rotation. Doty, himself

---

2 For the purpose of this article, we take “professional skepticism” to mean, as stated in Directive 2014/56/EC, “an attitude that includes a questioning mind, being alert to conditions which may indicate possible misstatement due to error or fraud, and a critical assessment of audit evidence.”
at the SEC Open Meeting on the PCAOB 2015 Budget, eventually reprised this concept when he underlined, among others, the auditors’ independence as a core value an investor expects and deserves in audits without making specific reference to the mandatory rotation [Doty (2015)].

2.2. The E.U.

We treat the E.U. as a unique area because, during the last decade, the E.C. has issued a number of different guidelines that concern all Member States. Under the E.U. guidelines, the key audit partner rotation is mandated after seven years in PIEs, as E.C. Directive 2006/43/EC states, and the key audit partner shall not be able to participate in the audit of the same entity for at least two years. Member States were given two years to implement the directive (until mid-2008) and, in case they thought it might be necessary, were allowed to impose more stringent requirements. After the recent financial crisis, the E.C. issued, in October 2010, a Green Paper entitled: Audit policy: lessons from the crisis [E.C. (2010)]. In particular, it was stated that, while a lot of attention was devoted to the role that banks, hedge funds, rating agencies, supervisors and central banks played in the run-up to the crisis, no attention was directed at the role that auditors of public companies might have played. One of the key points expressed in the Green Paper was that nowadays an even larger number of companies are under the audit responsibility of a smaller number of big audit companies after the demise of Arthur Andersen and the merger of Price Waterhouse and Coopers & Lybrand.

In 2012, the European Parliament proposed audit firm rotation as mandatory at 25 years. However, after Germany and Austria voted strongly against such a long tenure, it was to be reduced to 21 years [Ewelt-Knauer et al. (2012)].

<table>
<thead>
<tr>
<th>Country</th>
<th>Partner rotation</th>
<th>Audit firm rotation</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>5 years</td>
<td>No</td>
<td><a href="http://www.cpaaustralia.com.au">http://www.cpaaustralia.com.au</a></td>
</tr>
<tr>
<td>Austria</td>
<td>5 years</td>
<td>5 years for governmental owned companies</td>
<td>EY Transparency Report 2014 and <a href="https://www.crowehorwath.net">https://www.crowehorwath.net</a></td>
</tr>
<tr>
<td>Belgium</td>
<td>6 years</td>
<td>No (reappointment every 3 years with no limits)</td>
<td>Ewelt-Knauer et al. (2012)</td>
</tr>
<tr>
<td>Brazil</td>
<td>No</td>
<td>5 years</td>
<td>Harris and Whisenant (2012)</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>5 years</td>
<td>No</td>
<td><a href="http://www.crowehorwath.net">http://www.crowehorwath.net</a></td>
</tr>
<tr>
<td>Canada</td>
<td>7 years</td>
<td>No</td>
<td><a href="http://www.cga-canada.org">http://www.cga-canada.org</a></td>
</tr>
<tr>
<td>China</td>
<td>5 years</td>
<td>5 years for financial institutions and state-owned enterprises (there must be a tender process every 3 years)</td>
<td>Lennox et al. (2014) and <a href="http://economia.icaew.com">http://economia.icaew.com</a></td>
</tr>
<tr>
<td>Croatia</td>
<td>7 years</td>
<td>7 years for banks and 4 years for insurance and leasing companies</td>
<td><a href="http://anale.feea.uaic.ro/">http://anale.feea.uaic.ro/</a>) and Ewelt-Knauer et al. (2012)</td>
</tr>
<tr>
<td>Cyprus</td>
<td>7 years</td>
<td>No</td>
<td>EY Transparency Report 2013</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>7 years</td>
<td>No</td>
<td>Ewelt-Knauer et al. (2012)</td>
</tr>
<tr>
<td>Denmark</td>
<td>7 years</td>
<td>No</td>
<td>EY Transparency Report 2014</td>
</tr>
<tr>
<td>Estonia</td>
<td></td>
<td>7 years</td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>7 years</td>
<td>No</td>
<td>Ewelt-Knauer et al. (2012)</td>
</tr>
<tr>
<td>France</td>
<td>6 years</td>
<td>No (6 years with joint audit but may be renewed)</td>
<td>André et al. (2015) and Francis et al. (2009)</td>
</tr>
<tr>
<td>Germany</td>
<td>7 years</td>
<td>No</td>
<td>Ewelt-Knauer et al. (2012)</td>
</tr>
<tr>
<td>Greece</td>
<td>7 years</td>
<td>No</td>
<td>EY Transparency Report 2013</td>
</tr>
<tr>
<td>Hungary</td>
<td>7 years</td>
<td>No</td>
<td>Information retrieved directly via the EY network</td>
</tr>
<tr>
<td>Ireland</td>
<td>5 years</td>
<td>No</td>
<td><a href="http://www.cpaireland.ie">http://www.cpaireland.ie</a></td>
</tr>
</tbody>
</table>
With Directive 2014/56/EU and Regulation 537/2014, a new statutory audit framework was introduced. The objective of this reform was the enhancement of audit reports’ quality. The audit firm tenure for PIEs will be limited to 10 years and the same audit firm can be reappointed for another 14 years in case of joint audits (the company is being audited by more than one audit firm) or for another 10 years in the case of a public tender. After the expiry of the maximum tenure, the PIE may reappoint the audit firm for a further two years, leading to a “10+14+2” rule in case there is a joint audit during the second period or to a “10+10+2” rule in case a single firm provides the audit service. The partner rotation remains mandatory after the seventh engagement year; both firm and partner rotation period can be shortened by Member States. These specific requirements will automatically apply on 17 June 2016.

Between 17 June 2014 and 17 June 2016, Member States are permitted to adopt the audit firm mandatory rotation rule earlier and may even shorten the tenure of audit term. At the time of writing, some countries have already implemented the audit firm rotation rule. In the U.K., firm mandatory rotation rule was introduced on the 1 January 2015 and the Netherlands made it effective from 1 January 2016.

### 3. The pros and cons of mandatory rotation

Regulators, accountants, academics and the public in general have been debating the need for mandatory audit firm rotation.

---

**Table 1: Audit rotation around the world**

<table>
<thead>
<tr>
<th>Country</th>
<th>Partner rotation</th>
<th>Audit firm rotation</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy</td>
<td>7 years</td>
<td>9 years</td>
<td>Cameran et al. (2015a)</td>
</tr>
<tr>
<td>Latvia</td>
<td>7 years</td>
<td>No</td>
<td>EY Transparency Report 2014</td>
</tr>
<tr>
<td>Lithuania</td>
<td>5 years for PIEs; 7 years for some other legal entities</td>
<td>No</td>
<td><a href="http://www3.irs.lt/">http://www3.irs.lt/</a></td>
</tr>
<tr>
<td>Luxembourg</td>
<td>7 years</td>
<td>No</td>
<td>Ewert-Knauer et al. (2012)</td>
</tr>
<tr>
<td>Malta</td>
<td>7 years</td>
<td>No</td>
<td>EY Transparency Report 2013</td>
</tr>
<tr>
<td>Poland</td>
<td>5 years</td>
<td>No</td>
<td>Information retrieved directly via the EY network</td>
</tr>
<tr>
<td>Portugal</td>
<td>7 years</td>
<td>No (8-9 years for listed companies on a “comply or explain” rule)</td>
<td>Ewert-Knauer et al. (2012)</td>
</tr>
<tr>
<td>Romania</td>
<td>7 years</td>
<td>No</td>
<td><a href="http://www.crowehorwath.net">http://www.crowehorwath.net</a></td>
</tr>
<tr>
<td>Singapore</td>
<td>5 years for listed companies; 7 years for other PIEs</td>
<td>No</td>
<td>Information retrieved directly via the EY network</td>
</tr>
<tr>
<td>Slovakia</td>
<td>5 years</td>
<td>5 years</td>
<td><a href="http://www.crowehorwath.net">http://www.crowehorwath.net</a></td>
</tr>
<tr>
<td>Slovenia</td>
<td>7 years</td>
<td>5 years for the banking and insurance industry, only</td>
<td>Ewert-Knauer et al. (2012)</td>
</tr>
<tr>
<td>South Korea</td>
<td>No</td>
<td>No</td>
<td>Kwon et al. (2014)</td>
</tr>
<tr>
<td>Spain</td>
<td>7 years</td>
<td>No</td>
<td>Ruiz-Barbadillo et al. (2009)</td>
</tr>
<tr>
<td>Sweden</td>
<td>7 years</td>
<td>No</td>
<td><a href="http://lup.lub.lu.se/search/">http://lup.lub.lu.se/search/</a></td>
</tr>
<tr>
<td>The Netherlands</td>
<td>7 years</td>
<td>No</td>
<td><a href="http://ey.com/Publications">http://ey.com/Publications</a></td>
</tr>
<tr>
<td>Turkey</td>
<td>5 years</td>
<td>7 years</td>
<td>Information retrieved directly via the EY network</td>
</tr>
<tr>
<td>U.K.</td>
<td>5 years</td>
<td>10 years starting 1 January 2015</td>
<td><a href="https://www.gov.uk/">https://www.gov.uk/</a></td>
</tr>
<tr>
<td>U.S.</td>
<td>5 years</td>
<td>No</td>
<td>See below</td>
</tr>
</tbody>
</table>

---

for decades but, after the recent major financial scandals, the topic became the subject of discussion also among a broader number of national governments and institutions. From a theoretical point of view, the rule could have both positive and negative implications on audit quality. Below, we summarize the most important arguments in favor and against the application of the rule, drawing from both academic and regulatory literature. Please note that most of the arguments may apply to rotation of both the audit firm and partner even if this article is focused on the former.

3.1. The pros of mandatory audit firm rotation
Generally speaking, mandatory audit firm rotation is considered as a way to increase and enhance audit quality. Here we list the major benefits that supporters of mandatory auditor rotations have highlighted.

1. Enhancing audit client independence: one of the main benefits that should accrue from imposing periodical audit firm changes is the increase in audit independence, which is closely linked to audit quality (this is further discussed in section 4). Audit independence may be defined as being made up of two components: independence in appearance and in fact. The former indicates that the auditor only appears to be independent, whereas the latter indicates that the auditor has an independent mindset when planning and executing an audit [Dopuch et al. (2003)]. These two different components may be affected by the long-lasting relationship existing between the auditors and the clients. Many scholars [e.g., Brody and Moscove (1998) and Mautz and Sharaf (1961)] believe that, as time passes, members of the audit team begin to unconsciously feel themselves as part of the client management team and that this might impede the auditors from impartially assessing the transparency and reliability of the clients’ accounts, thus influencing independence in fact. Moreover, a long-term relationship between a single audit firm and an auditee may cause a reasonable investor to conclude that the auditor would not be capable of acting without bias, thus impacting on independence in appearance. As Turner, one of the former Chief Accountants of the SEC, explained: “auditor independence is really about only one thing – investors’ confidence in the numbers” [Turner (2000)]. A long-term relationship between auditor and auditee may decrease investors’ confidence in the numbers stated within the financial statements.

2. The “fresh look” benefit: another major benefit of imposing periodical changes of the audit firm emanates from what has been described as the “new fresh look” or “new fresh eyes” [Shockley (1981)]. When the same audit firm carries out its services for a long period of time, it may fall into a sort of professional routine. The auditor may be tempted, for example, to rely increasingly on previous audit tests and may overlook certain accounting areas that had been considered as “correct” in previous audits, thus potentially missing material errors and mistakes. Instead, when a new audit firm comes in, it approaches the financial statements of the client with a “new” pair of critical eyes, thus paying greater attention to all areas of the annual report.

3. Reduction in economic dependence: mandatory auditor rotation would also reduce the period of “economic dependence” on the client, which would in turn decrease the potential incentives for supporting the accounting policies of the client. As audit technology is characterized by significant start-up costs, incumbent auditors can earn client-specific “quasi-rents” [DeAngelo (1981)]. Since an incumbent auditor possesses cost advantages over potential competitors, it is able to set future audit fees above the costs of producing audits, thus allowing it to earn client-specific quasi-rents. If the relationship is terminated, the incumbent auditor will lose the wealth equivalent of the client-specific quasi-rent stream. In order to avoid the loss of client-specific quasi-rent, the auditor has the incentive to act opportunistically, e.g., to soften audit controls, if this permits the incumbent auditor to retain the client in the future.

4. Embarrassment effect: during the final years of an engagement, as the auditor knows that a new audit firm (a competitor) will come in to replace it and examine its working practices, it will be incentivized to incentivized audit quality in order to protect its reputation. This could eventually turn into a virtuous circle that would eventually enhance audit quality.

5. Increase in competition: one of the major concerns of regulators around the world is the potential lack of competition arising from the oligopoly structure of most
audit markets, where the Big Four (Deloitte, EY, KPMG and PwC) audit the great majority of clients. Introducing mandatory periodical auditor changes could reduce this phenomenon and increase competition; especially if this rule is combined with the choice of non-big audit firms as a substitute for one of the big auditors.

3.2. The cons of mandatory audit firm rotation
Here, the potential negative effects of introducing the mandatory audit firm rotation rule are described.

1. Switching costs: changing of an audit firm introduces certain costs that may be borne by both parties, the audit firm and the client. As highlighted by DeAngelo (1981), if the relationship between auditor and client ends, the incumbent auditor loses the wealth equivalent of the client-specific quasi-rent stream. The client is also forced to bear the costs of switching and the costs associated with “training” a new auditor.

2. Loss of specific knowledge: imposing periodical auditor changes results in the loss of client-specific knowledge possessed by the audit firm. That is, a certain amount of time needs to pass before the new auditor becomes familiar with the accounting systems, internal controls, etc., of the client (the “learning curve”). During this period, therefore, the auditor could potentially miss material errors and mistakes, with a consequent negative impact on audit quality. The more complex and industry-specific the business of the client is, the deeper and longer this learning curve will be.

3. The information cost: to provide a more holistic picture, the perspectives of the stakeholders also need to be taken into consideration [Ewelt-Knauer et al. (2012)]. In fact, in the context of voluntary auditor rotation, market players can easily find out whether a new audit firm has been appointed and thus determine whether there is a risk of “opinion shopping” (i.e., when a company replaces an audit company with another one in order to obtain a more favorable opinion). In a mandatory rotation environment, such information is more expensive since the stakeholders cannot easily determine whether the decision made was based on a voluntary or mandatory basis.

4. Audit quality
According to GAO (2003): “mandatory audit firm rotation (setting a limit on the period of years a public accounting firm may audit a particular company’s financial statements) was considered as a reform to enhance auditor independence and audit quality.” However, before discussing the studies that have attempted to understand the actual implications of the mandatory rotation rule, it is necessary to clarify the relationship between auditor independence and audit quality and how they have been measured by academic researchers.

According to DeAngelo (1981): “the quality of audit services is defined to be the market-assessed joint probability that a given auditor will both (a) discover a breach in the client’s accounting system, and (b) report the breach. The probability that a given auditor will discover a breach depends on the auditor's technological capabilities, the audit procedures employed on a given audit, the extent of sampling, etc. The conditional probability of reporting a discovered breach is a measure of an auditor’s independence from a given client.” As this definition clearly illustrates, audit quality and auditor independence are strongly linked.

As far as the empirical measurement of audit quality is concerned, we should note that this implies the assessment of three types of judgments [Moizer (1997)]: “(i) whether the amount and nature of audit work undertaken is appropriate for that particular client company; (ii) how technically competent the audit staff are to undertake the work properly; and (iii) how independent the audit firm is and hence how likely it is to report any unadjusted errors or omissions that it finds.” Making such a judgment implies having access to the audit working papers and interviewing the key personnel involved in the audit. This is not possible for an external observer, hence making it necessary to use some proxies to judge the audit quality of a given engagement. The main proxies used in the academic literature may be grouped into two categories:

- Measures based on the type of opinion issued by the auditor
- Measures based on the earnings quality of the auditee

There are, in fact, other factors that may be considered as measures of audit quality. For example, DeAngelo (1981) suggests that size could be an indicator of audit quality as, the larger the size, the more independent an audit firm is from one specific client.
4.1 The opinion
One of the ways in which audit quality could be measured is related to the ultimate output of the audit work, i.e., the opinion as expressed in the independent auditor reports. Many studies have used audit opinions in order to identify what is considered to be a “failure” in the audit work carried out. This refers to those episodes in which the stakeholders’ decisions might be misled by an inaccurate audit firm opinion. In particular, researches have focused on episodes of “false positive” and “false negative” opinions. A “false positive” opinion is issued in situations where a going concern report is provided for a company that eventually does not go bankrupt. On the other hand, a “false negative” opinion is issued in case a clean report is issued for a company that in the end, fails (Francis (2004)).

4.2 Earnings quality
Perhaps one of the most widely used measures of audit quality is clients’ earnings quality. The assumption behind this proxy is that audits of higher quality should lead to higher levels of quality of the accounting numbers published by the client. The great majority of the studies measure earnings quality by looking at the levels of “discretionary accruals” in the earnings figure. As the name itself suggests, these models aim to determine the accruals that are supposed to be “discretionary” or “abnormal,” deriving from accounting manipulations. In particular, these models estimate the discretionary component of the “working capital accruals” (abnormal working capital accruals – AWCA) as the difference between the expected value of working capital and the actual value. Other measures traditionally used by academic research to measure earnings quality (and therefore, in turn, audit quality) are the frequency of “small positive earnings” (based on the asymmetry that earnings distributions show around zero) and earnings conservatism (based on the assumption not always accepted – that higher levels of conservatism in the accounting numbers are a sign of higher earnings quality). Other measures of earnings quality involve market perception.

For example, earnings response coefficients measure the way changes in stock prices reflect changes in accounting numbers: higher levels of association signal stronger belief in accounting numbers by market players, suggesting, in turn, higher levels of earnings quality.

5. Empirical evidence
5.1 Literature review
From a theoretical point of view, as mentioned earlier, it is not clear whether the mandatory rotation rule can be considered as an effective means of enhancing audit quality: that is, whether potential benefits deriving from imposing periodical changes might counterbalance the expected negative effects. It, therefore, becomes crucial to examine the empirical evidence that has been gathered in the past in different research settings.

The majority of studies examining the implications of mandatory audit firm rotation have used data from situations where the rotation was voluntary, and hence were mainly studying how audit firm tenure affects audit quality. Results are mixed. For example, considering audit failure as an audit quality measure, Geiger and Raghunandan (2002) document that U.S. firms going into bankruptcy are less likely to have received a going concern audit opinion from audit firms with shorter tenure. Carcello and Nagy (2004), who also use a U.S. database of companies, find that fraudulent financial reporting is more likely when audit firm tenure is short, i.e., three years or less. On the other hand, using a sample of Belgian private companies, Knechel and Vanstraelen (2007) show that the decision of the auditor to issue a going concern opinion is not affected by the audit firm tenure in their bankrupt sample. In the non-bankrupt sample, they document some evidence of a negative association between auditor tenure and the issuance of a going concern opinion. Using earnings quality as a measure for audit quality, Chung and Kallapur (2003) and Myers et al. (2003) find that discretionary accruals are negatively related to audit firm tenure in U.S. companies. Similarly, Johnson et al. (2002), and Gul et al. (2007) find evidence of higher discretionary accruals in the early years of the audit firm’s tenure. Jenkins and Velury (2008) document a positive association between conservatism in reported earnings and the length of the auditor-client relationship, and an increase

---

5 For the purpose of this article, we are using the definition of earnings quality to mean that a higher level of earnings quality is associated with the earnings’ capacity in reliably reflecting the corporate performance and the medium- to long-term profitability.

6 In particular, this measure is based on the assumption that higher than expected (i.e., compared with the normal distribution of earnings) of small positive earnings (i.e., earnings scaled by total assets of between 0% and 2%) is a sign of earnings manipulation carried out by the management to avoid reporting losses.

7 The most used measure of earnings conservatism looks at the (higher) speed with which losses are incorporated into earnings as compared with gains.

8 As defined in 4.1., an episode of audit failure refers to the situations in which stakeholders may be misled by “false positive” and “false negative” opinions.
in conservatism between short and medium audit firm tenure that does not deteriorate over long tenure. Other studies, using Taiwanese [Chi and Huang (2005)] and U.S. data [Davis et al. (2009)], find that earnings quality increases in the early years of audit firm tenure, and later deteriorates. Comparing different audit firm engagement lengths in Belgium, Vanstraelen (2000) suggests that long-term audit firm engagements significantly increase the probability of issuance of an unqualified audit report. Finally, there are studies that suggest that the relationship between audit quality and audit firm tenure is not homogeneous for all firms in the U.S. [e.g., Li (2010), Gul et al. (2009)].

As the auditor incentives are different, conclusions drawn from voluntary replacement environments cannot be easily extended to mandatory rotation settings [see also Casterella and Johnston (2013)]. In fact, the number of potential future reappointments from the existing client is ideally (theoretically) equal to infinity in a voluntary setting, which implies different auditor incentives in comparison with a mandatory rotation setting (where the maximum term is fixed). In an attempt to overcome this limitation, some papers have tried to model a mandatory rotation setting on a theoretical basis, with conflicting conclusions.

For example, in a multiperiod model, Elitzur and Falk (1996) show that planned audit quality level diminishes over time and the level of the last period is the lowest, concluding that planned audit quality is negatively affected by the policy of audit firm mandatory rotation. Arruñada and Paz-Ares (1997) focus on the expected financial consequences of the auditor’s reporting decision. They conclude that the rotation rule does not modify the transaction costs of collusion and reduces both the probability of detecting “non-reporting auditors” (i.e., auditors that do not report irregularities after detecting them) and the level of sanctions associated with such detection. Gietzmann and Sen (2002) find that mandatory rotation should only be imposed in thin markets where a few clients are very important to the auditor, as in these markets the resulting improved incentives for independence outweigh the associated costs.

Some empirical evidence on the audit firm rotation rule was collected in settings where researchers can find forced auditor change, but not a proper mandatory rotation rule such as the Arthur Andersen (AA) collapse in 2002. Also, the results from these studies are conflicting. For example, while some studies find that forced audit firm rotation following the AA collapse is associated with better audit quality [Cahan and Zhang (2006), Krishnan (2007), Nagy (2005)], others document the opposite [Blouin et al. (2007), Krishnan et al. (2007)]. Despite the fact that such change was in fact compulsory, this specific episode has at least two differences with an actual mandatory rotation setting. The tenure allowed was not established at the moment in which the audit firm was appointed. Second, the level of control of the new auditor is presumably much greater than normal. Actually, the new audit firm is motivated to audit the new auditee with greater care, because the reputation of the previous auditor was severely damaged. For example, Cahan and Zhang (2006) show that the succeeding auditors viewed an AA audit as a unique source of litigation risk. In Korea, from 1991 to the early 2000s, an auditor change was to be imposed by a financial supervisory commission on companies deemed to have a high potential to manipulate accounting results. In this setting, Kim and Yi (2009) find lower levels of earnings management following a regulator-imposed audit firm change. However, Kim and Yi (2009) recognize the uniqueness of the Korean auditor replacement rule and note that their conclusions cannot be generalized to a mandatory rotation setting.

There are very few studies that consider real audit firm mandatory rotation settings. One of the reasons for this is the mere fact that audit firm rotation is mandatory only in a small number of countries (see section 2), thus making it very difficult to collect sound empirical evidence.

Ruiz-Barbadillo et al. (2009) analyze the Spanish setting, comparing a period where the mandatory audit firm rotation rule was effective (1991–1994) with a voluntary rotation period (1995–2000). Their findings do not show any significant change in audit quality between the two periods. However, in the Spanish setting, the rule was never actually implemented because it was dropped before the first rotations imposed by law could take place. In Korea, audit firm rotation was required every six years starting in 2006 up to 2010 [Kwon et al. (2014)]. Kwon et al. (2014) showed that the post-regulation period audit fees were much higher than those preceding the rule and there was no impact on financial reporting quality. Again, in this case, as in the case of Spain, the rule seems to have never been actually implemented. Studying the audit framework in China, Firth et al. (2012) focused on a setting where different kinds of rotation
The audit mandatory rotation rule: the state of the art

(i.e., audit firm and audit partner) are mandatory. Using modified audit opinions, the authors find a positive effect from mandatory audit partner rotation on audit quality for firms located in regions with weak legal institutions. On the other hand, mandatory audit firm rotation does not seem to have clear benefits. Firth et al. (2012: p.118) explain that they “classify an audit firm rotation as mandatory if the preceding audit firm changes because of its inability to provide audit services for the client.” Hence, most rotation cases used in their study are not comparable with the typical mandatory audit firm rotations that operate on a periodic basis. Consequently, the Firth et al. (2012) results cannot easily be extended to a typical audit firm mandatory rotation setting.

The aforementioned studies suggest that the literature fails to provide sound and consistent empirical evidence about the audit firm mandatory rotation rule.

Another stream of the literature tried to determine the actual effects of the mandatory rotation rule by looking at cases of audit partner changes, both mandatory and voluntary. Even if the main focus of this article is mandatory audit firm rotation, for the sake of completeness, we summarize below the main evidence gathered from partner rotation studies. It should be stated that in this case also, the results do not univocally support a rotation requirement.

Until now, the implications of partner rotation has not been empirically tested widely as audit partner names are not always publicly disclosed in the independent auditors’ opinion around the world, making it difficult for an external researcher to understand when the change occurs. So the major part of the empirical evidence on the topic was collected for countries where partner identity is disclosed or using proprietary data. An exception being Litt et al. (2014), where the authors used Sarbanes-Oxley (SOX) partner rotation requirements to infer partner rotation using the years of tenure with the same audit firm. Their study focused on U.S. companies, where the name of the signing partner is also not made public. They find evidence of lower financial reporting quality following an audit partner change, especially for larger clients. For non-Big Four partners, city-level non-industry specialist auditors and smaller audit offices, this effect is even more pronounced. As the same authors recognized, they cannot directly observe audit partner identities, and so their results are limited by the use of an approach that indirectly tries to ascertain the occurrence of audit partner rotation.

Using Australian data and a voluntary partner rotation regime, Carey and Simnett (2006) do not find evidence of deterioration in reporting quality (measured by abnormal working capital accruals) for long partner tenure. However, for long partner tenure, they find a lower propensity to issue a going concern opinion and some evidence of higher likelihood of just beating earnings benchmarks.9 Fargher et al. (2008), however, who also use Australian data, find results that are consistent with a positive effect of partner rotations. In Taiwan, Chi and Huang (2005) document that discretionary accruals are initially negatively associated with audit partner tenure and audit firm tenure, meaning a positive impact on audit quality. However, the associations become positive (and thus have a negative impact on audit quality) when tenure exceeds five years. The dataset used in this study covers a period when partner rotation was voluntary. Chen et al. (2008) find a positive relationship between reporting quality and partner tenure (once again, in a period where audit partner change was voluntary). Chi et al. (2009) address this issue following the implementation of mandatory partner rotation in Taiwan (while previous cited studies on the Taiwanese setting use data referring to the period before the implementation of the rule) and find results consistent with Chen et al. (2008).

In order to eliminate the limitations associated with the use of voluntary partner rotations, some researchers have used proprietary data. Using a small sample of U.S. proprietary data, Manry et al. (2008) show that audit quality increases with partner tenure, but only for certain types of auditees (relatively small clients having a fairly lengthy partner tenure). Finally, Bedard and Johnstone (2010), also using proprietary data on U.S. companies, find that the level of planned effort (as a proxy for audit quality: the higher the audit effort, the better the quality of financial statements) does not differ for clients having longer versus shorter tenured partners.

5.2 The Italian case
As stated earlier, Italy is one of the few countries in which both

---

9 The probability of just meeting/beating earnings benchmarks as the “zero earnings” or analysts’ forecasts is a widely used measure of earning quality (in particular, it is a measure of earnings quality that involves market perception, similar to those briefly mentioned in section 3). Assuming normal distribution of earnings numbers, a higher-than-expected likelihood of just meeting/beating such benchmark is considered a signal of earnings manipulation.
kinds of rotations (partner and firm) have coexisted for a number of years. Moreover, as the audit firm mandatory rotation rule has been in effect for more than 30 years, different studies were conducted with a specific focus on this setting.

5.2.1 The audit legal framework in Italy
The Presidential Decree D.P.R. 136/1975 introduced mandatory audit firm rotation for listed companies in Italy in 1975. The rule stated that audit firm engagement should last three years as a maximum, and that they may be reappointed for two further three-year periods, with a cooling-off period of five years (this was in fact defined as the 3+3+3 rule). This rule does not impose audit partner rotation. A second audit reform in 1998 (Legislative decree 58/1998) stated that the audit firm engagement should last three years as a maximum and that this can be renewed for an extra two three-year periods (leading again to the 3+3+3 rule). This rule, however, made no references to the cooling-off period, thus giving rise to many issues with interpretation. These were addressed by CONSOB (the Italian equivalent of U.S. SEC), which stated that the cooling-off period should last three years. There was still no mention of the issue of partner rotation.

In the years following 1998, many financial frauds occurred all over the world and in Italy, too: the most famous being the Parmalat case. This led the legislator to strengthen the audit legal framework in 2005 (Law 262/2005) and mandate a six-year tenure, renewable once with a cooling-off period of three years. For the first time, an audit partner rotation rule was also introduced in Italy, where audit partners were required to be changed every six years.

Because of the many interpretation issues arising from Law 262/2005, the following year, another legislative decree (No. 303/2006) was issued that clearly limited the maximum length of audit firm engagement to nine years and set a three years’ cooling-off period. Concerning the partner rotation, the rules remain unchanged: the partner must rotate every six years with a cooling-off period of three years.

The last reform was introduced in 2010, in which the maximum audit firm tenure was confirmed to be nine years with a cooling-off period of three years. As for the partner rotation, this was set at seven years with a cooling-off period of three years.

5.2.2 The implications of the mandatory rotation rule: evidence from the Italian case
Cameran et al. (2015b) examined a sample of non-financial Italian firms in the period 1985–2004 (i.e., the period of the 3+3+3 rule) and assumed that a better audit quality is associated with a higher level of accounting conservatism. The study found that the auditor becomes more conservative in the last three-year period, i.e., the one preceding the mandatory rotation. Since the auditor has incentives to be reappointed at the end of the first and the second three-year periods, the audit quality is lower in these periods, as compared with the third (i.e., the last) term. So, after a mandatory audit firm rotation, audit quality decreases.

Cameran et al. (2015a) focus their research on the effects of the mandatory rotation rule in a sample of listed companies in the period 2006–2009. The authors found that outgoing auditors do not reduce their effort (nor their quality), but final year fees are on average 7% higher than normal, which may hide opportunistic pricing. The fees of incoming auditors are on average lower by 16%, even though they show unusually higher engagement hours in the first year (17%), which could imply lowballing behavior. However, subsequent fees are abnormally higher and exceed the initial fee discount. Overall, the evidence provided by this paper shows that the costs resulting from mandatory rotation are not that irrelevant. Moreover, the study shows that these higher costs are not associated with higher audit quality. On the contrary, the authors show that earnings quality is lower in the first three years of the nine-year engagement than in the last years of the period.

Looking at the impact of the audit firm mandatory rotation rule on audit market competition, Cameran and Pettinicchio (2011) document a steady increase over time in the market share in Italy held by big audit firms, with an even stronger effect in the Italian market segments subject to mandatory rotation, suggesting that the rule per se does not necessarily increase audit market competition. The effects of the mandatory rotation rule on audit quality are also assessed, examining the types of audit opinions issued and the number of partner suspensions. Results show that the number of partner suspensions for poor-quality work is significantly higher in the first year of engagement compared with all other years, providing some evidence that audit quality may initially suffer as a result of auditor rotation.
On the basis of what has been stated earlier, we can conclude that the empirical evidence collected in Italy seems not to support the idea of audit firm rotation; not only is it expensive and leads to opportunistic pricing, but it also does not enhance earnings quality.

5.3 Is the audit firm rotation rule perceived as effective?
Given the high level of uncertainty associated with the actual costs and benefits of the mandatory rotation rule, it is worth examining another stream of literature. The studies mentioned so far have, in fact, generally not provided a clear-cut relationship between audit firm rotation and audit quality. The rotation rule may, however, be effective in contributing to the perceived quality. Specifically, this stream of research analyzes the ways in which long audit tenures and auditor switches are perceived by the market and it is conducted using both archival data and surveys with the aim of assessing the perception of the respondents on the effectiveness of the rule. The empirical evidence gathered from these studies is, in fact, useful to understand the implications that the introduction of the rule might have in terms of market effects. On the other hand, however, these studies suffer from the same limitations previously mentioned with regard to the evidence gathered on the impact on audit quality (i.e., evidence drawn from settings where the auditor change is voluntary and conflicting results).

For example, Ghosh and Moon (2005) document a positive association between perceived earnings quality and audit firm tenure, using earnings response coefficients\(^\text{10}\) as a proxy for investor perceptions of earnings quality. Thus, their results are consistent with the notion that, for investors and information intermediaries, auditor tenure improves audit quality. The same conclusion is supported by Mansi et al. (2004), who find a negative relationship between cost of debt and audit firm tenure. However, Boone et al. (2008) find that the equity risk premium decreases in the early years of tenure but increases with additional years of tenure for the then Big Five auditors. This contradicts the conclusion of the previously cited studies, as the relationship between perceived audit quality and tenure becomes negative for long audit firm tenure. In addition, the findings of Mai et al. (2008) demonstrate that shareholders view long auditor tenure as adversely affecting audit quality. All of the previously cited studies used U.S. databases. Cameran et al. (2015b) explore investor perception of audit quality in a real audit firm mandatory rotation setting (Italy) and find an increase in audit quality perception in the last engagement period.

Carcello and Reid (2014) conducted an empirical research on the perception that the news regarding audit rotation may have on the financial markets in the U.S. Specifically, they analyzed the stream of news occurring from 2011 to 2012, the time at which PCAOB was considering whether to introduce the mandatory rotation rule. Each of these events\(^\text{11}\) was linked to an increasing or decreasing likelihood of the implementation of the rule. The results indicated that the market seemed to react negatively to the possibility of introducing audit firm rotation. They also showed that the negative reaction was more pronounced for those companies with higher audit quality and for companies with a longer audit tenure.

Another part of the literature tried to assess perceptions of the rule through surveys. SDA Bocconi School of Management (2001) surveyed internal auditors, managers and the then Big Five auditors of Italian-listed companies, and reports that the first-year audit engagement requires more time by both the auditor and the client. However, the rule is judged positively by a large majority of internal auditors and managers, even though there is no evidence that the stock market reacts to news of audit firm changes [SDA Bocconi School of Management (2001, 2004)].

6. Conclusion
A crucial issue in audit regulation is whether to mandate audit firm rotation. Different countries operate different choices. For example, in Canada, mandatory firm rotation was introduced much earlier than it was in other countries (shortly after the Home Bank failure of 1923) and was in place until 1991, when it was removed. In other countries, like Spain, audit firm rotation was adopted and shortly after abrogated. One of the most peculiar evolutions of the rule was experienced in Costa Rica, where the audit firm rotation was introduced in 2005, appealed...
in 2006, ejected in 2007 and re-implemented in 2010. In recent years, the E.U. and the U.S. have been operating completely different choices in this regard. E.U. Regulation No. 537/2014 introduced an audit firm rotation rule, whereas after 10 years of debate, in 2013, the U.S. Congress ruled against the possibility of audit firm rotation.

From a theoretical point of view, mandatory audit firm rotation could have both positive and negative implications on audit quality. On one hand, it is considered a way to increase and enhance audit quality, while on the other, it is deemed costly and may cause, among other things, the loss of client-specific knowledge possessed by the audit firms.

One of the main benefits suggested by the supporters of the mandatory rotation rule is that it increases audit independence. They suggest that as time passes, members of the audit team may begin to unconsciously consider themselves as being part of the client management team and that this might impede the auditors from impartially assessing the transparency and reliability of clients’ accounts. Moreover, long-term relationship between a single audit firm and an auditee may cause external stakeholders (e.g., investors) to conclude that the auditor would not be capable of acting without bias. Apart from the independence of the auditor, another major benefit of imposing periodical changes of the audit firm emanates from what has been defined as the “new fresh look.” When the same audit firm carries out its services for a long period of time, it may fall into a sort of professional routine, whereas, when a new audit firm comes in, it approaches the financial statement of the client with “new critical eyes.” Mandatory auditor rotation would also reduce the period of the “economic dependence” on the client and this would, in turn, decrease the incentives of favoring the accounting policy of the auditee. In fact, an incumbent auditor possesses cost advantages over potential competitors, enabling it to earn client-specific quasi-rents. In order to avoid the loss of client-specific quasi-rent, the auditor has the incentive to act opportunistically, e.g., to soften audit controls, if this permits the incumbent auditor to retain the client in the future. Other benefits of imposing audit firm changes are linked to the “embarrassment effect” and the potential increase in audit market competition. Concerning the former, as the auditor knows that a new audit firm (a competitor) will come in to replace it and examine its workings, it will be incentivized to increase its effort and improve audit quality in order to protect its reputation. Concerning the latter, as most audit markets present an oligopoly structure, imposing a mandatory periodical-auditor change could reduce this phenomenon, especially if the change is from a big to a smaller auditor.

Concerning the potential negative effects of introducing the mandatory audit firm rotation rule, the most cited is the switching costs. Switching an audit firm will introduce costs that may be borne by the audit firm (e.g., familiarization with new client accounting systems) and the client (time for selecting and “training” the new auditor). Moreover, the rule causes the loss of client-specific knowledge possessed by the audit firms, since becoming familiar with a new auditee requires time. During this period of familiarization, the auditor could potentially miss material errors and mistakes, with a consequent negative impact on audit quality. Finally, in a setting where the rule is not implemented, market players may infer some kind of information (e.g., the opinion shopping attitude of the client) observing voluntary audit firm change: this is more difficult/expensive in a mandatory rotation setting where mandatory audit firm rotations also take place.

Despite the hypotheses above, however, the empirical evidence has not been so conclusive.

Even though many studies have been undertaken on this topic, at this stage, the implications of mandatory rotation are still unclear. One of the reasons for this is the mere fact that audit firm rotation is mandatory only within a small number of countries, thus making it very difficult to collect sound empirical evidence. The majority of studies conducted have focused on situations where the rule is not implemented, market players may infer some kind of information (e.g., the opinion shopping attitude of the client) observing voluntary audit firm change: this is more difficult/expensive in a mandatory rotation setting where mandatory audit firm rotations also take place.

A number of studies have also tried to determine the implications
of the mandatory audit firm rotation rule by looking at partner changes. The empirical studies focused on the partner rotation rule are once more mainly conducted in voluntary rotation settings and were unable to reach sound and univocal results.

The audit firm mandatory rotation rule seems to be an attractive solution to a very complicated issue: trust in audit activity. The well-known corporate scandals call for immediate action to restore public trust in corporate reporting. The idea of rotation has intuitive appeal, and the case for rotation is supported by anecdotal assertions like the following: “Situations where a company has appointed the same audit firm for decades seems incompatible with desirable standards of independence” [E.C. (2010)] or “Key to concern of independence was the level of ‘coziness’ the firm had with management of the company being audited... Many of the auditors of large companies... had long-running audit relationships with those companies” [PCAOB (2011)]. People contacted through a questionnaire-based survey generally agreed that the mandatory audit rotation rule constitutes a mechanism to guarantee auditor independence. From the point of view of the regulator, the enforcement in trust of audit activity is of utmost importance. But, so far, investigations into the impact of the rule at corporate and market level have not been able to prove that the benefits outweigh the costs.

References
AICPA, 1978, Commission on auditors responsibilities (Cohen Commission), report, conclusions, and recommendations, American Institute of Certified Public Accountants
André, P., G. Broye, C. Pong, and A. Schatt, 2015, “Are audit firms associated with higher audit fees?” European Accounting Review (forthcoming)
Abstract
The next decade in banking will see both evolution and revolution. Banks must reinvent themselves, not just to respond to the pressures of today, but to be flexible enough to adapt to the world of tomorrow. The most successful banks will be those that have transformed their business models. Banks will increasingly focus on profitability rather than revenues. As a result, they will be defined by narrower scope and simpler structures, but greater reach. They will serve fewer customer segments, but some will operate across more markets. They will deconstruct products, stripping them back to their component parts so that customers can rebuild them tailor-made. At the same time, they will introduce new products that are more aligned to emerging client needs. Some will move to outsource back-office functions that no longer provide a competitive advantage, and some will operate key banking services for new competitors.

Transformation is necessary because banks face an array of stakeholder pressures. They must find a way to deliver improved performance for investors who have tired of high volatility but low returns on equity. In doing so, banks will have to grapple with a low-growth environment across much of the developed world and slowing growth in the emerging world. To fund their transformation, they will need to become simpler and more efficient.

As banks position themselves to deal with a “new mediocre” era of low growth, they will have to continue to adapt to a post-financial crisis environment, where an often divergent global regulatory reform agenda shows no signs of abating and customer trust must be regained. With governments and regulators becoming increasingly assertive, and customers ever more demanding, this will not be easy. And it will be made harder in a world where the best and brightest talent is looking for a career beyond financial services.
1. The quest for profitable growth

We estimate that, if the average global bank grew revenues by 17% from FY13 levels, it would be able to deliver a 15% return on equity (RoE) without any further cost reduction. However, six years on from the global financial crisis, a global economic recovery is yet to be achieved. In fact, stagnation in the developed markets and slowing economies in the emerging markets mean that the world faces a “new mediocre” era of low growth. This creates significant problems for banks seeking to grow their way back to profitability.

Despite the prospect of continued low economic growth, we see four areas where banks will be able to generate higher revenues over the next decade: targeting new customers in emerging markets, developing new products and acquiring market share in developed markets, funding infrastructure investment and partnering with nonbanks.

1.1 Growth in emerging markets: opportunities, but not for all

The emerging markets were a focus of pre-crisis expansion for global banks, but many overreached and have had to exit some of their operations to refocus on their home markets. Despite this, we believe that, during the next five years, we will begin to see more banks expanding their geographic footprints.

The demographics are compelling. By sheer population size, the emerging world already dwarfs the developed world by five to one. By 2025, the population of Africa and Asia will increase by about 350 million and 450 million, respectively. This population will lead an increasingly urban lifestyle and its financial needs will evolve. Although economic growth in these markets may slow, they will still grow GDP by about 42% over the next five years – almost twice the rate of the developed world.1 Furthermore, these markets are set to attract increasing investment, and individuals will become more affluent. By 2025, the emerging markets will account for about 55% of the world’s disposable household income, compared with just 40% today. With this growth comes numerous opportunities for banks.

In retail banking, population growth and rising incomes will drive higher demand for transactional banking services.

1 International Monetary Fund, 2014, World Economic Outlook, October.
to operate in — with legal and regulatory frameworks that do not suit large global banks (Figure 1). An unequal regulatory playing field has accelerated the retreat of some banks from what are now seen as “non-core” markets.

In the longer term, we believe that increasingly formalized economic unions in Asia and Latin America may encourage regional banking unions. If this occurs, it will turn the tide against protectionism and make it easier for banks to operate across these regions. However, banks that have completely exited these markets over the short and medium terms may struggle to re-enter them. In the interim, therefore, it is essential that institutions with global aspirations maintain a toehold.

International banks will need to be selective about which markets they expand into and which market segments they serve. Global institutions will pursue only a few limited business lines, such as wealth management and high-value investment banking, in key markets. This will raise some questions for less advanced economies — who will finance regional trade? Who will fund growth in these markets if the local banks have limited balance sheets and domestic capital markets are not sufficiently mature? Who will provide banking services to an increasingly mobile labor force? Regional banks may be better placed to capture greater trade flows or meet retail banking demand. Alternatively, local players may increasingly have to be supported by government, becoming national champions that may themselves morph into regional champions.

Ultimately, all financial institutions must ask whether their existing product sets work for these markets. What new products might be required? Will they be able to distribute products to customers in rural areas? Will technology or partnerships (with banks or nonbanks) help them improve both product ranges and distribution networks? Expansion into new markets is not just about attractiveness but also about suitability. Banks must identify not only growth opportunities, but also where the regulatory environment will allow them to operate profitably (see Section 5). Before the crisis, banks overreached; what they do now must be different: they must be more selective.

1.2 Growth in developed markets: efficiency, market share and new products

In developed markets, the short-term focus is likely to be

profitability rather than revenue growth as banks try to address structurally higher cost bases. This is particularly important in stagnating European economies. To achieve this, banks will need to focus on improving both efficiency and productivity (i.e., cost and revenue per employee). Despite an emphasis on cost reduction, overall efficiency and productivity at many banks in developed markets have deteriorated since the crisis.

In the near term, in the low-growth economic environment of most developed markets, we expect the overall revenue pool for banks to stay constant. However, revenue growth will be achievable for individual banks that are able to increase their share of wallet. This means there will be winners and losers. The winners will be those that are able to significantly improve their customer propositions. This is highlighted in EY research showing that 30% of corporate banking customers saw one of their top challenges in dealing with banks as bureaucracy and a lack of flexibility.2 Banks will also need to improve trust. A recent report noted that 87% of U.K. small and medium-sized enterprises (SMEs) believe banks act only in their own best interests, not those of their customers.3 Recent retail, commercial and corporate banking scandals will reinforce that sentiment. Only by transforming customer experiences will banks be able to materially grow their share of wallet.

2 EY Corporate Banking Survey, 2013.
Over the medium term, banks can also achieve growth by developing products that cater to new customer needs. The rich world is getting older, and financial products that help the elderly release their wealth – often tied up in housing – will be ever more important. Although some such products already exist, they are often considered poor value. Similarly, the young in the developed world face increasing costs, including those for university and housing. With higher capital requirements, banks will have to find new ways to offer finance to such individuals – while reducing balance sheet risk, if possible.

1.3 Growth through funding infrastructure: increased fees

Infrastructure funding also offers growth opportunities. Estimates suggest that about US$57 trillion will be required for global infrastructure investment by 2030. This represents about 60% of total assets under management (AuM) in funds (Figure 2). Both developed and developing economies have major requirements for infrastructure, but banks may be challenged by the renewal of aging infrastructure in the former and development of new infrastructure in the latter. With increasingly constrained balance sheets, banks may be uncomfortable taking the risk of financing such long-term projects onto their balance sheets. However, there are still opportunities for them to grow their revenues while supporting these projects.

Banks can play a leading role in funding major projects in the early stages – when the funding demand is smaller, but the programs are not yet sufficiently mature for major fund managers to invest in. This may be through direct lending or, in regions where the capital markets are less developed, banks may do it by creating onshore investment opportunities for an increasingly affluent population.

In the later stages, rather than lending directly to infrastructure projects, banks can help direct investment from funds. They can work with broader groups of institutional investors, advising them and sharing their expertise in assessing the risks associated with such long-term projects – for example, working with sovereign wealth funds and helping them to identify opportunities to deploy their estimated US$5.2 trillion AuM.

1.4 Growth through nonbanks: new revenue streams

The fourth growth opportunity we see for banks is in partnership with nonbanks. In both developed and emerging markets, we believe there will be opportunities to grow revenues through joining forces with organizations from other sectors. The wealth of data that banks hold on their customers should facilitate a deep understanding of customer patterns and behavior. If they can unlock the value in this data, and convince customers that it is in their interests to let the banks share and use their data, banks can partner with telecommunication, technology and retail firms to offer customers real-time deals on non-financial products, earning fee revenue from partners. There will also be opportunities for banks to increase their revenues and expand their customer base by partnering with new intermediaries (see the next section). We also believe that banks will be able to generate additional revenues through fees or cross-selling, or partnerships with central and local governments – for example, as a channel for distribution of welfare payments or collection of local tax revenues, or by providing digital identification mechanisms for government programs. We already see some examples of such partnerships in the emerging markets, such as the Hunger Safety Net Programme in Kenya, and expect this will continue to develop as a revenue source.
Despite fears of an era of low growth, there are opportunities for banks to increase their revenues and boost profitability. Given past mistakes, the challenge for many banks will be to identify the most appropriate ones and to resist the temptation to overreach.

2. A new era of competition

New entrants have been nibbling at the banking value chain since before the global financial crisis (Figure 3). This has intensified following the crisis as trust in traditional banks has eroded and customers have turned to other organizations to meet their borrowing and investment needs. Some of these organizations have been welcomed by governments and regulators, particularly as a source of finance for small and medium-sized businesses. Banks must respond to defend their position.

We believe it is highly unlikely that these new competitors will be able to completely disintermediate traditional banks. It is currently unclear how far the organizations can expand their business models, and how they would deal with the inherent (and often little understood) risks in new intermediary models that have not withstood a crisis of their own. Nor is it clear that these organizations want to expand their operations so far that they attract significantly greater regulatory attention. The G20 recently announced that it would focus on addressing shadow banking risks during the coming year and, over the coming decade, we expect increased regulatory scrutiny will lead these organizations to face their own compliance problems.

However, despite the challenges they face, many of these new competitors have shown themselves to be better than banks at using customer information to deliver improved services to clients. Furthermore, greater regulatory scrutiny may even lead to greater trust in these firms, and although some new competitors may exit the marketplace, those that remain are likely to be much stronger. Banks must work out how to adjust to this new competitive environment.

2.1 Competition in retail banking

In retail banking, some of the greatest innovation in the past decade has been in payments. Alongside PayPal, which has grown into one of the world’s largest financial services organizations, a host of new innovators have emerged. Payments to small business that used to require a check can now be made by card if the business person carries a smartphone and a dongle. A number of mobile wallet applications for smartphones have been launched in recent years, allowing direct payment to individuals with only a mobile phone number.

Although mobile payments have only recently started to gather momentum in the developed world, they have long been popular in a number of emerging markets that have lacked traditional payment infrastructures. Across these markets, there are 251 live and 102 planned deployments of mobile money services for the unbanked, of which only 85 are financial service company initiatives.4

In lending and investments, a number of peer-to-peer (P2P) lending firms have emerged. Without the infrastructure and higher operating costs of traditional banks, including expensive branch networks, these P2P lenders are able to offer greater rates of return to investors and lower interest rates to borrowers.

There is also increased competition from full-service banks. For example, in the U.S., we have seen instances of direct banks offering services to customers while using the core banking systems of traditional banks. New bricks-and-mortar banks have also been launched, with a small number of full-service branches in areas with high foot-traffic. These banks have tried to differentiate themselves by offering higher standards of customer service – for example, extended opening hours or faster product application processes – even if they cannot compete fully on price.

This array of new entrants into retail banking is already forcing traditional banks to respond, whether by investing in improved customer service or the technology to deliver some of the services these new players are offering. Although we expect some consolidation of new entrants over the coming decade, they have already transformed customer expectations. We expect this to continue, with customers demanding better pricing and service. Banks will need to make the investments to meet these demands or risk losing business.

2.2 Competition in commercial banking

As in retail banking, there has been a revolution in the provision

of financial services to businesses. Not only has there been the development of P2P lenders providing debt financing, but also P2P equity investment platforms. In addition, some crowdfunding initiatives offer rewards to investors: rather than prioritizing a financial return, investors pre-purchase products or experiences.

Alongside the headline-grabbing P2P funders, other companies have emerged, including organizations that broker finance between institutional lenders and small and medium enterprises (SMEs) through online exchanges and organizations that provide supplier finance and offer online factoring and invoice discounting. Some non-financial companies are also tapping the market directly. For example, one U.K. retailer raised £50 million (approximately U.S.$77 million) from its customers through a fixed rate “retail bond” plus gift vouchers over five years. These new competitors have identified gaps in the market. Banks must decide how to exploit these gaps.

2.3 Competition in corporate and investment banking

Competition among investment banks is also increasing. With more products, such as derivatives, becoming increasingly commoditized as the over-the-counter market shrinks, institutions face greater margin compression. There is a technology “arms race” too – for example, as banks invest to reduce latency in their order management systems. Ultimately, the expense of competing, combined with declining margins, will lead to companies withdrawing from certain businesses. Only those with sufficient scale will survive.

Additionally, the demutualization of traditional exchanges has created additional competition. In response to what banks saw as increased fees, some investment banks established their own platforms to provide dealing services at lower rates than the traditional exchanges. In response, the traditional exchanges have sought to provide direct access to some investors, disintermediating the banks.

Beyond sales and trading, we expect banks to also face further erosion of their corporate banking business. As banks with constrained balance sheets struggle to lend to large corporates, we expect some funds to be ever more willing to fill this gap.

Although threatened by these new competitors, we believe banks can respond by building partnerships or making acquisitions. They can also become more effective at competing directly. Only in extremis, when they are unable to fund the investments to address this new era of competition, might they have to exit certain business lines.

3. Defining the new core of a bank

What is a bank? Fundamentally, it is an organization that takes deposits, lends and helps customers manage risks. Although this seems simple, over the past decade and more, a mix of consolidation and competition in global banking has created complex organizations, with diverse portfolios of products, operating across multiple business lines. The time has come for banks to simplify.

There is significant evidence of diseconomies of scope and scale within large universal banks (Figure 4), and we believe that a combination of the new regulatory agenda, changing customer behavior and a drive for operational efficiency within banks will force them to refocus on their core strengths. This means they will need to scale back areas of the business that do not create value or provide a competitive advantage. Some may exit business lines or turn to outside providers for particular services. As banks seek to rationalize their scope, we believe the next decade will be marked...
by the simplification of businesses, deconstruction of products and an end to the age of global universal banking.

3.1 Simplifying businesses
In the immediate aftermath of the global financial crisis, there was a chorus of bankers reaffirming their commitment to global universal banking because it helped smooth revenue volatility. A host of regulation, from structural reform to tougher capital and leverage ratios, has changed that. EY analysis shows that genuinely global banks reported average RoEs of about 7.5% in 2013, while large banks with a less diverse business and geographic footprint were able to achieve around 10.7%.

Investor sentiment will force further change. If banks are going to report single-digit returns, investors want them to be low risk. If banks are going to take on risk, investors want higher returns. And finally, banks themselves are beginning to recognize the diseconomies of scope that come with universal models, the costs involved in being a flow monster and the challenges of culture and governance associated with having an investment bank and a retail bank under one roof.

We have already seen some banks beginning to move away from the universal banking model and shrinking their investment banking divisions. Over the coming decade, we expect more banks to shrink in scope, focusing on highly profitable core businesses. Banks need to examine what is core to their business and where they create genuine value. We believe more banks will establish minimum hurdle rates of return for divisions. Parts of the business that do not create value – be it distribution channels, product manufacturing or core banking systems – will have to be improved. If they cannot be improved, banks must consider whether they are non-core and whether it would be better if they were provided to customers via an alternative supplier.

We also expect to see partnerships between domestic and international banks in emerging markets, where local institutions want to support their corporate and commercial clients as they expand overseas but lack the capabilities and capacity to do so. In return, these local banks will provide a distribution channel and fee-based revenue stream for more global institutions with broader capabilities.

As well as reassessing business lines, we anticipate banks reassessing their internal operations. Historically, banks have been poor at controlling the growth of their corporate center. Banks need to ensure that they actively control the cost of core operations. Despite the current focus on lean operations, there is a risk that when revenue growth returns, the corporate center will expand again. To help avoid this, banks should ask themselves whether a particular back-office function offers them a competitive advantage. If not, where regulatory and data-protection requirements do not prevent it, there is no reason for these functions not to be outsourced to IT and operations specialists. In some instances, this might be through a managed service arrangement but, in others, it is possible the majority of a bank’s IT and data management could be handled by a cloud-based service provider or an industry utility. Procurement and supplier management will become even more important. In the most extreme circumstances, we believe some banks with advanced systems might even offer their core banking platform, or other technologies or services from support functions as a service, to other financial institutions that are not in direct competition with them.

3.2 Deconstructing products
There is, in our experience, little correlation between the size and breadth of a bank’s product set and its market share. Furthermore, there is little evidence that product proliferation has improved choice for customers. Instead, they have been left bewildered by an array of products with a variety of terms and conditions, few of which are suited to them individually. Although many banks have started talking about a customer-centric (rather than product-centric) approach, few have achieved it.

The product-centric approach has done little to benefit banks. They have had to grapple with the costs of managing and distributing all those products and increased conduct risks; the complexity and diversity of products has contributed to recent misselling to both business and retail customers. This has eroded trust in institutions, while the concomitant fines have had a direct impact on bank profits – for example, U.S. banks have been fined around U.S.$90 billion for mortgage and asset-backed security mis-selling while payment protection insurance and interest rate swap mis-selling, has cost U.K. institutions an estimated U.S.$40 billion in fines and litigation.

Global regulators are increasingly focused on customer protection
and product suitability. The challenge for banks is to ensure their products are suitable when regulators are unwilling to approve individual products up front, but willing to penalize banks retrospectively. If banks are not confident of certain products’ suitability, they risk further misselling scandals in the future. This is particularly true as banks look to develop new products to meet the developing needs of customers (see section 1). It is compounded at institutions where banks have vast product sets that sales teams are unable to understand in detail. This means that staff are more likely to make mistakes when advising customers, thus increasing conduct risk.

We believe that for banks to be truly customer centric and to avoid the risks of misselling, products must become customer driven. A customer-driven approach is already seen in other industries, where customers are getting used to more personalized products and experiences. This has set customer expectations, and banks must adapt to match them.

A move to a customer-driven approach will require banks to rationalize and simplify their product sets. By breaking products into their component parts, banks will enable customers to tailor-make products to suit their needs. Banks that achieve this will be able to provide customers with truly differentiated offerings. EY’s Global Consumer Banking Survey 2014 suggests that banks can achieve it will be able to win an increased share of wallet (Figure 5). They will also be able to reduce the cost of their operations and better satisfy regulators that they are treating customers fairly. However, achieving it will require reform of banks’ internal systems, processes, operations and culture. Banks currently tend to view each variant of a product as a separate product – for example, mortgage accounts with different interest rates would be treated as separate products from a system perspective and would have individual manufacturing, marketing and distribution processes associated with them. Banks will need to change the way they view products – so they become simple “base products” but have dynamic attributes (such as interest rates or fees) associated with them.

Just as importantly, banks will also need to develop a clearer understanding of the value of a product across its life cycle to ensure that these customer-driven, dynamic products are priced appropriately. This may be the biggest challenge for most banks. Typically, many institutions have a poor understanding of the full profitability of products, exacerbated by ineffective cross-charging models, complex business structures and poor data analytics. However, a combination of simpler businesses and rationalized product sets should give these institutions a better understanding of profitability, enabling them to completely transform their customer offerings.

We believe that as banks respond to consumer, regulatory and investor pressures to simplify, it will become more critical that they understand where value is created in their organizations.

4. Technology reshaping banking

In the past decade, technology has completely transformed banking. Over the next decade, it will continue to do so.

While we do not anticipate many banks making the investment to replace their legacy core banking systems any time soon – in the past decade, barely a handful of banks have replatformed – we do believe that achieving a digital and technology transformation will require continued investment in middleware. This is critical to drive efficiency, productivity and speed to market. However, to ensure real change, this investment must be strategic rather than solely for tactical fixes.
With only 24% of global IT spend going to investment (less than 20% in Europe), the choice of where to invest will be even more difficult. Nevertheless, it is clear that technology will not only reshape the experience of banking customers, but will revolutionize institutions’ internal processes, making them more efficient and more productive.

4.1 Reshaping the customer experience
Many of the changes that banks must make to the customer experience will be driven by the growth in mobile and smartphone ownership, which are transforming customer expectations. Mobile payments have already transformed financial services in some emerging markets, where mobile phone ownership significantly outstrips penetration of bank accounts. In fact, in some emerging markets, ownership of smartphones is already beginning to exceed bank account penetration. It is only six years since apps first appeared on mobile devices, and customers now expect to be able to check their balance and make transfers through their phones, at the minimum. Online banking has evolved too – with some banks even offering banking services through Facebook.

Customer expectations have moved beyond “omnichannel” – which has generally been seen by banks as a more consistent multichannel strategy. They now want to interact with their bank whenever they want, however they want and wherever they want, and to be able to shift seamlessly between channels. Customers want a ubiquitous experience. In the next decade, we expect to see banks increase their investment in social media as this becomes a more common route to interact with banks. We expect technologies such as check and card imaging to become commonplace. We also believe multiday cycles for reissue of checkbooks or cards will no longer be acceptable. While some newer, more efficient institutions already have the capability to issue checkbooks and cards in-branch, the evolution of technologies such as digital printing may eventually enable customers to issue their own cards at home – if cards have not been usurped by mobile wallets.

In fact, with smartphone penetration across a sample of 48 developed and emerging markets at almost 45% (up from around 27% in 2011), mobile wallets may become the primary way customers access their bank accounts. At their most advanced, they could offer a single gateway to access multiple accounts across multiple banks. Increased use of smartphones, combined with contactless technology, will potentially transform remittances

---

Transforming banking for the next generation

— for which combined transaction fees and exchange rate margins can still cost more than 10%. They will also transform cash management in these markets. Should mobile wallets with enhanced biometric security become the norm, the cost of obtaining banking services in emerging markets will be dramatically reduced and penetration rates significantly increased.

However, it is not all about a shift to digital channels. While we expect self-service machines to replace cashiers and mobile wallets to become commonplace, we believe branch staff will still have a critical advisory and sales role. EY’s Global Consumer Banking Survey 2014 highlights that, while digital channels are the most frequently used, they are principally for transactional banking. Customers still rely heavily on personal branch staff for advice and buying and selling investments (Figure 6).

Nevertheless, the roles of branch staff in retail banking, as well as relationship managers in commercial, corporate and private banking, will be transformed by technology as tablets are increasingly used to deliver their training, to provide customer account information and to ensure they stay within a predetermined sales process. In such instances, technology will not replace people; it will enhance their capabilities and change the ways customers and staff interact.

Critically, as customers expect greater real-time services, some improvements will require cross-industry collaboration. We believe some of the greatest investments will be in real-time payments. The recent advances in mobile payments mean more customers expect to be able to transmit funds instantaneously.

However, these mechanisms only support relatively small transfers. We believe banks in a number of countries will make significant investments to develop real-time, high-value customer payment.

Better use of customer data will also transform the customer experience. Banks will be better placed to understand customer needs and behaviors. They will have greater connectivity with customers who, EY’s Global Consumer Banking Survey 2012 revealed, would be more willing to give more data to their bank if it led to their being offered more suitable products. With mobile banking apps potentially having access to GPS data, banks will in fact be able to offer personalized, real-time and location-based offers to their customers (see section 1 above).

4.2 Reshaping banking processes
Changes in customer expectations will force banks to invest in their core systems and processes to make them fit for purpose. But technology spend will not only be driven by customers. Banks will need to make some tough choices about investment priorities.

Banks will deploy a raft of new technology internally. For example, we anticipate increased automation and digitization of processes, reducing the need for manual intervention. We expect greater use of “gamification” to help banks improve governance, and increased use of monitoring and surveillance tools to reduce conduct-related risks. Similarly, the deployment of digital tools such as tablets will help monitor and ensure adherence to sales processes and reduce the scope for misconduct.

Banks will also transform the way they use data internally as they respond to demands from multiple sources. Teams in banks increasingly need more and better-quality data to drive decision-making — whether around resolution of customer issues or internal risk management issues.

Data will be particularly important in dealing with cybersecurity issues, which our European Banking Barometer 1H14 revealed as a top-three priority for major banks. EY’s Global Forensic Data Analytics Survey 2014 showed that 72% of respondents believed big data technologies had a role to play in fraud prevention and detection — but only 2% are already using them.

At the same time, banks are facing increased demand for data from regulators; for example, for the recent asset quality review in Europe, or for the comprehensive capital analysis and review in the U.S. Institutions are being asked to make increasingly frequent submissions of increasingly granular data. It has been argued by many that banks should look beyond merely complying with regulations to using these requirements to their advantage to support changing internal processes. To date, the volume and lack of clarity of regulatory change, combined with the cost of dealing with the aftermath of the global financial crisis, have made this difficult. As most post-crisis regulation becomes business as usual, banks will finally be better able to think strategically about how to transform their internal processes and how technology and data will help them do this.
It is clear that technology will reshape banking over the next decade. The challenge for the industry is to shift the balance of technology spend between maintenance and investment to drive internal change and protect against non-traditional competitors.

5. Defining the structure of a bank
Since the global financial crisis, banks have faced increased regulatory burdens. The pressure to reduce risk and complexity, and to enhance resolvability, is forcing banks to rethink their operational structures. Already, some banks have reduced their global footprint. However, increased pressure to grow revenues means many banks will remain committed to offering diverse services and operating across multiple markets. To overcome the operational and regulatory challenges this will create, banks need to ensure they have structures and operating models that are lean and flexible.

5.1 Operating across multiple markets
Before the financial crisis, cumbersome and complex legal and operational banking structures grew up as by-products of overseas expansion, acquisitions, changing tax legislation and the development of new banking products. The traditional approach of banks to operating diverse businesses across the confused landscape of divergent regulatory regimes has been to match complexity with complexity. However, we believe the tide is turning, and leading banks will be those that simplify their structures. In fact, the introduction of recovery and resolution plans (RRPs) in key jurisdictions will compel some banks to reappraise and simplify their entity structures, but we believe that, even when not mandatory, there are benefits in doing so.

The pre-crisis period was one of legal entity proliferation. It is now time to rethink current structures that may no longer fulfill their intended purpose and may magnify risks, lead to inefficient use of resources and unnecessary costs and, ultimately, erode value. Before the financial crisis, one global bank had in excess of 5,000 legal entities across the globe. Such a structure is expensive to maintain. We estimate that the average “carrying cost” of a single legal entity can range from U.S.$90,000 a year for a small entity to over U.S.$600,000 a year for a large entity. Even more troubling, our experience suggests that few at board level truly understand these structures.

Banks that take the opportunity to simplify their structures will benefit from increased operational efficiency, cost savings and compliance benefits. They will also be better positioned to operate and grow in a world where regulatory boundaries are less clear, regulatory demands are increasing and regulators are increasingly assertive and less tolerant of excessively complex organizations.

We also believe that, regardless of business model or geographic footprint, many banks have suboptimal operational footprints. They must overhaul their operating models. Their core is inefficient and overly complex (see section 3 above). Many banks operate business lines supported by their own separate support functions. We believe that, by breaking out these functions and reshaping them to serve the broader organizations, institutions will be able to drive greater efficiency. They will also build institutional resilience and give regulators greater confidence around their resolvability.

As banks redefine how they operate, they will need to redefine where they operate. We believe that the offshore footprint of global and regional banks is often suboptimal. In our experience, there has been an optimistic bias among banks (and other institutions) when making outsourcing and offshoring decisions. The expected cost savings are, often, initially lower than expected. Furthermore, it typically takes longer than originally anticipated to deliver those savings and to ensure robust processes. In fact, for many banks that have offshored activities, the price of achieving lower costs has been unexpectedly high. Many emerging markets have seen wage inflation that has reduced labor arbitrage benefits. Furthermore, many customers and employees have been disappointed by the quality of offshored services.

Finding the right combination of nearshoring and offshoring will be critical to banks’ expansion plans. They need to look beyond pure cost factors to develop operations that support competitiveness, innovation, growth and flexibility across back-, middle- and front-office functions.

Regulation will clearly play a part in deciding where to locate operations and activities. Banks must consider global and local regulatory environments – including concerns about how control of data and activities in an RRP world will evolve in the future. Furthermore, they must be confident about the resilience of
local infrastructure and inflation expectations, as well as political and country risks. The availability of talent will also be critical to location decisions.

Our recent discussions with banks across a range of emerging markets have identified shortages of skilled staff as a frustration to their growth ambitions. Such shortages will also become an increasing problem in developed markets, where fewer high-quality graduates want to work in finance. Banks across the world must find new ways to attract leading talent and deploy it where it is most needed.

As banks rethink the markets in which they want to do business over the next decade, so too must they re-examine their operational footprint. It is clear that banks need to simplify their structure, but they must do so in a way that helps achieve efficient growth.

6. Conclusion
Existing pressures from stakeholders, combined with the megatrends that are reshaping the world, will transform banks and their business models over the next decade. In some respects, they will extend their reach — expanding into new markets, targeting new customer needs — with more decentralized operations. In many other respects, they will be leaner — nimbler and more responsive to customer demands, refocused around core businesses and capabilities, and “buying in” non-core capabilities.

Redesigned as more flexible institutions, banks will not only be better able to adapt to the changing world, but will also be better placed to respond to unforeseen events in the future. The next decade will be one of transformation in banking. The leading banks will be those that redefine their core operations to support this change and build in the agility to allow them to cope with unexpected global shifts.

6.1 New business models
As banks look to grow, they will have to serve new customers — the unbanked in the emerging world and the underserved in the emerging and developed worlds. They will have to develop new products to meet evolving customer needs and find new ways of financing large infrastructure projects. They will also have to partner with other financial institutions, as well as nonbanks, to leverage new technologies and generate fees. In addition, they will have to re-engineer to optimize efficiency and productivity, particularly in developed markets.

It will not be a straightforward transition to these new business models. Banks must deliver business-wide transformation. This will be an iterative process as models are continually refined and banks respond to unforeseen developments. As banks embark on these multiyear change programs, they must understand the political and regulatory environment better, they must understand their customers better and, above all, they must understand themselves better.

First, banks must transform in response to regulation. Simplification will be central to this. Banks that take the opportunity to simplify their structures will benefit from increased operational efficiency, cost savings and compliance benefits. They will also be better positioned to operate and grow in a world where regulatory boundaries are less clear, regulatory demands are increasing and regulators are increasingly assertive and less tolerant of excessively complex organizations.

Regulation is also forcing banks to change the way they use and handle data. Yet, institutions that use data and enhanced analytics not just to meet regulatory requirements, but to create insights to meet customer needs, to understand product-
level profitability and to deliver operational excellence will be transformed.

Better data usage will be the foundation for better business models. It will give banks an understanding of their own core competencies. Knowing where they can create additional value and which competencies are missing – and with a clearer view of costs across the organization – will leave banks better placed to understand which operations they run efficiently and which operations they do not. Based on this understanding, they can seek additional revenues, whether by partnering with other institutions to provide specific customer products or procure managed services, or building alliances and collaborating across the industry to develop industry utilities to help them plug their own capability gaps.

Doing this will also require banks to transform their core technology, investing strategically to deliver business objectives and to circumvent the challenges of legacy systems. They must transform digitally – while mindful of cyber risks – to improve their customers’ experience, to improve customer segmentation and to enhance customer economics. Mobile channels will be central to all of these, especially if mobile can provide a single point of access to multiple bank accounts across multiple providers. However, this will not completely usurp person-to-person banking relationships, and banks must also harness digital to support more effective sales and advisory staff.

It will also be necessary to transform controls and compliance. In the short term, banks must enhance controls across key business processes and strengthen the front office, control function and internal audit to ensure improved compliance. Over the longer term, banks must become more radical by transforming their culture.

Cultural transformation will drive employee behavior that is better aligned to the interests of all a bank’s external stakeholders – governments and regulators, investors and customers. It will also make banks more attractive places to work for a new generation of employees.

But to achieve this, it will be necessary to transform employee propositions. To secure leading talent and comply with regulation, banks must focus on creating compelling propositions for employees that are not based solely on higher-than-average salaries. At an aggregate level, there appears to be little correlation between pay and performance for most bank employees. Furthermore, the “lifestyle” plans proposed by some leading global banks suggest that leadership is out of touch with its millennial recruits. A war for talent between nonbanks and banks is already intensifying, and tactical responses are unlikely to tip the balance back in favor of banks.

This will allow banks to differentiate themselves in the marketplace and develop ways to drive productivity and efficiency. Ultimately, the leading banks 10 years from now will not necessarily be defined by the very best products and services or the most efficient operations, although they will score highly on all these counts. They will be defined by their ability to manage the risks of change programs and to make the right investments in products and services. They will be defined by their ability to create an internal culture that wedds dynamism to best practice. They will be defined by their ability to deliver new fit-for-purpose business models.

The most successful banks will be those that master transformation.
Part 2: Tactical

Effectiveness of loan-to-value ratio policy and its transmission mechanism: empirical evidence from Hong Kong


Recursive collective action problems: the structure of procyclicality in financial and monetary markets, macroeconomies and formally similar contexts

Do “too-big-to-fail” banks take on more risk?

An overview of the risk-neutral valuation of bank loans

Smart beta: too good to be true?

Regulatory herding versus democratic diversity: history and prospects

Squandering home field advantage? Financial institutions’ investing in their own industries

The new Banking Union landscape in Europe: consolidation ahead?
Effectiveness of loan-to-value ratio policy and its transmission mechanism: empirical evidence from Hong Kong

Eric Wong
Market Research Division, Research Department, Hong Kong Monetary Authority

Kelvin Ho
Market Research Division, Research Department, Hong Kong Monetary Authority

Andrew Tsang
Market Research Division, Research Department, Hong Kong Monetary Authority

Abstract
This paper provides a non-technical summary of two recent empirical studies to shed light on key important issues regarding the implementation of loan-to-value (LTV) policy as a macroprudential tool, including its effectiveness, potential drawbacks and its transmission mechanism to improve financial stability. Empirical evidence suggests that LTV policy is effective in reducing systemic risk associated with boom-and-bust cycles in property markets. Although the LTV policy may be associated with higher liquidity constraints on homebuyers, we show that the mortgage insurance program (MIP) can mitigate this drawback without undermining the effectiveness of LTV policy. Thus, MIPs play an important role in enhancing the net benefits of LTV policy. Concerning the transmission mechanism, empirical evidence suggests that the policy pass-through to property market activities may be weak. By contrast, there is clear evidence that tightening the LTV cap would reduce household leverage and credit growth, and that lower leverage plays a major role in strengthening banks’ resilience to property price shocks. This finding supports the view that household leverage would be an optimal target of LTV policy.

1 The views and analysis expressed in this paper are those of the authors and do not necessarily represent the views of the Hong Kong Monetary Authority.
1. Introduction

Although there is a growing consensus that regulating maximum LTV ratios on mortgages (henceforth referred to as LTV policy) could help contain systemic risks associated with credit-property price spirals, empirical evidence with regard to some fundamental issues of this macroprudential tool remains scant. First, how effective is LTV policy in reducing systemic risk? Second, does LTV policy create significant liquidity constraints for potential homebuyers? Third, can an MIP offset this drawback of LTV policy without undermining the effectiveness of LTV policy?

Drawing on the empirical findings by Wong et al. (2011), sections 2 and 3 of this paper answer these questions based on Hong Kong’s experience with both LTV policy and MIPs, and econometric analysis of panel data from 13 economies respectively. In section 2, we present the history of LTV policy in Hong Kong, showing strong evidence that it has helped the Hong Kong banking sector weather the boom-and-bust cycles of the property market. We also present evidence that Hong Kong’s MIP has helped homebuyers overcome the liquidity constraints they may experience because of LTV policy without increasing banks’ credit risk. In section 3, a summary of the panel econometric analysis is provided. The result bears out Hong Kong’s experience and shows that LTV policy is effective in reducing systemic risks associated with credit-property price spirals, and that MIPs have not reduced the effectiveness of LTV policy.

While the empirical findings point to a pivotal role for LTV policy in safeguarding banking stability, there remain debates about the design of the operational framework. In particular, would LTV policy be optimal to target household leverage, credit growth or property prices in pursuit of financial stability? Theoretically, the answer crucially depends on the transmission mechanism of LTV policy, particularly (i) the extent of the policy pass-through on these three variables and (ii) their respective contributions to financial stability.

To shed light on this issue, section 4 investigates the short-term policy impact on property market activities for three selected economies, Hong Kong, Korea and Singapore. Section 5 provides novel empirical evidence from Hong Kong based on Wong et al. (2013) on the pass-through of LTV policy to borrowers’ leverage and credit growth, and their respective contributions to financial stability. Overall, the empirical results in these two sections support the view that the dampening effect of LTV policy on household leverage is more apparent than its effect on property market activities, suggesting that it would be optimal for LTV policy to target household leverage.

2. A brief history of LTV policy and the MIP in Hong Kong

LTV policy has long played a vital role in strengthening Hong Kong banking system’s resilience to asset price volatilities and in reducing the risk of cycle amplification through bank credit, rather than as a means of managing asset price cycles and market activities or targeting asset prices. Figure 1 provides a...
succinct visual summary of the developments in Hong Kong’s LTV policy, together with changes in property prices and mortgage delinquency ratios. The development of LTV policy in Hong Kong can be broadly divided into four phases, as summarized here.

Phase 1: before 1997
Prior to the adoption of the LTV policy in 1991, authorized institutions (AIs) in Hong Kong were allowed to grant mortgage loans covering up to an LTV ratio of 90% under the Third Schedule of the Banking Ordinance, the legal framework for banking supervision in Hong Kong. In view of the systemic risk that could arise from residential mortgage lending (RML), the Commissioner of Banking had intended to amend the Third Schedule to lower the 90% LTV threshold to 70%. The Commissioner of Banking consulted the banks on these intentions during 1991. Banks were cooperative, offering to adopt the 70% LTV policy voluntarily, removing the need to amend the Third Schedule. The 70% maximum ratio has since been fully endorsed by the Hong Kong Government as a prudential measure and has evolved into a banking industry standard intended to guard against overexposure to the property market.

Phase 2: from 1997 to 1999
Against the backdrop of a sharp rise in residential property prices in 1996, signs of speculative activities and the rapid increase in RML, the Hong Kong Monetary Authority (HKMA) issued guidelines to all AIs, on 28 January 1997, recommending the adoption of a maximum LTV ratio of 60% for properties with a value of more than H.K.$12 million – “luxury” properties. In the wake of the Asian financial crisis, Hong Kong’s property prices fell significantly – by more than 40% from September 1997 to September 1998 – yet the mortgage delinquency ratio remained below 1.43%, which is low by international standards. This fact alone suggests that LTV policy reduces the credit risk faced by banks and assures the quality of banks’ mortgage loan portfolios.

Phase 3: from 1999 to 2008
In line with measures intended to stabilize the property market implemented by the Government, the HKMA restored the maximum LTV ratio of 70% in October 2001. Meanwhile, the HKMA allowed AIs to refinance the mortgage loans of homeowners with negative equity for up to 100% of the current market value of the mortgaged property. Notwithstanding this relaxation, the HKMA emphasized that the 70% LTV policy remained as a long-term prudential measure.

Because of a significant drop in household income after the Asian financial crisis, prospective homebuyers faced significant obstacles in the property market, which led to calls for the relaxation of the 70% LTV policy. In 1999, the Hong Kong Mortgage Corporation (HKMC) launched an MIP, under which mortgage loans of up to an LTV ratio of 90% were made available to homebuyers meeting certain eligibility criteria. The MIP is designed to protect participating banks against credit losses on the portion of loans that exceeds the 70% threshold in the event of defaults. It also avoids the potential drawback of LTV policy that some homebuyers may not qualify for a mortgage because of substantial down payment requirements, even if they have sound financial conditions. The increase in the number of homebuyers participating in the MIP during this period demonstrates that the MIP has helped a significant number of households overcome liquidity constraints (Figure 2) and that concerns about liquidity constraints should not be lightly dismissed.

One concern is that the MIP may reduce the effectiveness of LTV policy because it enables households to increase their leverage ratios, thereby boosting the risk of mortgage defaults, in theory, and hence of bank credit losses. In reality, however, the HKMC’s MIP portfolio registers a lower delinquency ratio than Hong Kong’s banking sector, indicating that, thanks to prudent underwriting criteria, the MIP has not undermined the LTV policy but has actually improved the stability of Hong Kong’s banking system.

---

4 AIs are institutions authorized under the Banking Ordinance to carry on the business of taking deposits. All AIs in Hong Kong are supervised by the HKMA.
5 The HKMA was established on 1 April 1993, by merging the Office of the Exchange Fund and the Office of the Commissioner of Banking. Its main functions and responsibilities are governed by the Exchange Fund Ordinance and the Banking Ordinance. Banking supervision is one of its functions.
6 On 2 November 1995, the Hong Kong Government confirmed at a Legislative Council meeting that the 70% LTV ratio should be adopted as a long-term regulatory policy.
7 For details, see HKMA (2009).
8 The HKMC, which is owned by the Hong Kong Government, was established in 1997. Its primary missions are: (1) enhancing the stability of the banking sector by serving as a reliable source of liquidity, thereby reducing the concentration and liquidity risk of mortgage lending by banks, (2) promoting wider home ownership and (3) facilitating the growth and development of the debt securities and mortgage-backed securities markets in Hong Kong.
9 The criteria include maximum levels for the debt-to-income ratio, loan amounts and maturities. The delinquency ratio of the HKMC’s MIP portfolio reached a historical high of 0.39% at the end of September 2003, whereas the ratio for the Hong Kong banking sector was 1.05%.
10 The Journal of Financial Perspectives
Phase 4: 2009 to the present

Strong capital inflows and unusually low interest rates since early 2009 have fueled property prices in Hong Kong sharply, particularly at the upper end of the property market. As a prudential measure, the HKMA issued guidelines in October 2009 requiring all AIs to reduce the maximum LTV ratio for properties with a value of H.K.$20 million or more from 70% to 60%.

To further safeguard banking stability and help banks manage credit risk more prudently, the HKMA tightened LTV limits with five subsequent rounds of prudential measures between August 2010 and February 2013. As a result, the current maximum LTV ratio is set to be 50% for residential properties with a value of H.K.$10 million or more, 60% for those with a value between H.K.$7 million and H.K.$10 million and 70% for those with a value below H.K.$7 million. The maximum LTV ratio for non-owner-occupied residential properties and properties held by a company is 50%, regardless of property values. For commercial and industrial properties, the maximum LTV ratio is 40%. For net-worth-based mortgages, the maximum LTV ratio is 40% for residential properties and 30% for commercial and industrial properties.

Regardless of property types or values, the maximum LTV ratio is lowered by 10% points for mortgage applicants whose income is mainly derived from outside Hong Kong. If the mortgage loan applicants already have outstanding mortgages, a further 10% point deduction in maximum LTV ratio is applied if their income is mainly derived from outside Hong Kong and if they apply for net-worth-based mortgages.

3. Panel econometric evidence on the effectiveness of LTV policy and MIPs

This section provides cross-country evidence on the effectiveness of LTV policy and MIPs based on econometric analyses of panel data from 13 economies. This study aims to test two hypotheses. First, countries with LTV policy tend to have lower sensitivity of mortgage default risk to property price shocks and second, MIP does not reduce the effectiveness of LTV policy. To this end, the economies are separated into three groups: (i) with both LTV policy and MIP, (ii) with LTV policy but no MIP and (iii) no LTV policy.

Two fixed effects models, which postulate that the change in the mortgage delinquency ratio is negatively correlated with real property price growth and real GDP growth, are estimated. To facilitate comparisons between the three groups of economies, the estimated coefficient for property price growth, and that for real GDP growth, are assumed to differ across the groups.

Regarding the first hypothesis, the sensitivity of the mortgage delinquency ratio to property prices is found to be negative and lower (in absolute terms) in economies with LTV policies than those without. A 1% drop in property prices would increase the delinquency ratio by 0.35 basis points in economies with LTV policies, and by 1.29 basis points in economies without LTV policies. Two fixed effects models, which postulate that the change in the mortgage delinquency ratio is negatively correlated with real property price growth and real GDP growth, are estimated. To facilitate comparisons between the three groups of economies, the estimated coefficient for property price growth, and that for real GDP growth, are assumed to differ across the groups.

Figure 2: Annual drawn down loan amount and usage rate of the MIP in Hong Kong

Note: the MIP usage rate is defined as the ratio of the mortgage loan amount drawn down under the MIP to the total mortgage loan amount drawn down in the Hong Kong banking sector. All currencies are in Hong Kong dollars, unless specified otherwise. Source: HKMC

11 Including stand-alone car park space.

12 The economies included in this study are Australia, Canada, Greece, Hong Kong, Korea, Malaysia, the Philippines, Portugal, Singapore, Spain, Thailand, the U.S. and the U.K. The sample covers the period Q1-1991 to Q2-2010.

13 Hong Kong, Korea, Malaysia and Singapore have adopted LTV policies, according to the Bank for International Settlements (BIS) (2010) and information obtained from their respective central banks/supervisory authorities.

14 An aggregate mortgage-to-GDP ratio and the change in interest rate are included as control variables. The effect of other institutional factors, such as recourse rules and personal bankruptcy regulations, are assumed to be captured by country fixed effects.
To visualize the policy effect on banking stability, a simulation exercise is conducted for Hong Kong’s banking sector. The simulation counter-factually assumes that a relaxation of the maximum LTV ratio from 70% to 90% was to occur before 1997. All banks are assumed to aggressively exploit this relaxation to extend mortgage loans to an average LTV ratio of 90%. A shock of a 40% decline in real property prices is then imposed in the simulation. The simulation (Figure 3) suggests that, if the maximum LTV ratio were to have been relaxed before 1997, the delinquency ratio after the property price shock would have been 1.7% (at the 95% confidence level) at the end of 1998 compared with the actual level of 0.84%.

There is supporting evidence for the second hypothesis. Specifically, empirical results find that MIPs are not a significant determinant of the sensitivity of the mortgage delinquency ratio to property prices for economies with the LTV policy, indicating that MIP should not reduce the effectiveness of LTV policy.

4. Empirical evidence on the policy pass-through to property market activities

Regarding the transmission mechanism of LTV policy to improve financial stability, this section investigates whether LTV policy is an appropriate tool for stabilizing property market activities. We assess this issue by studying Hong Kong, Korea and Singapore, where LTV caps were tightened in the past when there was concern about the risk of overheating property markets. By quantifying the immediate effect of the tightening LTV caps on property market activities in these economies, the policy pass-through to property market activities can be evaluated.

For this purpose, a time series model is separately estimated for the three selected economies using data from Q1 1981 to Q2 2010. In the model, property market activities are explained by their lag terms, real interest rates and two dummy variables. These variables are constructed to test the change in the average value of property market activities between the six-month period before the tightening LTV caps and after the tightening. The model is applied for estimating the policy impact on (i) the real property price growth, (ii) the deviation of actual property prices from their trend and (iii) the number of property transactions (in logarithmic form).

The estimation result suggests a significant dampening effect of LTV policy on real property price growth only for the case of Hong Kong. For the other two property variables, their movements are found to be insensitive to LTV caps tightening in all the three economies. Taken together, the policy pass-through to property market activities may be weak. A similar assessment, however, finds a significant dampening impact of LTV tightening on mortgage debt leverage (measured by the ratio of mortgage loans to GDP) for all the three economies, suggesting that the effect of LTV policy on systemic risk may be primarily transmitted through effects on household leverage.

5. Empirical evidence on the policy pass-through to leverage and credit growth

This section attempts to advance our understanding of the transmission mechanism of LTV policy further by quantifying the policy impact on borrowers’ leverage and credit growth (i.e., henceforth referred to as the direct and indirect effects respectively) and their respective contributions to financial stability using data from Hong Kong from Q2 1999 to Q4 2012.

15 The trend level is derived using the Hodrick-Prescott filter. The gap is expressed as a percentage of the actual level of property prices.

16 Theoretically, the direct effect improves the resilience because mortgagors would hold a larger equity buffer at origination, contributing to a lower likelihood of negative equity and thus lower default risk. The indirect effect primarily avoids banks underwriting excessively fresh mortgage loans, which are generally subject to higher default risks due to a relative low portion of equity.
The empirical analysis helps assess whether household leverage or credit growth would be the optimal target of LTV policy.

5.1 The significance of the direct effect

One salient feature of LTV policy is that authorities operate the tool by adjusting the maximum LTV ratio instead of the actual LTV ratio in the market (LTV\textsubscript{M}). Theoretically, however, LTV\textsubscript{M} is determined together with other mortgage terms both by banks and homebuyers [Zumpano et al. (1986)]. An important question, therefore, is to what extent the LTV cap, serving as a macroprudential instrument, would be factored in when banks and homebuyers determine LTV\textsubscript{M}.

To answer this question, a regression model of the determinants of LTV\textsubscript{M} is estimated. The model postulates that a higher LTV cap, a higher property price return (relative to its volatility) and rental yield and a lower debt-servicing ratio tend to be associated with a higher LTV\textsubscript{M}, as banks and mortgagors tend to accept a higher LTV ratio when the property market is buoyant, debt servicing burden is low and the LTV cap is less stringent.

The empirical result is broadly in line with our expectation. In particular, the LTV cap is found to be one binding factor affecting LTV\textsubscript{M}, with the long-run elasticity being estimated to be 0.33. To gauge the responsiveness of LTV\textsubscript{M} to the LTV cap, Figure 4 shows the cumulative contributions of the determinants to the change of LTV\textsubscript{M} since September 2007. The five rounds of LTV cap tightening from 2009 to 2012 are estimated to reduce LTV\textsubscript{M} significantly. Figure 5 plots the actual LTV\textsubscript{M} against a hypothetical series of LTV\textsubscript{M} under a counterfactual “no policy” scenario, suggesting that, had the HKMA not tightened LTV caps, LTV\textsubscript{M} may hover above 60% instead of 52% at the end of 2012, indicating a significant direct effect.

5.2 The significance of the indirect effect

The significance of the indirect effect is studied by estimating the demand and supply equations for mortgage loans in Hong Kong with a framework that allows for, but does not impose, disequilibrium. This empirical specification reflects two considerations. First, theoretically, LTV policy is likely to affect both demand for and supply of mortgage loans. Estimating a demand-supply system could, therefore, facilitate a clearer identification of the policy impact. Second, loan market

---

17 The LTV\textsubscript{M} under the counterfactual “no policy” scenario is estimated by removing all the policy impact of LTV cap tightening (i.e., the dark red area in Figure 4).

18 LTV cap tightening may reduce demand for mortgages, as homebuyers may be forced out of the property market because of higher liquidity hurdles or lower returns on equity for property investment. LTV cap tightening may also reduce credit supply because it may lead banks to lend less than they otherwise would.
disequilibrium, which has been widely documented in both the theoretical and empirical literature, could contribute to a state-dependent effect of LTV policy on credit growth (see Figure 6 for illustration).

Table A1, in the Appendix, presents the specification of the demand and supply equations for mortgage loans. The estimation result is summarized as follows:

- The estimation result reveals that disequilibrium could occur in the Hong Kong mortgage market, suggesting that


20 The two-stage least square (TSLS) method is adopted to estimate the model instead of applying the method of ordinary least squares because estimators of the latter are statistically inconsistent.

the demand or supply can be the sole binding factor in determining the credit volume.

- \( \text{LTV}_i \) is found to be a significant determinant of both the demand for and supply of mortgage loans, with the policy impact on loan supply being estimated to be significantly larger than that on loan demand. Figures 7 and 8, which show the supply and demand estimates respectively under the actual and counterfactual “no policy” scenarios, find that the dampening effect of the five rounds of LTV cap tightening from 2009 to 2012 is much larger on the loan supply than on demand.

- Estimation results (a) and (b) together suggest that the effect of LTV policy on loan growth in Hong Kong is likely to be state dependent, such that LTV policy is more effective in limiting credit growth when there is excess credit demand, but less so when excess credit supply occurs, suggesting a state-dependent feature of the policy effect.

Figure 6: A supply-and-demand diagram to illustrate the effect of LTV policy under scenarios of excess supply and excess demand in loan markets.

Note: the figure illustrates the implication of loan market disequilibrium for the effect of LTV policy on credit growth. Assume that LTV cap tightening shifts the demand from \( D \) to \( D' \) moderately and supply from \( S \) to \( S' \) more significantly, in case 1, where demand exceeds supply (implying credit supply is the binding factor) at the prevailing mortgage interest rate \( i_L \), the effect of the tightening solely reflects the supply-side impact, while the demand-side impact is invisible. In this case, the loan volume decreases considerably from a to b. In case 2, where supply exceeds demand at the prevailing mortgage interest rate \( i_H \), the effect of the tightening solely reflects the demand-side impact, while the supply-side impact is invisible. The loan volume decreases marginally from c to d. In this hypothetical case, LTV policy is expected to be more effective when there is excess credit demand but less so when excess credit supply occurs, suggesting a state-dependent feature of the policy effect.

Figure 7: Estimated supply of mortgage loans under the actual and counterfactual “no policy” scenarios
Effectiveness of loan-to-value ratio policy and its transmission mechanism: empirical evidence from Hong Kong

5.3 The contribution of the direct and indirect effects to financial stability

To quantify the contribution of the direct and indirect effects to strengthening the banking sector resilience to property price shocks, we first estimate an econometric model of the determinants of the average problem loan ratio for mortgage loans in Hong Kong. The model postulates that the average problem loan ratio is positively correlated with the proportion of mortgage loans in negative equity in banks’ mortgage portfolios. Other things being equal, a mortgage loan portfolio with a higher average LTV ratio at origination and a higher share of new mortgage loans to total loans tend to have a higher proportion of mortgage loans in negative equity and thus a higher default rate should negative property price shocks occur. LTV policy could, therefore, reduce the mortgage default risk by dampening the average LTV ratio at origination (by the direct effect) and the share of new loans to total loans (by the indirect effect).

Based on the empirical model, we then impose a hypothetical property price shock (assuming a 60% drop in property prices) and the estimated impact on the problem loan ratio in the eight quarters starting from Q1 2013 under the actual and “no policy” scenarios. Figure 10 shows that the problem loan ratio would increase from 0.03% in Q4 2012 to about 0.95% in Q4 2014 in the actual scenario compared with a much higher ratio of 2.32% under the counterfactual “no policy” scenario. The result suggests that the five rounds of the LTV cap tightening from 2009 to 2012 improve the banking sector resilience to property price shocks.

Our core interest, however, is the relative contributions of the direct and indirect effects to the 1.37% reduction of the estimated problem loan ratio from the counterfactual no policy scenario to the actual scenarios. A decomposition analysis using the Shapley approach shows that, of the estimated 1.37% reduction in the problem loan ratio, the direct and indirect effects account for 1.21 and 0.16 percentage points respectively. In other words,

---

21 A problem loan ratio of mortgage loans is measured by the sum of the three-month delinquency ratio and the rescheduled loan ratio for mortgage loans in Hong Kong.
22 The model also includes the unemployment rate and the proportion of mortgage loans with a distressed level of DSR as explanatory variables. Similar empirical models, which hypothesize that negative equity and a high level of debt-servicing burden are two triggers for mortgage defaults, have been developed by Foote et al. (2008) and Elul et al. (2010).
23 Assuming the property price shock is associated with a 300-basis-point increase in the reference interest rate, a 20% decline in household income and the unemployment rate increasing to 8.5%.
24 For details of the Shapley approach, see Shorrocks (1999) and Israeli (2007).
the effect of LTV policy on reducing the sensitivity of mortgage default risk to property price shocks is mainly through its impact on leverage.

6. Conclusion
This paper assesses some important issues regarding the use of LTV policy as a macroprudential tool, including its effectiveness, potential drawbacks and its transmission mechanism to improve financial stability. Hong Kong's experience and the empirical findings of the econometric analysis of the panel data suggest that LTV policy is effective in reducing systemic risk associated with boom-and-bust cycles in property markets. Although the significant number of homebuyers participating in Hong Kong's MIP indicates that LTV policy can lead to liquidity constraints for some households, empirical evidence shows that the MIP can mitigate this drawback without undermining the effectiveness of LTV policy. Thus MIPs play an important role in enhancing the net benefits of LTV policy. More importantly, potential liquidity constraints should not be considered a compelling reason for not adopting an LTV policy to contain the systemic risk associated with property price shocks.

The empirical findings regarding the transmission mechanism of LTV policy shed light on one important policy question – should LTV policy be assigned to target household leverage, credit growth or property market activities? There is strong evidence that tightening LTV caps would reduce household leverage, and that the effect on leverage plays the major role in reducing mortgage default risk. By contrast, there is no clear evidence that tightening LTV caps dampens property market activities. Taken together, the effect of LTV policy would be transmitted mainly through impacts on the household leverage rather than on property market activities. Regarding the policy impact on credit growth, although LTV cap tightening since 2009 is found to dampen credit growth, the state-dependent feature may suggest that calibrating this tool to curb loan growth may pose challenges for policymakers, as the calibrating needs an accurate estimate of loan demand and supply.

References
Kent, R. J., 1980, “Credit rationing and the home mortgage market,” Journal of Money, Credit and Banking 12:3, 488-501
Effectiveness of loan-to-value ratio policy and its transmission mechanism: empirical evidence from Hong Kong

Appendix

<table>
<thead>
<tr>
<th>Variable</th>
<th>Remark</th>
<th>Expected Impact</th>
<th>Estimation result</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demand equation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual change in the market LTV ratio (LTV&lt;sub&gt;m&lt;/sub&gt;)</td>
<td>A lower LTV&lt;sub&gt;m&lt;/sub&gt; implies a higher down payment requirement, which could force some marginal homebuyers out of the property market [Zumpano et al. (1986)], implying a positive relationship between LTV&lt;sub&gt;m&lt;/sub&gt; and the demand for mortgages.</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Returns on equity (RoE) for property investment: 1/(1-LTV&lt;sub&gt;M&lt;/sub&gt;)&lt;sup&gt;2&lt;/sup&gt; times net property return&lt;sup&gt;3&lt;/sup&gt;</td>
<td>From a property investor’s perspective, a lower LTV&lt;sub&gt;M&lt;/sub&gt; constrains investors’ ability to take higher leverage to enhance their RoE, contributing to lower demand for properties. So, RoE would have a positive impact on the demand for mortgages.</td>
<td>+ **</td>
<td>***</td>
</tr>
<tr>
<td>An interactive term of a dummy variable for capturing the effect of the SSD and ROE</td>
<td></td>
<td>- **</td>
<td></td>
</tr>
<tr>
<td>A dummy variable for stricter DSR requirements in 2010</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>A dummy variable for DSR tightening in 2012</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Unemployment rate</td>
<td></td>
<td>- ***</td>
<td></td>
</tr>
<tr>
<td>A dummy variable for Chinese New Year</td>
<td></td>
<td>- ***</td>
<td></td>
</tr>
<tr>
<td><strong>Supply equation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual change in LTV&lt;sub&gt;u&lt;/sub&gt;</td>
<td>An increase in LTV&lt;sub&gt;u&lt;/sub&gt; has a positive impact on loan supply. This specification asserts that the actual price of a mortgage loan is determined not only by the mortgage rate, but also by other contractual terms, such as LTV ratio and maturity [Kent (1980) and Stiglitz and Wesis (1981)]. Banks can, therefore, shift their supply of mortgage loans by adjusting these non-price mortgage terms.</td>
<td>+ **</td>
<td>***</td>
</tr>
<tr>
<td>Risk-adjusted return of mortgage loans on capital</td>
<td></td>
<td>+</td>
<td>***</td>
</tr>
<tr>
<td>Annual growth rate of residential property prices</td>
<td></td>
<td>+</td>
<td>***</td>
</tr>
<tr>
<td>Available funds: annual growth rate of Hong Kong dollar deposits</td>
<td></td>
<td>+</td>
<td>***</td>
</tr>
<tr>
<td>A dummy variable for stricter DSR requirements in 2010</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>A dummy variable for DSR tightening in 2012</td>
<td></td>
<td>- **</td>
<td></td>
</tr>
</tbody>
</table>

Table A1: The specification of the demand and supply regression equations for mortgage loans in Hong Kong

Notes:
1. ***, ** indicate statistically significant at the 1% and 5% levels, respectively.
2. It can be shown that 1/(1-LTV<sub>M</sub>) equals the ratio of the property value to equity (i.e., the amount of down payments) for property investment.
3. Defined as 12-month property price return + property rental yield – (effective borrowing rate for best lending rate-based mortgages times LTV<sub>M</sub>).
4. Defined as (1-profit tax rate) times net mortgage return divided by an estimated amount of regulatory capital required per H.K.$, where net mortgage return is effective mortgage rate minus total cost of mortgages (= the sum of funding cost, administrative cost and expected credit loss).

Pieter Klaassen
Group Risk Control, UBS
Idzard van Eeghen
Chief Risk Officer, Royal Bank of Scotland N.V.

Abstract
We introduce a new performance scheme for banks, inspired by the Du Pont scheme for corporates, which clarifies the relationship between return on equity (RoE), risk-adjusted return on capital (RAROC) and return on assets (RoA). The scheme highlights how common financial ratios risk factors influence the development of RoA, RAROC and RoE. The scheme can be applied by managers, analysts and regulators to analyze the performance of an individual bank, as well as the performance of the banking sector as a whole. In addition, it can be used by bank managers to set coherent targets for various key financial ratios that tend to be managed separately within a bank, to achieve a target RoE, RAROC and RoA. We illustrate our performance scheme by applying it to analyze the main drivers behind the development of the performance of the U.S. commercial banking sector during the past 23 years.

1 The views expressed in this article are those of the authors and not necessarily of their employers.
1. Introduction
How can banks generate healthy returns? This question occupies the minds of many bank managers, regulators and analysts, as banks have to increase their capital at the same time as they aim to improve their returns on capital. The commonly used performance metric to measure bank performance is return on equity (RoE). However, this metric completely ignores risk. Since returns can be boosted by taking more risk, at least in the short run, this is a serious limitation. A performance measure that takes risk into account is the risk-adjusted return on capital (RAROC), but this performance metric has not captured the imagination of many bank managers and analysts yet. This is probably due to the lack of availability and comparability of this metric across the industry, and the lack of clarity on how such a metric relates to returns for shareholders.

More generally, a problem with bank performance metrics is the missing link with traditional financial variables and ratios that drive performance. For industrial firms, the Du Pont scheme establishes this link by decomposing RoE into business factors that drive performance, including operating margin, capital velocity and leverage. Analyzing these factors helps us understand the difference between a firm's RoE and that of its peers, and how performance could potentially be improved. Similar schemes to analyze bank performance [Grier (2007), European Central Bank (ECB) (2010), McGowan and Stambach (2012)] typically disregard important factors that reflect risk-taking, regulatory capital requirements and the cyclical nature of bank performance.

In this article, we introduce a new performance scheme for banks, inspired by the Du Pont scheme for industrial firms, which clarifies the relationship between return on assets (RoA), RoE and RAROC. In contrast to existing performance schemes, our performance scheme clarifies how relevant risk factors influence RoE. The scheme can be used by analysts and regulators to analyze the performance of an individual bank, as well as the performance of the industry. It can also be used by bank management to set consistent targets for various key financial ratios that tend to be managed by different managers within a bank to achieve a target RoE, RAROC and RoA.

In the next section, we introduce the relevant performance metrics and our performance scheme for banks, and use them to analyze bank performance in the U.S. between 1992 and 2014. For the application of our performance scheme to the U.S. banking sector, we take the viewpoint of a commercial bank. We will subsequently comment on how the scheme could be modified to reflect other bank activities, such as investment banking or asset management.

2. RoE, RAROC and risk-based required capital
The key performance metric that most analysts use is RoE. This is defined as profit after tax (PAT) divided by equity. Since equity analysts are interested in shareholder returns, equity is typically taken to be the average common shareholder equity over the reporting period. We use the book value of equity in our performance scheme as this is generally available, although the market value of equity is preferable from a theoretical point of view.

RAROC is defined as:

\[
\text{RAROC} = \frac{\text{Revenues - operating costs - expected loss}}{\text{Risk-based required capital}}
\]

RAROC adjusts a bank's return for risks in two ways. First, the numerator incorporates the expected loss associated with lending, rather than the actual loan impairments that are included in the income statement of the bank. Expected losses are calculated on the basis of long-term average default rates and recovery rates associated with the bank's actual loan portfolio. It thus represents a long-term view on loan losses. RAROC thus represents a view on profitability that is cycle neutral by considering long-term average credit losses instead of actual loan impairments that are influenced by the phase of the economy.

Second, it relates the adjusted income as measured in the numerator to risk-based required capital (RRC). RRC is a measure of how much capital a bank needs as a consequence of the risks it is exposed to, as opposed to the amount of available capital. Risk is high when the potential for large unexpected losses is high, and the bank then needs more capital to safeguard itself against the risk of insolvency. Regulators require that banks hold more capital when their risks increase. With Basel 2, 2.5 and 3, these regulations have become increasingly more detailed and risk sensitive. However, the Basel regulations do not capture all risks, and some calculated risks are simplifications...
of the real risks. Hence, a number of banks have developed internal models to measure their risks more accurately and comprehensively [Klaassen and Van Eeghen (2009)]. These internal capital measures are referred to as “economic capital.” Ideally, these internal capital measures should be used in RAROC calculations. However, they are not always disclosed; and when disclosed, a comparison between banks is not always possible. Analysts, therefore, often have to work with required or available regulatory capital as (proxy) measures for RRC.

RAROC is used to assess whether a bank’s returns provide sufficient compensation for the risks that it is exposed to. RAROC is also used to assess and compare the performance of different business units within a bank that allocates RRC to business units on the basis of their contribution to the bank’s overall RRC.

3. A performance scheme for banks
An important feature of a good performance scheme is that the performance can be imputed to underlying factors that can be managed by the institution, and that are coherently related to each other (i.e., no double counts or gaps).

We distinguish three stages in our performance scheme:

1. Calculation of RoA
2. Derivation of RAROC from RoA
3. Derivation of RoE from RAROC

The graphical form of the scheme is presented in Figure 1.

The three key performance metrics are in formula form:

\[
\text{RoA} = \frac{R - C}{EA} = \frac{R}{EA} \left(1 - \frac{C}{R}\right)
\]

\[
\text{RAROC} = \frac{R \cdot C \cdot EL}{RRC} = \left(\frac{\text{RoA} \cdot EL}{EA} \cdot \frac{EA}{TA} \cdot \frac{RA}{RRC}\right)
\]

\[
\text{RoE} = \left(\frac{\text{RAROC} \cdot \text{EL}}{RRC} \cdot \frac{RRC}{\text{Equity}} \cdot \frac{PAT}{PBT}\right)
\]

RoA is primarily determined by the bank’s choice of business activities and cost-efficiency, since the factors that determine the RoA are:

- Gross return on assets (R/EA): earning assets are the assets that contribute directly to revenues. Non-earning assets include intangible assets, goodwill, tax claims and other non-interest bearing claims. The ratio reflects the bank’s asset productivity and is higher for activities with little price competition, or that are more risky (high revenues being compensation for the risk). The ratio will also be higher if the bank generates significant revenues from activities that do not contribute (materially) to earning assets, such as transaction services and asset management.

- Cost/Income (C/R): a lower ratio implies that fewer costs need to be made to generate a dollar of revenues. The ratio reflects the bank’s cost-efficiency.

RAROC represents the return on risk-based required capital, with the return based on expected credit losses, i.e., ignoring unexpected credit losses that are of a cyclical nature. The RoA is an important driver of the RAROC. Other factors are:

- Expected loss as a percentage of earning assets (EL/EA): reflects the asset quality of the bank. A high ratio indicates that the bank is active in lending to debtors that have a high risk of default (e.g., credit cards and subprime lending). However, we note that this ratio only provides an indication of the credit
quality of assets that are accounted for as held-to-maturity. For assets that are accounted for at fair value, such as assets in the trading book, the expected loss is reflected in the fair value already. Hence, an increase of assets that are accounted for at fair value may lead to a decrease in the EL/EA ratio without a change in asset quality.

- **Earning assets as a percentage of total assets (EA/TA):** is an indicator of a bank’s efficient use of assets. If a bank has many non-earning assets the ratio would fall and form a drag on its RAROC.

- **Risk-normed leverage (TA/RRC):** indicates to what extent the assets of a firm should be funded with capital, given the risk profile of the assets. In other words, how much leverage a firm can use. This ratio is lower the more risky the assets are, since risky loans require more capital. This leverage ratio is a normative ratio, because it is calculated on the basis of how much capital is required, not how much capital a bank actually has.

RAROC increases when (ceteris paribus) RoA increases, when asset quality improves or when risk-normed leverage increases. Optimizing RAROC may give rise to some complex trade-offs. For example, if a bank wants to increase its RAROC by expanding its high-risk loan portfolio and thereby its RoA, then such an expansion will increase its EL and decrease its risk-normed leverage, because required capital will increase. The key question will be whether the increase in RoA sufficiently compensates for the higher EL and lower acceptable leverage. A number of sophisticated banks have provided their business and risk management units with RAROC-tools to aid them in finding the right balance between risk and return.

The final step is to derive RoE from RAROC. This is accomplished by adding three more factors to the equation:

- **Unexpected losses (UL) as a percentage of RRC (impairments – EL)/RRC:** if impairments are higher than expected losses, this constitutes an unexpected loss, which reduces RoE. Impairments are typically higher than EL in an economic downturn, and lower than EL in good times. As such, this ratio provides an indication of how economic circumstances have affected the bank’s results and RoE.

- **The inverse capital buffer ratio (RRC/equity):** holding equity in excess of RRC may reflect the desire of the bank’s management to have a capital buffer on top of RRC. A high capital buffer will decrease leverage and reduce RoE.

- **The impact of taxes (PAT/PBT):** higher taxes result in a lower ratio and a lower RoE.

It follows that RoE and RAROC are directly and positively related to each other. RoE increases when RAROC increases, but also when the capital buffer decreases and when there are favorable economic circumstances (i.e., when actual impairments are lower than expected losses).

When we consider the entire performance scheme, we can directly see the chain of cause and effect. For example, a lower cost/income ratio will improve RoA and, as a result, improve RAROC and RoE. The risk-normed leverage ratio has an impact on RAROC and, via RAROC, on RoE. The inverse capital buffer ratio only has an impact on RoE.

The performance scheme comprises ratios that are commonly tracked within a bank and relate to specific management actions or choices. In that respect, it is very similar to the Du Pont scheme for industrial firms.

**4. Application of the performance scheme: analyzing performance of the U.S. commercial banking sector**

Aggregated financial data for all commercial banks in the U.S. can be obtained from the website of the Federal Deposit Insurance Company (FDIC; www.fdic.gov). The number of commercial banks included in the FDIC database has decreased gradually between 1992 and 2014, from 11,463 to 5,642, while total assets have more than quadrupled from U.S.$3.5t to U.S.$14.5t. This consolidation trend had started much earlier [see Berger et al. (1995) for a study of the period 1979–94 and Jones and Critchfield (2005) for the period 1984–03].

Based on the available data, we have made the following assumptions in our analysis:

- Profit after tax is measured as “income before extraordinary items,” thereby excluding “extraordinary gains.”
- Equity equals “total equity capital” and includes (a small amount of) perpetual preferred stock.
- RRC is not available in the data, and, therefore, we need to approximate it. We have set RRC equal to intangible assets
plus 8% of risk-weighted assets. Between 1992 and 2014, U.S. banks calculated risk-weighted assets based on Basel I, which has limited risk sensitivity as it only includes credit and market risk and, within credit risk only distinguishes between broad lending categories (sovereigns, corporate, retail mortgages, etc). Hence, the proxy we use for RRC may differ materially from the risk-based capital requirement that banks calculate internally.

- EL to EA in each year equal to the average ratio of impairments to risk-weighted assets over the period 1992–2014 times the ratio of risk-weighted assets to earning assets at the beginning of the year. The ratio thus reflects the long-term average of actual impairments, but varies with the risk weight (and thus riskiness) of assets.

Figure 2 shows the expected loss ratio, and thus the risk profile of the credit portfolio, increased between 1994 and 2000, notwithstanding that realized impairments were below their long-term average. After a peak in impairments in 2001–02, associated with the dotcom crisis, the expected loss ratio slightly decreased before increasing again in 2005–06. After the very significant impairments as a result of the financial crisis, the expected loss ratio declined notably as banks took firm action to clean up their balance sheets.

As Figure 3 shows, RoE for the U.S. banks in aggregate has decreased from about 15% between 1993 and 2006 to below 10% since the financial crisis. In the same period, the yield on 10-year U.S. Treasury securities decreased by approximately 4%, and while the RoE of banks bounced back after the financial crisis, the excess return over U.S. treasuries has not returned fully to pre-crisis levels.

To analyze the causes for the decrease of RoE, we first compare the development of RoE with the development of RAROC during the same period (in Figure 4).

We observe that, up to 2002, RAROC and RoE display a similar pattern, but in the years thereafter, RAROC decreases whereas RoE remains flat. The main reason for this is that the capital buffer decreased, and hence the ratio RRC/equity increased, between 2002 and 2008, which benefited RoE. We observe in the run-up to both the dotcom crisis in 2001–02 and the financial crisis in 2008, that capital buffers are decreasing; after the dotcom crisis, the capital buffer was only partially restored, but after the financial crisis, the capital buffers increased significantly. While lower capital buffers helped to support RoE, the re-establishment of a capital buffer had the opposite effect.
An additional contributor to the difference between RAROC and RoE is the unexpected loss, which impacts RoE but not RAROC. This is most clearly seen in the large unexpected losses in 2008 to 2010, as a result of which, RoE recovered more slowly than RAROC. We have depicted the impact of the unexpected loss (IMP-EL)/RRC with a negative sign in Figure 4 to clearly show the positive correlation with RoE.

To obtain additional insight into the drivers of the changes in RAROC over time (and implicitly also RoE), Figure 5 depicts the development of the main individual components influencing RAROC = RoA and the risk-normed leverage ratio TA/RRC.

Figure 5 shows that the risk-normed leverage gradually decreased until 2007, implying that banks were taking more risks. Until 2002, the RoA increased, suggesting that the higher risks were also reflected in higher returns. However, between 2002 and 2007, RoA and risk-normed leverage both decreased, resulting in a significant drop of the RAROC. It was only by reducing their capital buffer that banks could maintain their RoE, as was shown in Figure 4.

As Figure 6 shows, the RoA increased until 2002, mainly due to higher cost-efficiency (1 - C/R). In the period thereafter, lower revenues (R/EA) started to reduce RoA, but RoA really fell sharply during the financial crisis, as a result of large trading losses that have been accounted for in revenues, and thus depressed both the R/EA and (1 - C/R) ratios. After the financial crisis, when trading losses had ceased, RoA bounced back but, from 2010 onward, RoA fell again as the pressure on revenues caused a continuation of its downward trend. Moreover, the cost-efficiency did not return to pre-crisis levels, notwithstanding various cost reductions by banks; litigation costs related to the financial crisis proved to be very significant [Chauduri (2013), Lloyd (2013)]. These litigation costs materialize with a considerable time lag and it may, therefore, take some time before the cost-efficiency improves again.

The analysis for the financial performance of the aggregate U.S. commercial banking sector thus shows the following:

- The increase in both the risk-normed leverage and the capital buffer since the financial crisis and, to a lesser extent, after the dotcom crisis, indicates that U.S. banks have reduced their risk profile in response to these crises. These changes are much less visible if we were to consider only the total leverage ratio (total assets/equity), since the increase in the risk-normed

---

**Figure 4:** RoE and RAROC for aggregated U.S. commercial banks over the period 1992-2014

**Figure 5:** RAROC and RoA for aggregated U.S. commercial banks over the period 1992-2014
leverage (TA/RRC) is offset by the increase in the capital buffer (decrease in (RRC/equity)) in this ratio.

RoA has been under pressure since the financial crisis due to lower returns on earning assets, thereby negatively impacting RAROC and RoE. The lower return on earning assets appears to be a global phenomenon, and is also observed in various European banking markets where the decrease, in fact, started earlier and is much more pronounced (see for example Lusignani and Onado (2014), Van Eeghen and Klaassen (2013)).

The increase in RoE since the financial crisis is mainly caused by the fact that credit losses have been below long-term average losses, but the underlying trend, as reflected in a lower RAROC, signals fundamental performance challenges. The conclusion is, therefore, sobering: if the decreasing trend of R/EA cannot be stopped, and increasing leverage is ruled out by regulators, then improving cost-efficiency is the only route open to banks to improve returns. Nevertheless, the challenges for U.S. banks pale in comparison with those of European banks, whose profitability has suffered by a much greater degree after the financial crisis [Deutsche Bank Research (2013)].

---

5. Using the performance scheme for scenario analysis and target setting

The presented performance scheme enables a structured analysis of the main drivers behind the RoE, RAROC and RoA performance metrics and their interrelationships. In addition, it allows for a direct evaluation of the impact of a change in a financial ratio in one of the branches in the performance scheme on RoA, RAROC and/or RoE. With a target RoE in mind, it can, therefore, be used to determine target levels for the individual financial ratios that contribute to the RoE.

Figure 7 provides the performance scheme for the aggregate U.S. commercial banks for 2014. For banks RAROC and RoE in 2014 were, respectively, 13.82% and 9.20%. On the one hand, RoE had the benefit of a tail wind in 2014, as loan impairments were significantly below their estimated long-term expected value ((Imp – EL)/RRC = –4.33%). On the other hand, the relatively large capital buffer (represented by the ratio RRC/equity being significantly below 1) depressed the RoE.

We will now analyze various scenarios. As a first illustration, we evaluate the impact of higher RRC (triggered, for example, by new insights and refinement of internal risk models or, if based on RWA, an increase in RWA), while taking a long-term view on...
credit losses. In this scenario, we make the following adjustments to the scheme: unexpected credit losses are set to zero and we increase the percentage with which we multiply RWA from 8% to 10% to obtain a higher risk-based required capital. The resulting performance scheme is presented in Figure 8 (changes from the actual situation are highlighted in red).

The performance scheme shows that the higher capital requirements result in a lower risk-normed leverage, which reduces RAROC from 13.82% to 11.77%. The mirror image of higher capital requirements is a lower capital buffer (RRC/equity) that would normally compensate for the impact on RoE. However, in this example, we assume that banks and/or their regulators want to maintain the equity buffer, so we have kept that constant at 0.73. With unexpected losses set to zero, the end result is that RoE reduces to 5.97%.

In case management aspires to achieve a RoE of, say, 10%, what action can be taken to achieve this? Various options exist, ranging from seeking assets with higher returns, reducing costs, decreasing non-earning assets and lowering taxes. In Figure 9, we “solve” the problem by reducing costs and improving the operating margin from 38.18% to 53.72%. This margin is approximately 10% higher than the best margin achieved by U.S. commercial banks since 1992, a veritable challenge.

In reality, individual banks may need to critically evaluate their chosen markets, business activities and operating model to be able to achieve a significant improvement of RoE. This will likely affect most, if not all, ratios in our scheme, instead of just the cost-efficiency that we have used in the illustration. The necessary changes in ratios that are found as a result of such an evaluation can subsequently be used to set performance targets for separate areas in the bank.

6. General use of the performance scheme
The above analysis can be performed at an industry level and at the level of an individual organization, or even business unit. For a bank performance scheme to be of practical value, its key driving factors must be intuitive, actionable and risk sensitive. In the presented scheme, the factors are well known within the industry and responsibilities can be assigned fairly easily. For example, managing the capital buffer is typically the responsibility of the treasury, whereas management of expected (and unexpected) credit losses is the responsibility of portfolio management and/or credit risk management. In many cases, an optimal balance between two or more factors needs to be found; for example, between revenues (R/EA) and risk-normed leverage (TA/RRC),
since changing one is likely to lead to an offsetting change in the other(s). Pricing and other tools may be provided to the commercial units to aid them find the optimal balance.

Some factors may be branched out further to deepen the analysis. For example, the ratio R/EA may be split to distinguish between interest and non-interest income. Moreover, for a bank with multiple activities, a separate branch can be introduced in the performance scheme for each activity at the RAROC node. The RoE will then become a weighted average of the RAROC of the various activities, multiplied by the group-wide tax efficiency and capital buffer ratios.

In our presentation, we have taken the viewpoint of commercial banks. However, the performance scheme can easily be generalized to other types of banks and activities. For example, for asset management, earning assets could be replaced by assets under management. For trading activities, they could be replaced by a measure such as VaR. As such, we believe the performance scheme to be quite versatile to accommodate various situations and organizations.

**Conclusion**

The presented performance scheme allows bank analysts to analyze the driving factors of a bank’s performance and the relation with (changes in) the risk profile of the bank. The scheme can be applied to individual banks, as well as at the industry level, as we have done in this paper.

The scheme is also relevant for regulators to obtain better insight of the drivers of change in the financial health of the banking industry. Furthermore, new regulations may negatively affect various performance drivers, of which the impact may be better understood by use of the scheme.

Finally, and perhaps most importantly, the presented performance scheme may help a bank’s leadership to formulate a coherent internal risk appetite and set explicit targets for key performance drivers that directly link to the desired shareholder returns. The clarity of the link between performance targets and RoE will help staff to have a clear understanding of what is being expected from them and why. Such clarity will also contribute to a convincing story for investors. The fact that risk appetite, in the form of asset quality and risk-normed leverage, is an explicit driver of RoE, should furthermore give comfort to bank regulators and depositors. The performance scheme thus provides clear cues to all stakeholders on how banks can generate healthy returns again.

**References**


Chauduri, S., 2013, “U.S. banks’ legal tabs is poised to rise, new estimates for crisis-era settlements total $107 billion or more,” Wall Street Journal, 28 October


Jones, K. D., and T. Critchfield, 2005, “Consolidation in the U.S. banking industry: is the ‘long, strange trip’ about to end?”, FDIC Banking Review 17:4


Recursive collective action problems: the structure of procyclicality in financial and monetary markets, macroeconomies and formally similar contexts

Robert Hockett
Edward Cornell Professor of Law, Cornell Law School

Abstract
The hallmark of a collective action problem is its aggregating multiple individually rational decisions into a collectively irrational outcome. Arms races, “commons tragedies” and “prisoners’ dilemmas” are well-known, indeed well-worn examples. What seem to be less widely appreciated are two complementary propositions: first, that some collective action problems bear iterative, self-exacerbating structures that render them particularly destructive, and second, that some of the most formidable challenges faced by economies, societies and polities are iteratively self-worsening problems of precisely this sort. Financial markets, monetary systems and macroeconomies in particular are rife with them – as are other complex systems subject to group-mediated procyclicalities or “feedback” effects.

I call the mentioned challenges “recursive collective action problems,” and show that a great many familiar regulatory and policy challenges – including asset price bubbles and busts, consumer price hyperinflations and debt deflations, “paradoxes of thrift” and “recessionary spirals” – constitute instances of this general phenomenon. I also hazard suggestions as to how best to address such challenges. Key to the effort is first to recognize their shared structure, second to recognize that collective action problems require coherent collective agency for their solution, and third to recognize that the collective agents in question must act to render no longer individually rational such decisions as aggregate into collectively irrational outcomes. I close with specific examples of what problem-solving strategies informed by the “three recognitions” will tend to look like. The implications for macroeconomic and “macroprudential” finance-regulatory policy in particular are manifold. If we but attend to the shared nature, structure and pervasiveness of recursive collective action problems, I conclude, we can recoup much in the way of wealth and well-being that is now needlessly lost.
1. Introduction
Since Hobbes’s day at the latest, and probably for much longer still, people have been aware of certain challenges that confront groups of rational agents, precisely by virtue of their being (a) groups of (b) rational agents. These challenges typically stem from either (a) the very fact or (b) certain possible structures of plural and consequently distributed rational agency. Challenges of the first sort are nowadays often called “coordination problems.” They are rooted in the need of certain kinds of uniformity on the one hand, and the absence of any inherent tendency for such uniformity spontaneously to emerge on the other hand. Challenges of the second sort are now typically called “collective action problems.” They stem from certain possible divergences between what it is individually rational to do, absent coordination, on the one hand, and what would be possible divergences between what it is individually rational to do, absent coordination, on the one hand, and what would be both collectively and, therefore, individually optimal to do, were reliable means of coordination available, on the other hand.1

The dominance of these terms – “coordination problems” and “collective action problems” – as names for what they now designate is in some ways regrettable. One reason is that, as is implicit in the foregoing, so-called collective action problems constitute a subclass of coordination problems, since collective action itself is a form of coordination; yet coordination and collective action are often discussed under their separate rubrics as if they bore no structural features in common. Another, related reason to regret the received terminology is that, on a “natural,” untutored understanding of the words, a coordination problem is a collective action problem as surely as a collective action itself is a form of coordination; yet coordination and collective action are often discussed under their separate rubrics as if they bore no structural features in common. Another, related reason to regret the received terminology is that, on a “natural,” untutored understanding of the words, a coordination problem is a collective action problem as surely as a collective action itself is a form of coordination; yet coordination and collective action are often discussed under their separate rubrics as if they bore no structural features in common.

Notwithstanding these reservations, however, for purposes of this discussion, we shall play with the terminological hand we have been dealt, and our focus will be on a particularly vexing subclass of collective action problems whose solutions require a particularly robust form of coordination. These problems are particularly rife in credit, monetary and broader financial markets, not to mention the macroeconomies that depend on them, meaning in turn that the mentioned robust form of coordination must find expression in the regimes pursuant to which these markets are supervised.

Now, as suggested earlier, the hallmark of a collective action problem is its tending to aggregate multiple individually rational decisions, absent coordination, into a collectively irrational outcome – an outcome that is ultimately suboptimal for each individual agent. In other words, it involves some situation in which individually rational action, when taken by multiple individually rational actors, ultimately proves self-defeating for each actor precisely because all actors (rationally) take it. The trajectory from individually rational decision to individual self-defeat is in this sense mediated, in these cases, by multiplicity itself; in that everyone doing the rational thing is precisely what brings on the suboptimal outcome. And so an individual or collective decision to endorse or permit every individual to act rationally in the relevant manner would itself be an individually or collectively irrational – because ultimately self-defeating – decision.2

There is in this sense a certain incoherence, a dilemmic or “tragic” character, inherent in any situation characterized by the presence of a collective action problem. “Commons tragedies” and “prisoners’ dilemmas” are probably the most oft-encountered examples of the paradoxical, “tragically” ironic phenomenon in question – examples the very names of which hint at the crossed-purposes problem that these problems present.3

Absent coordinated rationing of a scarce resource such as shared grazing land – “the commons” – is once said to have been, for example, “rational,” in a particular sense that we shall more fully and critically elaborate later, for each individual shepherd to permit his cattle to graze without restriction. For if one is competing for, rather than jointly managing, a

---


2 Here is one sense in which official endorsement of “unregulated markets” by, e.g., certain political parties, is, if sincere, collectively irrational. Likewise in the case of any individual’s endorsement of the same.

3 The “prisoner’s dilemma” has been a staple of game-theoretic discussion and its associated literature for so long as to have long since become tiresome. It is, of course, treated, among many other places, in von Neumann and Morgenstern (1944), supra note 1. The “commons tragedy” is nearly as well-worn, particularly since its canonization in ever-cited Hardin, G., 1968, “The tragedy of the commons,” Science 162(3859), 1243-1248.
scarce resource, and, crucially, cannot unilaterally change this background condition, then more grazing for one’s cattle is nearly always “better,” while less for one’s own is simply more for one’s competitors. Yet—and here, of course, is the tragedy—everyone acting thus rationally and grazing their cattle without restriction then depletes the commons more rapidly than it can be replenished, ultimately impoverishing everyone.

Similarly, absent coordinated testimony on the part of the “prisoners” in the “dilemma” that is named for them; it is in a certain sense rational for each prisoner to confess and thus “rat out” the other. Such is the “dominant strategy” for each prisoner in their capacity as an individually rational agent. Yet both prisoners’ adoptions of their dominant strategy then seals the doom of each one of them. Hence, once again there is tragedy.

This much, I take, is more or less widely appreciated, certainly among choice theorists operating in various disciplines, and even, to a lesser extent, among laypersons. What seem to be less widely appreciated, however, are two complementary propositions. The first is that some collective action problems bear iterative, self-exacerbating properties. The second is that some of the most formidable and all too common challenges faced by economies, societies and polities are iteratively self-worsening collective action problems of precisely this sort. Indeed, most, if not all, of the procyclicality that often plague financial and monetary systems and macroeconomies, denying them tolerable equilibria, seem to involve them.

As for the iterative, self-exacerbating character of some collective action problems, this appears to stem from the fact that some actions taken by some actors at one moment can rationally warrant specific responsive actions, taken by other actors within the same domain of action at subsequent moments, which can then “feed back” procyclically into yet further rationally responsive decisions taken within the same domain by the original actors. This cycle can continue with no natural resting point—no equilibrium—and result in situations that are suboptimal at best and disastrous at worst. The iterativity in these cases is, in this sense, as rooted in plural agency as is the aggregative nature of the decisions in question, while also being rooted in that same other-regard and mutuality—the mutual responsiveness—which characterizes so much of human behavior.

As for the commonality of collective action challenges of this iterative, self-exacerbating type, we shall see this in the particular examples that I shall consider. One thing we will see in particular is that financial markets and macroeconomies are rife with them.

I call these iterative cases of the classic collective action problem “recursive collective action problems” (recap). They are collective action problems by virtue of their aggregating multiple individually rational decisions into collectively irrational outcomes. They are recursive collective action problems by virtue of their mutuality-rooted iterativity and iterativity-rooted self-exacerbation properties, with each new individually rational but collectively self-defeating action in the sequence effectively inducing further such individually rational actions along the same ultimately self-defeating lines. My aim in this paper is both to limn the shared structure of these problems in a bit more detail, and to demonstrate how a great many familiar regulatory and policy challenges—including those presented by asset price bubbles and busts, consumer price inflations and debt deflations, bank runs and financial panics, “paradoxes of thrift” and “liquidity traps”—all constitute instances of the phenomenon under consideration.

I also hope in the paper to hazard some helpful suggestions as to how best to address recap challenges. Key to the effort in this case is, first, to recognize these challenges’ shared structure; second, to recognize that collective action problems require coherent collective agency for their solution; and third, to recognize that the collective agents in question must, in general, act precisely to render no longer individually rational such individually rational decisions as tend to aggregate into collectively irrational outcomes.

---

4 The iterativity of some choice scenarios that confront choosers with collective action problems appears to be more widely appreciated than the self-exacerbation properties that attend many of the same. This might owe something to the moral of the story told by Axelrod and Hamilton (1981) [Axelrod, R., and W. Hamilton, 1981, “The evolution of cooperation,” Science 211:4489, 1390–1396], an early work placing iteration center stage. As the title of this paper, shared by Axelrod’s subsequent (1984) monograph of the same name, suggests, the focus here is on how iterated “games” afford opportunity for cooperative practices to emerge and develop, rather than how much worse things will steadily grow in many cases if it does not. Much the same can be said of Lewis’s [supra, note 1] treatment of convention as the spontaneously emergent solution to a recurrent coordination problem. The present paper can in this sense be viewed as a complement to Lewis, Axelrod and Hamilton, and their progeny, focusing more on the urgency of the need for “cooperation” or “coordination” (especially of a particular kind discussed in section 4) than on the opportunity for the same afforded by “repeat playing.”
The paper is accordingly structured as follows. Section 2 first characterizes the structure of recaps in a bit more detail, focusing in particular on the forms of rationality and mutuality that they involve. Section 3 then lays out a number of specific instances of the recap apt to be familiar to many, even if their shared structure as recaps is not. My hope is that this will both (a) render recaps, as abstractly characterized in section 2, more concretely appreciable, and (b) aid policymakers in framing solutions to the particular problem examples selected, all of which are particularly salient at present – a time marked by post-bubble, post-bust debt-deflation. Picking up on that last aim, section 4 turns from the common form that all recaps share to the common form that solutions to recaps must accordingly take. Once again, the hope is both to illuminate something important about particular forms of collective action and to afford practical guidance to would-be problem-solving policymakers. Section 5 then concludes and looks forward.

2. Structure of the problem
It will be helpful to commence our discussion by characterizing, now with a bit more care, the structure and basic attributes that seem to be borne by all problems of the kind under consideration. Then, in section 3, we will consider a number of specific instances of this common problem type.

A recursive collective action problem, we suggested earlier, is a situation in which (1) multiple decisions that are individually rational in the absence of coordination (2) aggregate into collectively irrational outcomes, the outcomes of which then (3) render it rational for agents to take yet more decisions along the same lines as in (1), thereby compounding the irrationality at work in (2), ad infinitum. Conditions (1) and (2) are what render these problems “collective action” problems. Condition (3) is what renders them “recursive,” possessed of the familiar “feedback-fed” self-worsening property that deprives some interactions among multiple agents of tolerable equilibria. We can say a bit more about each of these conditions, all three of which are individually necessary and jointly sufficient to constitute a recursive collective action problem.

2.1 Individually rational
The form of individual rationality involved in a recursive collective action problem is not particularly exotic. It is simply the form familiar to economists and choice theorists more broadly, pursuant to which chosen means are reasonably believed by the choosing agent to conduce to desired ends. Typically some form of “maximization” is involved in the choosing, such that the agent’s choices are consistent with maximizing the degree to which some antecedent preference or set of commensurable preferences is satisfied. The shepherds familiar to the commons tragedy, for example, aim to maximize the availability of nutrients to their animals. The prisoners perplexed by the prisoners’ dilemma, for their part, aim to minimize terms of imprisonment, hence to maximize time free of prison.

There are, of course, all manner of objections to “instrumental,” or “means” rationality of this familiar variety. There are claims that it has no positive explanatory value, for example, in view of the ease with which putative ends can be recharacterized post hoc as to “rationalize” means and the decisions that have led to them. So are there Aristotelian-style claims to the effect that ends, too, can be (now normatively) rational or irrational, meaning that instrumentalist accounts of rationality à la Hume are too limiting in view of the concept of rationality’s ineluctably normative valence.

None of these concerns, critically important as they are in other contexts, need trouble us here once they are acknowledged. For our purpose in the present discussion is simply to show that, in some circumstances, in which the applicability of this form of rationality appears to be uncontroversial, everyone doing the

---

6 The consensus choice for locus classicus of this take of rationality appears to be Hume, D., 1739, A treatise on human nature. Since Weber (1903) (Weber, M., 1903, The Protestant Work Ethic and the spirit of capitalism), it has been common to refer to this form of rationality as “instrumental.”
7 See, e.g., Weber (1903), idem.
8 This and related concerns are nicely laid out and discussed, for example, in Jose Luis Bermudez (2009) (Bermudez, J-L., 2009, Decision theory and rationality), as well as in some other of the many works out there on the subject. Among the more searching and pioneering of contemporary investigations of decision theory and rationality are the papers collected in Davidson, D., 1980, Essays on actions and events.
individually rational thing in isolation can issue in everyone doing the collectively irrational thing in aggregate, in a manner that not only prevents maximization of what each agent individually prefers, but actually can maximize what each individual disprefers.

2.2 Collectively irrational
The collective irrationality involved in a collective action problem is understood by essential reference to the individual rationality whose end it frustrates. That is to say, it involves subversion of precisely that end which the agents in question arerationally seeking in their disaggregated, individual capacities. This is precisely why the individually rational decisions that aggregate into an irrational outcome are often said to be collectively irrational, or collectively self-defeating, as well as why the resultant choice situation is said to be “tragic,” ironic or paradoxical. The understanding upon which those characterizations are predicated is that the individual decisions are (rationally) aimed at something – some end – and that the particular circumstance in which those decisions are being taken is then structured in such a way as renders the end more elusive precisely when all individuals (rationally) pursue it.

In light of collective irrationalities being thus indexed, in the context of a collective action problem, to individual rationality as just characterized in 2.1, it is, of course, subject to the same sorts of objections we noted before to apply to individual rationality as there defined. Also as there, though, so here, none of these important objections ought to prevent our plowing ahead for present purposes. For our purpose, again, is simply to show that, in some circumstances, in which the applicability of this form of rationality is not particularly controversial or contested, everyone doing the individually rational thing can result in everyone doing the collectively irrational thing, such as then not only prevents maximization of what is preferred, but again actually maximizes that which is dispreferred.

2.3 Recursively self-exacerbating
What renders a recursive collective action problem “recursive” is its self-amplification characteristics, these in turn being rooted in part in the “maximization” behavior associated with rationality as characterized in 2.1 and 2.2. The recursion at work here, in other words, stems from the form of rationality described in 2.1, just as the irrationality described in 2.2 trades on it. Indeed, the element of recursion stems from the crosswise interaction of individual rationality and collective irrationality themselves – the mutuality of some human decision-making – when decisions that aggregate can be iterated.

The process, in essence, runs thus: individuals first act rationally to maximize some desired variable, call it x. In so doing, they employ means that, absent the presence of other actors, are well adapted to that task of maximization. Multiple individuals acting thus individually rationally, however, not only frustrates the maximizing purpose, but actually works to minimize that purpose’s satisfaction. Individuals have less of what they seek – less x – after all of them seek more, than if none of them had sought to maximize – rather than, say, “satisfice” – in the first place.\(^\text{10}\)

When this happens, however, individuals who do not adjust their preferences simply repeat their maximizing actions, since their ends are still unfulfilled and, indeed, less fulfilled than they would have been had not everyone acted individually rationally. Hence the process repeats itself, ad infinitum or at least ad suboptimum, such that each round of individually rational decision-making, paradoxically, carries each individually rational agent further from maximal x. No tolerable equilibrium, in other words, characterizes such sets of interactive decisions. They are, in this sense, procyclical – a fact that bears critically upon their salience for macroeconomic and finance-regulatory purposes.

3. Some common instances of the problem
We noted, in introducing this discussion, that recursive collective action problems are pervasive, and that they are particularly common in the guise of many procyclicalities that vex financial markets and macroeconomies. It will be helpful now to run through some well-chosen examples. There are several reasons to do so. One is to render the phenomenon under discussion – recaps and their characteristic structure – more concrete and accessible to intuition. Another is to illuminate the nature of these

\(^{10}\) In referencing the concept of “satisficing,” I am, of course, alluding to the work of Herbert Simon, particularly as encountered in Simon, H., 1947, Administrative behavior and Simon, H., 1956, “Rational choice and the structure of the environment,” 63 Psychological Review 63:2, 129-138. The idea is that cognitive limitations in many circumstances render optimization behavior taking the form of boundless maximization quite futile, such that decision-making often does better to proceed on the basis of some threshold of acceptability rather than some maximandum. Careful philosophic treatment of Simon’s idea can be found in Slote, M., 1989, Beyond optimizing: a study of rational choice.
Recursive collective action problems: the structure of procyclicality in financial and monetary markets, macroeconomies and formally similar contexts

familiar instances of our general problem by drawing attention to their shared structure as recaps — certain aspects of which structure find implicit expression in elements of these problems’ popular names that I will scare-quote on first mention. Finally, and relatedly, a third reason for offering the particular examples treated here is to provide an indication of how we might best solve these pervasive, destructive and presently salient problems by a common method that responds to their common structure, the subject of section 4.

3.1 Arms “races”
Let us begin with a very simple example, relative to whose basic core all subsequent examples seem to constitute only slightly more complex variations. So first consider two parties, each of whom aims to maintain parity with — or perhaps, under conditions of uncertainty, to maintain a certain incremental “margin of safety” relative to — the other by stockpiling an equal, or marginally greater, number of weapons than the other. The ultimate aim in each party’s case is to maintain “security.”

Given this end, it is instrumentally rational, absent coordination, for party 1 to define its aim relative to party 2’s stockpile — say, by aiming for rough parity with party 2’s stockpile, or for parity with party 2’s stockpile plus some small increment, “d,” to compensate for uncertainty as to just how large party 2’s stockpile actually is. The problem here, of course, is that, as in all collective action problems, what is sauce for the goose is likewise sauce for the gander. It is just as rational for party 2 to define its aim relative to party 1’s stockpile — say, by aiming for parity with party 1’s stockpile plus increment d — as it is for party 1 to define its aim relative to party 2’s stockpile. The parties’ aims are in this sense mutual — they are mutually regarding — and that is indeed part of the problem; the sense in which the problem is group-mediated.

And so, if you’ll pardon a pun, we are off to the races. Starting from a position of rough but uncertain parity, party 1 adds d to its stockpile “just to make sure.” Party 2 responds by adding 2d to its stockpile — d to match, another d “just to make sure.” Party 1 then responds to the response by adding 2d — again d to match, plus another d’s worth to achieve a buffer. And so on. The race is self-reinforcing and self-exacerbating. It is recursively procyclical.

Arms accordingly accumulate steadily in number and expenses grow with them, ad infinitum, yet the parties are no better off, security-wise, late in the game than they are early in the game. Indeed, they are likely much worse even on the security front alone, irrespective of pecuniary expense, given both (1) the dangers of stockpiled weapons and (2) the temptations to launch preemptive wars during brief intervals of relative advantage, occasioned by all arms races. Hence, the race is enormously wasteful — a sheer deadweight loss for both parties in pecuniary and in security terms. That is to say, that it subverts the very ends in terms of which the parties’ means rationality is to be understood.

Of course, the example is stylized. Real arms races are more complex and nuanced than this. But the essential idea applicable to all variations on the stylized example is nevertheless clear: the aim of each party is greater security attained cost-effectively; and yet the race, by steadily increasing the quantum of arms and encouraging preemptive aggression during brief intervals when advantages are had, renders all less secure, while also occasioning enormous pecuniary expense. And things only grow worse in these very respects with each iteration.

Critically, moreover — a point that will figure quite prominently later — neither party can unilaterally quit, let alone stop, the race without thereby relinquishing its aim of security. That is why unilateral disarmament is nearly universally eschewed. It is as individually irrational to stop racing, in other words, as it is individually rational and collectively irrational to continue racing. You’re damned if you do, collectively and hence ultimately individually, and damned if you don’t, individually. This is precisely the sense in which arms races, like all collective action problems, are “tragic” (in the Attic Greek rather than the modernly trivialized, synonymous-with “disaster” sense of the word).

---

11 See generally Schelling, T., 1960, The strategy of conflict; Halperin, M., and T. Schelling, 1961, Strategy and arms control; and Schelling, T., 1966, Arms and influence. Much more, of course, has been written on the structure and dynamics of various species of arms race, much of it for obvious reasons during the period extending from the late 1940s through the late 1980s. Indeed much of the early work in “game theory” itself, a good bit of it sponsored by the Rand Corporation, appears to have been occasioned by the arms race dynamic of the “Cold War.” And earlier contributions, dating back to the 1930s, themselves of course came along during times of significant international tension.

12 For some such nuance, see Schelling (1960) and Halperin and Schelling (1961), idem.
3.2 “Positional goods” competitions

Now suppose we are not talking about nations and weapons, but neighbors and homes. If a number of neighbors inhabit a society in which all wish to attain and hold status relative to others, and status is signaled by home size, each inhabitant of the society in question might wish to own a home slightly larger – and certainly no smaller – than those of their neighbors. Beginning from an initial position of rough but perhaps uncertain parity, then, one neighbor might contract to add a floor or some other addition to their house. Other neighbors might then respond in kind. Still other neighbors might then follow suit. Thus, foiled in their initial attempt to either get a leg up or to ensure that they are not a leg down, the first neighbor might now add yet another addition, triggering further responsive moves by the others. And so on, along the familiar group-mediated, mutually responsive, self-augmenting “procyclical” trajectory.

In time, these continual rounds of home additions might come to be experienced as significant burdens by the homeowners. Nobody really wants all that space for purposes of habitation, let alone the pecuniary expenses incurred in acquiring it. They just all want, like the residents of Lake Woebegone, to be “above average” – or at least to be sure that they are not below average. The problem is that this competition can never be won, and it is very expensive – not to say wasteful – to continue on the resultant treadmill. At the same time, as with the arms race, so here no one can unilaterally end the race — all one can do is surrender and “unilaterally disarm,” which is precisely what renders the race a species of “treadmill.” Positional goods arms races of this sort occur not only in the realm of housing, of course, but across many spheres of potentially conspicuous consumption activity.13 Hence, there is reason to suppose that all might be rendered better off if such “races” might be contained via some commonly applicable method. We will come to that in section 4.

3.3 Labor-shedding recessionary “spirals”

Now let us consider a number of firms that are in competition with one another. At some point, some “shock” to the macroeconomy – a brief fuel price spike or stock market drop, for example – might yield a diminution in demand for these firms’ products that would, absent any responsive measures taken by the firms, be merely temporary. Because the duration is uncertain, however, and because these firms are in competition with one another, such that one firm’s greater profitability than another’s will typically result in higher capitalization by investors, each firm will be tempted to shed some of its labor. This, after all, will be profitable, at least in the short to medium term, given the decline in revenue wrought by the demand diminution wrought by the “shock.”14 And seeking higher profit and growth is widely considered to be rational in this context. To relinquish profits is to lose investors, incur a higher cost of capital and thereby risk individual downward spiraling into insolvency.

But now, of course, every firm’s rational actions, in the sense just described, results in aggregate demand contracting yet further economy-wide. For unemployed laborers do not purchase as much as employed laborers. The resultant contraction in demand now replicates the effect of the earlier “shock.” Revenues decline further, meaning that individual firms find it rational to shed yet more labor to protect profitability. This then collectively contracts demand all the further. The cycle continues. In effect, then, we have here another arms race, with labor-shedding playing the same role as arms-stockpiling. And just as in that case, so here, no one participant can stop the process. To try that is to try unilateral disarmament, thus to risk losing precisely what you shed labor to retain – profitability. So the process repeats itself – each round recurs – in procyclical fashion, “spiralizing” downward toward an intolerable equilibrium.

---

13 The term “positional goods arms race” appears to originate with my colleague Robert Frank, who has done more than almost anyone to draw attention to the phenomenon in question. See, e.g., Frank, R., 1986, Choosing the right pond; Frank, R., 2012, The Darwin economy; and a host of other works in between. The characterization of these “races” as recursive collective action problems appears to originate with the present author. See Hockett, R., 2013, “The libertarian welfare state, Challenge 56:2, 100–114. The locus classicus in introducing competitively “conspicuous” consumption is of course Veblen, T., 1899, The theory of the leisure class. Another evocative example of a positional goods competition, that involving standing audience members at a theatrical performance, figures in a recent op-ed by one of the present author’s mentors in support of a legal initiative launched by the present author to address that massive recursive collective action problem which is the continuing slump in the mortgage market and consequent continued debt deflation in the national macroeconomy. See Shiller, R., 2012, “Revising real estate requires collective action,” New York Times, June 23, available at http://www.nytimes.com/2012/06/24/business/economy/real-estates-collective-action-problem.html?_r=0. For more on the sense in which slumps and debt deflations constitute recursive collective action problems, see infra, Parts 3.3 and 3.8.

14 Of course the example is, like those that precede it, again stylized. It is sufficiently costly to shed and then reemploy labor that firms expecting the diminution in demand to be brief will not bother laying off labor. When, however, the diminution is expected to continue for some time to come, or when sufficient uncertainty attends the question, it will grow correspondingly “rational,” cost-wise, to shed the labor.
Of course, as with the arms race example, so here the envisaged scenario is stylized. Real slumps are more nuanced than this, with many firms seeking to avoid layoffs for as long as they can for any number of reasons, some firms being less affected than others, and so forth. But the recursive and “paradoxical” structure of the problem is discernible nevertheless even in more nuanced reality. What is individually rational to do, and irrational not to do, proves to be collectively irrational to do and individually irrational not to attempt to redress. It is self-defeating, as well as self-worsening, via a process of iteration. Such is the structure of self-worsening “downward spirals” and ongoing “slumps” in a macroeconomy.

3.4 Bank “runs”
Now consider a system of fractional reserve banking without deposit insurance. All depositors in a certain bank know that the bank does not have sufficient liquid assets on hand to accommodate all of them should they all decide to withdraw all of their funds at one instant. This does not ordinarily trouble them, however, because (1) it is in general highly unlikely that all depositors will wish to withdraw all funds at one time, and (2) the bank’s investing a large portion of all deposits in assets less liquid than cash enables the bank to earn returns that are passed on in part to depositors in the form of interest borne by their accounts.

But suppose now that some rumor or other piece of information apparently bearing upon the bank’s ultimate solvency, which for whatever reason is not altogether implausible on the surface even if happening to be factually false, begins to spread among depositors. Each depositor knows that, in the event that the bank’s solvency really is shaky, the bank will not be able to pay all of its depositors what it owes them. Hence, any depositor who is early in the queue to withdraw funds will be less likely than other depositors to lose anything. At the same time, should the information that has occasioned concern ultimately prove untrue, any such depositor can simply place their funds back in the same or some other bank without loss once the truth is established.

So they lose virtually nothing by going ahead and withdrawing early, while they risks losing everything if they do not.

What all of this means, of course, is that all depositors can find it individually rational to seek to beat all of the others to the bank to withdraw, and indeed individually irrational not to do so. No one depositor can end the run unilaterally. It also means that, with each withdrawal by someone other than you, the likelihood grows that you might lose something or everything should your bank fail. In effect, then, a form of recursive “bum’s rush,” much like an arms race, can commence, with every individual withdrawal rendering further such withdrawals that much more prudent – more rational – and hence likely. The run is in this sense self-reinforcing. It is recursive.

If a race like this does commence, and the bank is actually solvent at bottom, the depositors will then have brought on the very eventuality in contemplation of which they have acted. The liquidity crisis with which they confront the bank will, in such a case, have transitioned into a full-blown solvency crisis, leaving most depositors empty-handed in the aftermath. And this can be so, again, even where the bank is, absent the run, perfectly solvent at its core. What is more, inasmuch as the “information” that precipitates such a run on one bank often can be as little as “news” to the effect that there have been runs on some other banks, the “run” phenomenon can bring down an entire banking system – indeed, before deposit insurance, it often did. Such is the sense in which bank runs are often described as “contagious.” So contagion itself can be underwritten by what is, at bottom, a recursive collective action problem.

As with the arms race and labor-shedding race, then, so with the race to the bank during a run: multiple individually – and at each iteration, increasingly – rational decisions here aggregate into a collectively – and with each iteration, increasingly – irrational, self-defeating outcome for all but a very few depositors. The only exceptions are those who are first in the queue, who might be analogized to arms-racers who succeed in launching preemptive strikes against other arms-racers. It is perhaps theoretically possible for someone to “win” such races, but, by far, most parties “lose,” with the likelihood of winning sufficiently low for each racer as to render the race worse, in a probabilistic sense, for each racer than if there was no race at all. And yet also here, as again in the

15 As further elaborated in footnote 14.
16 They suffer merely the inconvenience of the withdrawal and ultimate redeposit.
17 See, e.g., the colorful instances in past U.S. banking history recounted in Gary Gorton, G., 2010, Slapped by the invisible hand.
18 For an account of that “contagion” that marked the “Asian Financial Crisis” of the late 1990s, see, e.g., Hockett, R., 2002, “From macro to micro to “mission creep,” Columbia Journal of Transnational Law, 153, 177-190.
3.5 Consumer price “inflations” and “hyperinflations”
Now suppose that it is Tuesday, and that you ordinarily purchase your groceries on Fridays. Some of your neighbors typically do likewise, while others typically shop on Thursdays and still others shop on Wednesdays. All of you now suddenly notice that prices advertised on the grocery store billboards are a bit higher today than they were yesterday. Some of you had also noticed, last week, that prices then were a bit higher than they were several days earlier. Either today or tomorrow or Thursday, it might begin to look likely, then, that prices will be a bit higher later this week than they are now — even if, in fact, whatever induced the earlier price rises is no longer at work by this point.

Against this backdrop, it might well appear to be individually rational for you to purchase your groceries today, Tuesday, rather than waiting till Friday. Prices might very well, after all, be higher by then. Matters will look likewise to your neighbors: today, Tuesday, looks better than Wednesday or Thursday as well — for again, prices might well be higher by then. So you and your neighbors rationally accelerate your purchase of groceries. You buy today, Tuesday, rather than later in the week when prices might be higher.

But note that, by acting rationally in this manner, by purchasing your groceries today rather than later in the week, your collective actions can bring on or worsen the very eventuality in the contemplation of whose possibility you have chosen to shop now — to wit, higher prices. All of you are driving prices yet higher than they would otherwise be by accelerating your demand for the products you are buying. But that fact cannot warrant refraining from purchasing on the part of any of you acting separately, for none of you can single-handedly prevent prices rising any more than you can stop the most salient losses wrought by an arms race by unilaterally disarming. You are accordingly once again damned if you do, collectively, and damned if you don’t, individually. And, matters of course, continue to unfold in this manner, and indeed worsen, indefinitely. For your all-inducing price acceleration gives you reason to accelerate purchasing behavior yet further, thereby inducing yet more rapid price acceleration, and so on, procyclically.

3.6 Asset price “bubbles”
Now consider again the immediately preceding example — that of a consumer price inflation — but change the object of purchase from groceries to financial assets of one sort or another. In addition, for at least some cases, add one more element to the story — viz. the availability of credit at low cost to purchase the assets in question, which later one can then sell after appreciation to pay off one’s debt while still pocketing a non-trivial margin. In such cases, we shall have told the tale of an asset price bubble, which is merely a price inflation in respect of financial assets rather than groceries. In the credit-augmented case, for its part, we shall simply have told the tale of a particularly dangerous — because even less amplitude limited — rendition of such inflation. Indeed, where credit comes into the picture, the asset price bubble can readily become something more than a mere inflation. It can become a full hyperinflation, by which price rises occurring at, say, velocity “v” during one iteration directly induce further rises at velocities vastly exceeding v over subsequent iterations.

There are two reasons for financial assets being particularly vulnerable to this pattern, both of which constitute sides of one coin. First, people often — perhaps typically — purchase financial assets these days less to consume or even to “hold” them than to “flip” or resell them. Hence, there is no natural “satiation” point where their purchase is concerned, as there is in the case of most consumer goods. Buyers will buy for as long as the prices are rising, and indeed aim to profit by sales after such price rises in ways that they typically do not, save in exceptional circumstances, in the case of consumer price inflations. Second, some such dynamic as this is what characterizes any self-reinforcing, procyclical consumer price inflation or “spiral,” as they used to be called. On the one hand, it is individually rational, once prices seem likely to rise somewhat in future, to act preemptively by accelerating one’s purchasing decisions. Insofar as one’s expectations of such price rises are plausible, moreover, it is symmetrically irrational not to do so. And yet, everyone continuing on the treadmill collectively brings on, and steadily worsens, precisely those losses that stepping onto the treadmill is meant individually to minimize.

19 For present purposes, “low-cost” credit can be either or both (a) credit available at low interest rates and (b) credit available on the basis of little down payment or collateral — hence high-leverage availability.
because, in light of the first point, people will rationally borrow in order to buy for as long as the asset price rises exceed rises in credit costs – something that is, again, much less common in consumer goods markets, even when consumer credit is available – there is no “natural,” consumption-rooted limit to the price levels that might be attained by financial assets. These can rationally rise for as long as credit remains sufficiently cheaply available to purchase them.20

Since a credit-fueled asset price bubble is just a (particularly dangerous) form of inflation, it bears the same rationality structure as inflation. It is individually rational, instrumentally speaking, for market participants to borrow in order to buy financial assets for as long as their prices rise faster than credit costs – for as long as, that is, there is a “spread” for speculative buyers to arbitrage or “leg.” By the same token, however, it is just as collectively irrational, again instrumentally speaking, for market participants to keep participating. For at some inherently uncertain point, the credit will run dry, people will begin selling off assets to pay down their debts left from earlier credit-financed purchases, and prices will then commence dropping at least as precipitously as they rose, leaving debt deflation and consequent recessionary spiral in their wake.21 And yet, once again, no individual can unilaterally terminate either process.

It should also be noted, if only in passing, how the individual rationality of decisions to purchase assets during the course of a bubble can, in effect, radiate outward, rendering other decisions at least temporarily rational as well. It is more rational, for example, at least for a time, to lend to an otherwise less worthy borrower during a credit-fueled asset price hyperinflation, at least if the asset to be purchased with the loaned funds is itself to collateralize the loan; for the expected value of the loan rises with the value of the collateral. It is likewise more rational, for similar reasons, to “extend” oneself further as a borrower under such circumstances, as well as to assign higher credit ratings as a rater to borrowers. And if one is in competition with others, it is not only individually rational to do these things, but individually irrational not to do them – for reasons similar to those that oblige firms to shed labor during recessions.

It also bears noting that it is even individually rational under the circumstances here described, for some of the same reasons, for some “functional” regulators to permit more on the part of borrowers and lenders than they would otherwise do – at least if, like other functional regulators, they are regulating individual persons or institutions for their individual safety and soundness rather than entire financial systems for their “systemic stability.”22 That last proviso offers a hint as to what we shall have to conclude in section 4 in the way of how to address recaps. The key is to supplement functional “microprudential” regulators, whose mandates tend toward procyclicality in many cases, precisely by virtue of the recursive collective action problems to which their constituencies are subject, with system-focused, countercyclically acting, “macroprudential” regulators, who are optimally situated to adopt regulatory strategies of the type discussed in section 4.23

3.7 “Fire sales” and asset price “busts”

Now consider the bank run example adduced in section 3.3. And as we did just now in section 3.6, alter the object of the activity in question. Change the asset from which one “runs” from a bank deposit of the section 3.3 variety to a financial instrument of the section 3.6 variety, saleable on the securities markets. In such a case, we shall have a “run” on assets, a.k.a. “fire sale,” which is structurally identical to a run on a bank.

A holder of the asset receives plausible information – perhaps no more than rumors – concerning the issuer’s or asset’s ultimate soundness. They know that, if others receive the same information and begin selling their holdings, the asset will quickly lose value in the market. It is accordingly in their interest – it is individually rational, instrumentally speaking – to sell their holdings before others sell theirs. That way they minimize pecuniary loss and salvage value. This is all the more true


21 The latter is, of course, discussed supra, section 3.3. For more on debt deflations, see infra, sections 3.7 and 3.8.


23 See again sources cited idem.
insofar as they can always purchase the asset back – possibly even at lower, post-panic cost – should the earlier information prove ultimately unfounded.

Once again, however, what is sauce for the goose here is sauce for the gander: it is just as individually rational for most or all other holders of the asset to aim to be quick to shed holdings as it is for the first agent. But lo, everyone acting instrumentally rationally in this way brings on the very result that was feared – a precipitous drop of the asset’s value and associated collective calamity. And each such drop induces further such drops at accelerating rates, per the familiar recursive, feedback-fed dynamic. As in the bank run example, moreover, all of this will be so, even if the rumors prove ultimately to have been unfounded, and no individual can stop the rumor-fed process from occurring.

At the same time, however, the run on assets example stands to the bank run example rather as the asset price bubble example stands to the consumer price inflation example: losses are magnified here because busts follow bubbles in which agents have incurred significant debt in making their bubble-magnified purchases. Those debt obligations do not fall with the prices of assets that the borrowed funds have purchased, meaning that people are left with significant “debt overhang” in the wake of these busts in a manner they are not after typical bank runs. They are left “underwater.”

This is precisely why recessionary spirals of the kind considered above in 3.4 are most severe after credit-fueled asset price bubbles and busts – which the latter themselves are among the most forceful of “shocks” of the sort mentioned there. Those with debt overhang do not spend as do those without overhang. They (slowly) pay down debt. Hence, Irving Fisher’s profound diagnosis of the depression of the 1930s as a “debt deflation.”

Much of the world is living through such a debt deflation even at the time of this writing – a deflation following on from that asset price hyperinflation that went into reverse, then to bust, over 2007–09.

It also bears noting here that, just as in the case of the asset price bubble, so here with the bust, that which renders it instrumentally rational for individuals to act in manners that aggregate into collectively dysfunctional outcomes also can render it rational for other agents to act reinforcingly. It is financially rational, for example, for a creditor who sees their debtor’s collateral plummeting in value either to accelerate the debt or to place a “margin call” demanding that the collateral be topped-off – all the more so if mark-to-market accounting practices are widely followed, or even imposed, by regulators.

It is similarly rational for microprudential regulators who key liquidity or capital requirements to asset risk to demand more buffer of regulated entities whose asset values begin dropping during such runs. But that, in turn, further fuels such fire sales – liquidations – of assets. Creditors and microprudential regulators accordingly act procyclically here just as they do during the price run-ups that are the antecedent bubbles. They do so because they are individually rational while caught in the grip of a recursive collective action problem. The problem is ultimately soluble only by means of the kind sketched in section 4.

3.8 “Debt deflations” and “paradoxes of thrift”

Now recall the labor-shedding example of section 3.4, and consider the sense in which it amounts to a means by which agents rationally endeavor to cut expenditures – labor expenditures – and thereby save funds. Well, there are, of course, other ways to cut expenses. One can simply refrain from spending on other things – things such as consumer goods. Any time that one holds off on spending – including spending on investment goods – in turn, one saves. And at times, it can be individually rational to do so – as the labor-shedding example itself might have made plain. In the wake of an asset price bust of the sort considered in section 3.7, for example – particularly during a debt deflation of the sort also there noted – there is every reason for individuals to hold off spending and repair personal balance sheets.

But lo, as in the case of labor-shedding, so here: what is individually rational can be collectively and self-augmentingly self-defeating, insofar as everyone saving causes further economic contraction, hence lower aggregate spending still, hence yet more contraction, and so on. Ironically, each person saving more can result in all people – that is, the macroeconomy – saving less.
Recursive collective action problems: the structure of procyclicality in financial and monetary markets, macroeconomies and formally similar contexts

in these cases. For gross income shrinks, individual incomes thus shrink, and ultimately there’s less to save.

This particular tragic irony bears a name — it is the celebrated Keynesian “paradox of thrift.” The paradox is paradoxical precisely because what is individually rational proves to be collectively irrational. And the irony is tragic, precisely because it would be just as individually irrational, absent some means of acting collectively, for any one person to eschew the individually rational action that aggregates into the collectively irrational outcome. Such are, of course, the paradox and tragic irony that afflict all collective action problems. It is simply more acute when the problem is recursively self-worsening.

4. Addressing the problem through its structure
Recursive collective action problems as just characterized are both formidable and, often, devastating. That they are potentially devastating is clear from the brief sampling of particular manifestations provided in section 3, all of which involve not merely irony, but indeed “tragic” irony. That they are formidable is presumably clear from their “paradoxical” – their “ironic” – characteristics as analyzed in section 2: they are collective irrationalities that spring from the aggregation of individual irrationalities, the latter of which we cannot expect – or even sensibly wish – to extirpate. What to do then?

The key, I suggest, lies in focusing upon two of these problems’ three basic features: first, the collective aspect of their collective irrationality; and second, the rational aspect of their individual rationality. To solve our most formidable recursive collective action problems, in short, requires our first assuming collective agency and then collectively acting to render certain individual decisions – those that aggregate into collective calamity in each kind of recap – no longer instrumentally rational. The remainder of this section briefly elaborates on this fundamental idea.

4.1 Collective rationality via collective agency
For collectively irrational outcomes to be possible, something has to be missing – some prerequisite to collective rationation and resultant action. What is missing, in fact, is a locus or situs at which collective rationation and associated rationality can operate. In short, what is missing is collective agency, at least in the sphere of activity in question – arms stockpiling, home buying, asset purchasing, or what have you. To solve a collective action problem in these spheres requires collective action itself in these spheres, hence some form of collective agency. The relevant collectivity must be able to act in concerted fashion, either directly or via some agent duly authorized to act in the name and on behalf of all in the collectivity. Where the collectivity in question either is, or is part of, a polity or some other aggregate of persons in whom the attributes of sovereignty vest – that is, a state – the most common form of such agency is a government or government instrumentality. In a state or other polity, government is the collective agent par excellence. It is the collective agent under whose ultimate collectivity-vested authority, and with whose supplemental assistance of various kinds, all individual and other, substate collective agents will operate.

Even the most diehard of libertarians in effect acknowledge this collective agency function of government – when, for instance, they say that the “first duty” or “essential role” of governments is “national defense.” The reason is that collective defense itself poses a classic collective action problem – namely that subspecies typically referred to as the “free rider problem.” It is individually rational, the story here runs, for each member of a collectivity in need of defending to let others do the job or foot the bill where defense is concerned, then to enjoy the positive externalities generated by their doing so. Border defense, after all, is a largely non-excludable public good. But this means that defense will be systematically underprovided if its provision is left to voluntary contributions of effort or funds; for it means that the latter is individually irrational. Hence even in libertarian fantasies, for collective defense at least, the collectivity appropriately authorizes an agent – a government – to “provide for the collective defense.” In so doing, they solve a classic collective action problem.

25 The locus classicus in this case is Keynes, J. M. 1936, The general theory of employment, interest and money. For a recent discussion of the causal connection between present day debt deflation and savings-undercutting thrift, see Alpert, Hockett, and Roubini, idem.

26 The locus classicus in this case is, of course, Olson, M., 1962, The logic of collective action, the ubiquity of which is such as to lead the present author almost to wish that he did not have to cite it. For a skeptical take on the utility of “free rider problems” as explanations or justifications of sundry policy measures, see Tuck, R., 2008, Free riding.
How governments discharge the appointed function, however, is as instructive for present purposes as the fact that they do it. For one way of looking at conscription, mandatory subscription — a.k.a. taxation — or both, is a means by which those collective agents known as governments render it no longer individually rational to do that which, when all do it, results in collective, and hence individual, calamity. Collectivities in these instances, in other words, charge their governments with rendering it no longer instrumentally rational for individuals to attempt to free-ride or shirk. Here lies the key to understanding how best to address all other recaps, including those sampled in section 3.

4.2 Conforming individual to collective rationality by collectively changing the individual calculus

Collective agency is necessary, but not sufficient to solve a collective action problem. What else is required? The key to our answer lies in that other constitutive feature of the collective action problem additional to aggregation — viz. the individual rationality of the decisions that aggregate into collectively irrational outcomes in these situations. The duly appointed collective agent must, in short, act in the name of all to change the calculus of each, such that certain erstwhile individually rational decisions that aggregate into collectively irrational outcomes cease to be individually rational.

How to do that? It is actually quite simple. Return for a moment to the collective defense example just considered. The individually rational decision that aggregates, when replicated by multiple actors, into calamity, absent a collective agent in the form of a government, is the decision to shirk or to free ride. What does a collective agent then do in this circumstance to avert collective calamity? The government simply changes the free rider’s calculus in either, or both, of two ways. It either (1) conscripts the free rider, in effect forcing them to take up arms on pains of imprisonment or worse, (2) conscripts the free rider’s resources, again on pain of imprisonment or worse, so as to fund the collective defense, or (3) does both. Either way, it acts to render free-riding no longer “free,” hence no longer individually rational.

In effect, government instrumentalities must do the same thing in solving each of the problem instances considered above in section 3. All that differs from case to case is the particular means employed, which of course varies according to those salient particulars which distinguish one recap from another.

In the case of an arms race, for example, parties wishing to stop the operation of the treadmill simply merge into — they jointly constitute — the requisite collective agent via the medium of contract (in this case, treaty), then mutually authorize subagents to inspect certain sites and verify mutual compliance with the agreement. This they can do with relative ease in view of their small number. Insofar verification then can be reasonably relied upon, it becomes no longer rational to produce or procure more weapons. For doing so will incur significant costs — those occasioned both by the new weapons purchases and by subsequent iterations once the race is renewed — while affording no offsetting benefits, any of which are quickly cancelled out by retaliatory procurements by competitors.27

Positional goods arms races are handled in a subtly different, but functionally equivalent, manner to that employed in the case of actual arms races. Because individual market participants are not themselves states or distinct loci of sovereignty, they need not content themselves with simulated collective agency in the form of contract or treaty — nor are they apt to be able to do so if numerous. Instead, then, they delegate to their already existent collective agent — their shared government — the task of rendering “racing” no longer individually rational. How? The simplest way probably is to employ graduated taxation schemes that render each later stage in the iterative “racing” process proportionally more expensive to each “racer” than earlier stages. That way, escalation in the race is phased out, and each erstwhile competitor is able, in effect, to rely upon others to attribute their decision to stop racing to the presence of the taxation regime, rather than to pretext incapacity to compete.28

The best-known collective solution to the bank run problem, as well as one such solution to a run on financial assets problem, involves rendering it no longer individually rational to “run.” In the bank run case, deposit insurance is, of course, the response in question. By assuring each depositor that any loss, up to some reasonable ceiling, occasioned by a bank’s failure will be fully compensated, the deposit insurer eliminates the principal factor — possible loss — that renders “running” individually rational. It thereby prevents mere liquidity troubles from morphing into full-on insolvencies. While in theory this function can be, and

27 See generally work by Schelling and Halperin & Schelling cited supra, note 11.
28 See generally work by Frank, Hockett and Shiller cited supra, note 13.
It should also be noted that deposit insurance, though inherently countercyclical at least in virtue of its run-stemming properties, can be designed in more and in less countercyclical renditions. The U.S. Federal Deposit Insurance system, for example, until 2005, operated procyclically where its method of premium assessment was concerned. For it assessed premia only when the insurance fund fell below a stipulated floor level, which, of course, typically occurred only during times of multiple bank failures — meaning assessments were made principally during times of stress in the banking industry. Happily, the U.S. Congress rectified this in 2005 and 2006, through legislation that effectively renders the assessment system countercyclical. In effect, then, this legislation implicitly recognizes deposit insurance’s function as a solution to a recursive collective action problem.

Turning now to the case of a run, not on bank deposits, but on financial assets, the collective agent now acts to render it no longer individually rational to run in a manner analogous, but not identical, to that afforded by deposit insurance. Here, the agent — typically, but not necessarily, a central bank — announces its willingness to purchase all assets of a particular type at some floor rate which, on the one hand, amounts to a “penalty rate” relative to the asset’s price prior to the run, while, on the other hand, constituting a rate well above market rate during the run.

Playing the “market maker of last resort” in this manner, the collective agent operates from “two sides,” as it were, to render it no longer individually rational to dump assets. From the down side, the collective agent renders it no longer worthwhile for the asset-holder to sell the asset at a price below the announced threshold, hence effectively preventing the asset from plummeting below that threshold pursuant to the familiar self-fulfilling prophecy dynamic. From the up side, in turn, the collective agent affords, via the penalty in the penalty rate, incentive for asset holders to hold on to their assets rather than selling them even to the collective agent — at least, where there is reason to suppose that the asset is being systematically undervalued by the panic-struck market, such that its value will recover in time once it has weathered the storm via other holders continuing to hold the assets by virtue of the collective agent’s guarantee.

It also bears noting that, as in the case of deposit insurance, so here the collective agent who acts in the requisite manner to address an acute recursive collective action problem acts countercyclically and, in that sense, macroprudentially. The agent acts countercyclically in blunting the recursion, and macroprudentially in acting on behalf of the full collectivity of participants in the “macro-market” in question — the financial system as a whole. Any finance regulatory regime that would aim to be complete, must include at least one such countercyclically acting macroprudential regulator — a regulator both cognizant of this function and able to discharge it along the lines sketched in this section.

In the case of inflations, debt deflations, paradoxes of thrift, credit-fueled asset price bubbles and layoff-led recessionary spirals, again the key to solving the problem is for the collective agent to render no longer individually rational those decisions that tend to aggregate into the collectively irrational consequences in question. In these cases, however, the means of doing so combine individual decision-targeting with more macro-level action.

Begin with the credit-fueled asset price bubble. The key to this phenomenon, recall, is the spread between low borrowing costs and high capital appreciation rates during the boom, which it is individually rational for market actors to arbitrage, or “leg.” One means of rendering participation in the bubble dynamic no

---

29 See, e.g., Sontag, supra note 17; for more both on the historic role of clearing houses in addressing runs, and on the design of the Federal Reserve System as a form of publicly backstopped, quasi-private clearinghouse arrangement.
30 See Hockett and Hockett and Omarova sources cited supra, notes 20 and 22. See also Bagehot, infra note 31, who first articulated the importance of the “penalty rate.”
31 Idem. The canonical articulation of the vaunted “lender of last resort” role played by central banks is, of course, Bagehot, W. 1873, Lombard Street: a description of the money market. The ways in which “market-making” has become the new “lending” for these purposes is discussed in Hockett and Hockett & Omarova as cited supra, notes 20 and 22, as well as in Mehring, P., 2011, The New Lombard Street.
32 See again Hockett and Hockett and Omarova sources cited supra, notes 20 and 22.
longer individually rational, then, is for the collective agent, now acting as a countercyclical macroprudential financial regulator, to close that spread from either, or both, of the two ends. The authority can thus tax capital gains at gradually higher rates, tax asset “flipping” transactions à la Tobin, tax credit extensions at gradually higher rates or simply place hard limits on extensions of credit or incurs of leverage — via lending limits, debt-to-income (DTI) or loan-to-value (LTV) ratio maxima, or reserve and capital minima.33

Turning from individual decision-targeting to more macro-level action, the collective agent can also rein in credit by blunter means — for example, by raising interest rates via traditional monetary policy instruments.34 Alternatively, the regulator might target particular classes of asset that seem to be prey to bubble dynamics at a given time, by “shorting” those assets in the markets on which they are traded.35 This would lie somewhere between more individual decision-targeting action such as margin requirements on the one hand, and more macro-level action on the other. In effect, it would be a bubble-side counterpart to run-side last resort market-making.

Now consider a consumer price inflation. Here, the more familiar means of collectively addressing the problem is for the collective agent — in most jurisdictions, the central bank or monetary authority — to render accelerated purchases no longer individually rational by reining in the credit-money supply via traditional monetary policy measures: more restrictive discounting, interbank rate setting and open market operations to close that spread from either, or both, of the two ends. The hope is that the authority will induce market participants into holding off on their purchases longer, thereby draining off price-pushing pressures brought on by excessively high-frequency purchasing activity. In so doing, the authority in question is, of course, once again acting countercyclically. And in doing that, it is implicitly acting to solve a recursive collective action problem.

Though the method of “tight money” is more broad-brush and blunt than the more carefully targeted macroprudential tools mentioned in the previous paragraph but one, it bears noting that those more narrowly targeted tools could, in effect, be at least partly approached in the realm of consumer goods and services markets. A central bank or monetary authority could, for example, as noted in the previous paragraph, engage in open market operations that target financial instruments other than treasuries, thereby tamping down inflations that afflict some submarkets more than others.36 By the same token, it could place a floor under some goods or services while simultaneously pushing back on inflationary pressures in respect of other goods and services. In a sense, innovation of this sort is, of course, already under way, via a form of the “market-maker of last resort” function noted, in the case of Fed treatment of housing-associated financial assets.37

Turning now to debt deflations, paradoxes of thrift and layoff-led recessions, here again the collective agent can render no longer individually rational such decisions as aggregate into the collectively dysfunctional outcome through a combination of more narrowly targeted and more broadly sweeping methods. In the case of a debt deflation and associated paradox of thrift, for example, the debt overhang that renders it individually rational to hold off on spending can be trimmed from either, or both, of two ends — the debt end and the equity end, so to speak. To pare back debt overhang from the debt end, narrowly targeted means — in particular, principal reductions — do the trick.38 To work from the equity end, either narrowly targeted means such as those now being employed by the Fed to maintain a floor

33 See again Hockett and Hockett and Omarova sources cited supra, notes 20 and 22.
34 See again Hockett and Hockett and Omarova sources cited supra, notes 20 and 22.
37 Idem.
under home prices, or broader monetary and fiscal policies, can be employed, the choice among them to be determined by reference to the comparative unintended follow-on effects apt to be associated with each.39

In the case of a layoff-led recession and associated paradox of thrift, again, the collective agent can employ either or both narrow and broader means of rendering the decisions that aggregate into the collectively dysfunctional outcome no longer individually rational. Broadly, for example, the fiscal and monetary policy measures already alluded to have the effect of increasing aggregate demand in a manner that boosts incomes and, thereby, renders purchasing and, ultimately, employment activity more attractive. Similarly, were a government to adopt countercyclical, “employer-of-last-resort” hiring policies providing employment to all who sought it, but for temporary intervals could not find it, it would be employing a relatively broad-sweeping measure tending directly to increase purchasing power and indirectly to increase wages hence purchasing power, thereby rendering consumer spending activity more individually rational economy-wide.40

More narrowly contoured means toward the same ends include, for example, negative interest, stamped money and dated voucher programs to discourage hoarding, and tax incentives to induce greater hiring and/or employee retention by employers.41 In all such cases, again, the collective agent in question would be acting countercyclically in addressing a recursive collective action problem. We could go on to adduce further such examples, but presumably by this point the case has been made.

5. Conclusion
We have covered a good bit of ground here – one hopes, for the present, enough such ground. None of the policy challenges addressed here are unfamiliar, nor is the notion of a collective action problem in connection with which we have considered them. What might be new here, however, is our bringing them together in thought in a manner that helpfully illuminates both their shared structure and, therefore, the sorts of means requisite to solving them. The value added, in short, assuming there’s any, lies in our highlighting two facts. First, that these particular problems are, in fact, classic collective action problems – indeed, particularly devastating renditions of the same by virtue of their recursive properties. And second, that these hitherto intractable problems are accordingly soluble by recourse to familiar collective means – means whose legitimacy as responses to classic collective action problems is readily made plain, and which work by rendering no longer individually rational such decisions as tend in aggregate to issue in such problems.

Assuming that highlighting those facts has been valuable, however, it still is the case that more remains to be done. For one thing, there might well be other hitherto intractable challenges that face polities, which are in the nature of recursive collective action problems but have not been noticed as yet to be such, including by the present author. For another thing, there might be additional means of addressing such problems that have thus far escaped notice, including again that of the present author. It is to be hoped, then, that by (a) identifying the species of challenge here considered, (b) drawing out the shared structure common to all instances of this species of challenge and finally (c) identifying what all solutions to such challenges must therefore have in common, we have at least facilitated the discovery of both more such challenges and more means of addressing the same.


40 A suggestion along these lines is offered and elaborated in Hockett and Omarova, supra note 22. See also Wisman, J. D., 2013, “The growth trap, ecological devastation, and the promise of guaranteed employment,” Challenge 56:2, 53–78; and Wray, R., 2013, Modern money theory.

Do “too-big-to-fail” banks take on more risk?¹

Gara Afonso  
Senior Economist, Federal Reserve Bank of New York²

João A.C. Santos  
Vice President, Federal Reserve Bank of New York

James Traina  
Senior Research Analyst, Federal Reserve Bank of New York

Abstract  
The notion that some banks are “too big to fail” builds on the premise that governments will offer support to avoid the adverse consequences of disorderly bank failures. However, this promise of support comes at a cost: large, complex or interconnected banks might take on more risk if they expect future rescues. This article studies the effect of potential government support on banks’ appetite for risk. Using balance sheet data for 224 banks in 45 countries starting in March 2007, the authors find higher levels of impaired loans after an increase in government support. To measure support, they rely on Fitch Ratings’ support rating floors (SRFs), a new rating that isolates potential sovereign support from other sources of external support. A one-notch rise in the SRF is found to increase the impaired loan ratio by roughly 0.2 — an 8% increase for the average bank. The authors obtain similar results when they assess the effect of increased support on net charge-offs and when they narrow their sample to U.S. banks only.


² The authors thank Christian Cabanilla, Nicola Cotorelli, Mark Flannery, David Marqués-Ibañez, Stavros Peristiani, William Riordan and Tony Rodrigues for valuable comments. They are grateful to Alex Entz for research assistance. The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.
1. Introduction
In 1984, U.S. regulators made the unprecedented move of insuring all of Continental Illinois’s liabilities. The Comptroller of the Currency indicated, during the hearings after Continental’s resolution, that regulators would not allow the 11 largest banks in the U.S. to fail. Ever since, there have been many concerns with banks deemed “too big to fail.”

These concerns derive from the belief that the too-big-to-fail status gives large banks a competitive edge and incentives to take on additional risk. If investors believe the largest banks are too big to fail, they will be willing to offer them funding at a discount. Together with expectations of rescues, this discount gives the too-big-to-fail banks incentives to engage in riskier activities. This, in turn, could drive the smaller banks that compete with them to take on further risks, exacerbating the negative effects of having too-big-to-fail banks in the financial system.

The debate around too-big-to-fail banks has given rise to a large literature. Part of this literature attempts to determine whether bank investors, including depositors, believe the largest banks are too big to fail. Some studies seek to answer this question by investigating spreads on bank bonds (Flannery and Sorescu (1996), Sironi (2003), Morgan and Stiroh (2005), Anginer and Warburton (2010), Balasubramnian and Cyree 2011, Santos (2014)). Other studies consider spreads on bank credit default swap contracts [Demirgüç-Kunt and Huizinga (2013), Li et al. (2011)], bank stock returns [Correa et al. (2012)] and deposit costs [Baker and McArthur (2009)]. Yet others focus on the premiums that banks pay in mergers and acquisitions [Brewer and Jagtiani (2007), Molyneux et al. (2011)].

Another part of that literature investigates whether too-big-to-fail banks behave differently by looking at balance sheet data [Gropp et al. (2011)], syndicated loans [Gadanecz et al. (2012)] and bank z-scores [Brandão Marques et al. (2013)], among other measures.

Our paper is closer to the latter studies in that we are also interested in finding out whether the too-big-to-fail status affects bank behavior. Specifically, we study whether banks that rating agencies classify as likely to receive government support increase their risk-taking.

An important novelty of our paper is the way we measure the likelihood of a bank receiving government support. Previous studies, including Haldane (2010), Lindh and Schich (2012), and Hau et al. (2013), attempt to infer support from the difference between Moody’s all-in credit ratings (long-term bank deposit ratings, which capture a bank’s ability to repay its deposit obligations and include external support) and Moody’s stand-alone ratings (bank financial strength ratings, which exclude external support). The difference between Moody’s all-in credit and stand-alone ratings is commonly known as a ratings “uplift.”

Using uplifts, however, presents two potential issues. First, a change in uplift may arise from movement in either of the two underlying ratings, with completely different implications. Second, uplift incorporates any type of external support, including from governments, parent companies and other institutions.

To avoid the first concern, some studies rely on support ratings (SRs) issued by Fitch Ratings [Gropp et al. (2011) and Molyneux et al. (2010), among others]. As with uplift, SRs also include institutional, cooperative, local government and regional government support. We sidestep both problems by considering a new Fitch rating. Starting in March 2007, Fitch began to issue SRFs, which reflect its opinion of potential sovereign support only (including a government’s ability to support a bank). The main advantage of using this rating is that, in contrast with earlier approaches used in the literature, the SRF explicitly captures government support. That is, it does not incorporate other forms of external support, such as the institutional support of a high-holder in a banking organization to a bank within its own hierarchy.4

3 Continental Illinois, which was the seventh-largest bank by deposits, experienced runs by large depositors following news that it had incurred significant losses in its loan portfolio. Concerns that a failure of Continental would have significant adverse effects on other banks that had deposits with it, led the Federal Reserve Board, the Federal Deposit Insurance Corporation (FDIC) and the Comptroller of the Currency, together with 24 U.S. banks, to announce a U.S.$7.3 billion bailout. The rescue package comprised a U.S.$2 billion capital injection by the FDIC and the group of 24 banks and a U.S.$5.3 billion unsecured line of credit from the banks.

4 Fitch Ratings (2013a) explicitly defines SRFs as based on potential sovereign support (not on the intrinsic credit quality of the bank). In the case of the landesbanks, Fitch assumes that Germany’s and the German states’ creditworthiness are linked. For example, in August 2013, Landesbank Baden-Württemberg (LBBW) had an SRF of A+ even though Fitch does not rate the State of Baden-Württemberg. The assessment implicitly assumes that the creditworthiness of the support “is underpinned by the strength of the German solidarity system, which links the state’s creditworthiness to that of the Federal Republic of Germany (AAA/Stable)” [Fitch Ratings (2013b)].
The results of our investigation show that a greater likelihood of government support leads to a rise in bank risk-taking. Following an increase in government support, we see a larger volume of bank lending becoming impaired. Further, and in line with this finding, our results show that stronger government support translates into an increase in net charge-offs. Additionally, we find that the effect of government support on impaired loans is stronger for riskier banks than safer ones, as measured by their issuer default ratings (IDRs).

Our findings offer novel evidence that government support does play a role in bank risk-taking incentives. The results are also important because they already include the effects of the government interventions undertaken throughout the latest financial crisis. At the same time, however, not enough time has elapsed since the crisis for our results to reflect the impact of the regulatory changes enacted in its wake.

The rest of our paper is organized as follows. The next section introduces our measure of government support. Section 3 describes the data sources and characterizes our sample. Section 4 introduces our methodology. Section 5 discusses our results. Section 6 presents robustness analysis. Section 7 concludes with some final remarks.

2. Measuring the likelihood of government support

There are a number of different methods for measuring sovereign support based on rating agency assessments. Previous work uses two ratings published by Moody’s to derive a measure of government support [Haldane (2010), Lindh and Schich (2012) and Hau et al. (2013), among others]. Moody’s issues bank deposit ratings based on its opinion of a bank’s ability to repay punctually its deposit obligations. These ratings are all-in credit ratings that reflect intrinsic financial strength, sovereign transfer risk (for foreign currency deposits) and both implicit and explicit external support elements. Moody’s also issues bank financial strength ratings, which exclude sovereign risk and external support. Uplifts – calculated as the difference between these two ratings – provide an estimate of the implicit guarantees. This measure incorporates any type of external support (not just sovereign support), including institutional backing from parent companies. To control for this support, some recent studies exclude all bank subsidiaries from their samples and focus their analysis on high-holders of banking organizations only [Brandão Marques et al. (2013), among others]. Uplifts also capture cooperative, local government and regional government support.

Although intuitive, this methodology assumes a linear functional form for the difference between these two ratings, but the relationship between external support and stand-alone ratings may be more complex. It also makes it difficult to identify the source of variation in uplifts. For example, suppose there is a one-notch increase in the stand-alone rating, but no change in the all-in credit rating. Uplift would decrease, indicating weaker external support when, in practice, there has been no change. Moreover, even if both ratings were to change, differences in Moody’s publication timing would lead to spurious variation in external support.

An alternative approach relies on ratings issued by Fitch that explicitly measure external support, independent of the intrinsic credit quality of the bank. SRs rely on Fitch’s assessment of a
supporter’s propensity and ability to support a bank. Supporters can be of two types: sovereign states and institutional owners. Studies that use SRs include Gadanecz et al. (2012) and Gropp et al. (2011).

In addition to SRs, Fitch issues SRFs, based on its opinion of potential sovereign support only (including a government’s ability to support a bank).5

The main difference with respect to SRs is that SRFs do not incorporate external support other than sovereign support, such as the institutional support of a high-holder in a banking organization to a bank within its own hierarchy. Isolating the support coming from the government is crucial to addressing the question of whether too-big-to-fail banks increase their risk-taking, because, in contrast to other sources of external support, sovereign support is typically unpriced and not risk-sensitive. Figure 1 shows a comparison of these ratings-based approaches to measuring sovereign support.

To stress the difference between these two ratings, let us consider the case of Bank of America. Table 1 shows the history of changes in SRs and SRFs for Bank of America Corporation (the parent company) and Bank of America National Association (the largest national bank within the organization). Fitch expresses SRs on a five-notch, 1-to-5 scale, where a rating of 1 denotes a bank with extremely high probability of external support. SRFs use the AAA long-term scale, where AAA ratings indicate an extremely high probability of government support. SRFs include one additional point on the scale, “no floor” (NF), bringing the total number of notches to 20. According to Fitch, NF designates no reasonable presumption of potential support and translates to a probability of support of less than 40% [Fitch Ratings (2013a)].

From 16 March 2007 to 16 January 2009, Bank of America Corporation (the parent) had the lowest level of external support (SR = 5), while Bank of America National Association enjoyed the highest level of external support (SR = 1). By looking at SRs only, we cannot disentangle if the strong support of Bank of America National Association comes from the Government or from the parent company. To answer this question, we turn to its SRF. The SRF of Bank of America National Association was A– over this period, indicative of strong government support.

The evolution of Bank of America National Association’s SRFs also shows how sovereign support to the national bank heightened two notches in January 2009 and lessened one notch in December 2011, while external support (measured by SRs) remained constant. The difference in granularity between these two ratings is yet another advantage of using SRFs instead of SRs.

---

5 According to Fitch Ratings (2013a), support typically extends to the following obligations: senior debt (secured and unsecured), including insured and uninsured deposits (retail, wholesale and interbank); obligations arising from derivatives transactions and from legally enforceable guarantees and indemnities; letters of credit and acceptances; trade receivables; and obligations arising from court judgments.
since the former allows for higher precision and more variability in support.

A similar measure based on S&P ratings is currently not available since S&P does not issue ratings that allow measurement of sovereign support.

3. Data and sample characterization

3.1 Data

The data for this paper come from several sources. We use Bureau van Dijk’s Bankscope to gather balance sheet data on banks in our sample, including our key measures of bank risk-taking—impaired loans and net charge-offs. In addition, we use two data sets from Fitch Ratings: one containing information on government SRs (described in detail in section 2) and the other containing information on bank strength ratings (long-term IDRs). IDRs reflect Fitch’s opinion on an entity’s relative vulnerability to default on its financial obligations. IDRs are Fitch’s primary issuer rating for financial institutions and are expressed on an AAA long-term scale, where AAA ratings denote the lowest expectation of default. IDRs incorporate not only intrinsic strength, but also external support. Even though stand-alone ratings are a cleaner measure of a bank’s intrinsic strength than IDRs, we cannot rely on these ratings in our analysis because of the lack of a consistent time series during our sample period.

3.2 Sample characterization

To construct our data set, we start with the universe of banks that have SRFs, which Fitch began issuing on 16 March 2007. Though the most recent ratings are easily accessible online, historical ratings need manual collection. Our sample includes daily SRF observations for 612 banks (bank holding companies, commercial banks and savings banks) from 16 March 2007 to 15 August 2013. The data spans 92 countries, with 182 banks from the U.S.

Our sample of changes in SRFs comprises increases and decreases in ratings. The first change in our sample occurred on 2 July 2007, and the last one on 14 August 2013. There are 446 changes in SRFs (234 increases and 212 decreases) across 234 unique banks and 177 unique event dates. On average, each change shifts the rating about two notches.

We find support for the commonly held belief that foreign countries tend to provide stronger support to their banks than the U.S. does, with the average SRF of a foreign bank being about four times larger than that of a U.S. bank. Interestingly, this pattern changes dramatically when we zoom in on the set of banks with an SRF different from an NF rating: the “supported” banks. We find that the average sovereign support remains slightly humped in foreign countries (according to Fitch’s ratings), but the pattern changes significantly for the U.S., where, over the last six years, average government support has increased markedly. Since 2010, average sovereign support for U.S. banks has been stronger than that for foreign banks.

This difference in patterns seems to be driven by the larger proportion of U.S. banks that have a probability of government support lower than 40%. The data shows that 80% of banks in the U.S. have NF ratings, compared with 21% in foreign countries. Whether or not government support to banks is more prevalent in the U.S. than abroad depends on whether we take NF ratings into account. Making this distinction matters because it portrays a different picture of how government support has evolved in the U.S.

For information on credit quality and exposure to default, we use long-term IDRs issued by Fitch. For each bank in our sample, we obtain the history of changes in IDRs from 1 January 1988 to 15 August 2013. To present summary statistics on a comparable sample, we restrict our attention to IDR observations for which we also see an SRF. Figure 2 shows the

---

6 Historically, Fitch issued individual ratings on an A–E scale to assess a bank’s creditworthiness on a stand-alone basis. Similar to Moody’s bank financial strength ratings, these ratings aimed to capture the strength of a bank if it was unable to rely on external support. On 7 March 2011, Fitch announced a revision to the methodology used to calculate the stand-alone ratings, as well as a change from a 9-point scale (using letter ratings such as A and A+/B) to a lowercase variation of the traditional 19-point long-term rating scale (using letter ratings such as AAA and AA+). On 20 July 2011, Fitch introduced new stand-alone ratings called viability ratings, designed to reflect the same core risks as individual ratings but with renewed definitions and greater granularity.

7 As standard in the ratings literature, we assign numeric values to the notches on the rating scale, where a value of 19 denotes a AAA rating and zero an NF rating.

8 The heat map in Figure 3 highlights the unique character of the NF rating. At first glance, since SRFs act as a floor for IDRs, one might think the NF rating is located one notch below D on the SRF scale. However, the distribution of IDRs for banks with NF SRFs is significantly different from IDRs for banks with SRFs expressed on the AAA scale. While banks with SRFs ranging from CCC to AA+ typically have an IDR between zero to two notches higher, a bank with an NF SRF is more likely to have a BBB or A– IDR rating. This suggests a definition of average government support that excludes banks with NF ratings.
distribution of SRFs (left) and IDRs (right) for the sample of 612 banks.

Recall from sections 2 and 3 that SRFs reflect government support while long-term IDRs incorporate both intrinsic and external support. As such, a bank’s SRF acts as a floor for its IDR. Figure 3 highlights this relationship by presenting the distribution of IDRs by SRFs. The intensity of each symbol denotes the frequency (that is, a darker square indicates a more frequent relationship).

As expected, many bank ratings lie on the diagonal, indicating that Fitch’s assessment of a bank’s relative vulnerability to default and of a government’s propensity to support a bank are identical. The rest of the observations are on the upper diagonals of the heat map (Figure 3), which denote that the overall strength of a bank exceeds its sovereign support. It is also interesting to note that banks rated with a probability of sovereign support of less than 40% (SRF = NF) are rated with IDRs ranging from D to AA+. Having risky banks among those with a probability of sovereign support of less than 40%, suggests that risk alone does not drive the probability of government support. This would be the case, for example, for small banks that may not receive government support regardless of their overall financial strength.

Finally, we use the Bankscope database to augment the ratings data with quarterly information on bank characteristics spanning Q1 2007 to Q3 2013. Fitch issues SRFs at the entity level, so we keep in our sample parent banks and their subsidiaries when there are multiple entities for a consolidated bank in Bankscope. The matched sample consists of 11,929 bank-quarter observations for 601 banks.

Because of the global nature of our data, we are missing balance sheet information for approximately 59% of our bank-quarter observations for which we have SRFs. To alleviate this problem, we linearly interpolate adjacent data if they are missing for less than one year in duration. Interpolation recovers approximately 15% of our potential data, reducing the proportion missing to 44%. After matching and interpolation, we further limit our sample to banks with information on total assets, impaired loans, net charge-offs, tier-1 capital and trading assets. This step leads

---

9 Results are qualitatively similar in the analysis without interpolation.
to a final data set with 1,739 bank-quarter observations.

Most banks in the sample (75%) have investment-grade ratings. Many (38%) also have government support of BBB– or above. The median bank has total assets of U.S.$22 billion, while the average bank has assets of U.S.$200 billion. Size, however, changes significantly by level of government support, with highly supported banks being typically larger. The bank with a C–CCC rating (the lowest SRF in our sample) has close to U.S.$4 billion in total assets, while those with an AA–AAA rating are almost 100 times larger on average. Figure 4 shows this pattern, which is consistent with the literature that documents a positive relationship between size and government support.

Banks with a higher probability of government support also have more trading assets on average. However, as shown in Table 2, we do not find a similar pattern with return on assets (RoA), impaired loans, net charge-offs or tier-1 capital. In our sample, the average bank has an RoA of 0.27%, an impaired loan ratio of 2.48%, a net charge-off ratio of 0.59%, and a tier-1 capital ratio of 10.89%. Table 2 tabulates descriptive statistics for our sample.

4. Methodology and empirical strategy

The goal of our analysis is to investigate whether banks with higher government support engage in riskier activities. To test this hypothesis, we use a panel of bank-level data. After matching and interpolating, we further limit our sample to banks with information on total assets, impaired loans, net charge-offs, tier-1 capital and trading assets. This restriction leads to a final panel data set with 1,739 bank-quarter observations. Although 85% of our bank-quarter observations correspond to domestic banks, our sample retains a global nature, spanning 224 banks in 45 countries.

We first measure the riskiness of a bank’s activities by the ratio of impaired loans to total assets. We also present results for alternative measures of risk, including ratios of net charge-offs, net income, tier-1 capital and trading assets to total assets.10 Specifically, we investigate whether the ratio of impaired loans to total assets relates to government support of banks. Since we

---

10 Data on these risk measures are from Bankscope. In particular, we use the following series: DATA2170 (impaired loans), DATA2025 (total assets), DATA2115 (net income), DATA2140 (tier 1 capital), DATA2150 (net charge-offs) and DATA29190 (total trading assets).
The Journal of Financial Perspectives

Do “too-big-to-fail” banks take on more risk?

We expect that a bank's response to sovereign support might take time to show up on its balance sheet, we estimate specifications of our model with progressively higher lags for all right-hand-side variables. To that end, we estimate the following model:

$$R_{b,t} = \beta \cdot SRF_{b,t} + \delta \cdot IDR_{b,t} + \eta \cdot Assets_{b,t} + \mu \cdot \text{OtherRisk}_{b,t} + \gamma \cdot Z_{t} + \tau \cdot X_{t} + \varepsilon_{b,t}$$  \hspace{1cm} (1)

where $b$ indexes banks, $t$ denotes time in quarters, and $i = \{1, ..., 11\}$ indicates the number of lags. The dependent variable Risk$_{b,t}$ is a measure of bank riskiness. In our baseline specification, we measure riskiness as the ratio of impaired loans to total assets. SRF$_{b,t}$ denotes the SRF of bank $b$ at the end of quarter $t$, IDR$_{b,t}$ indicates the long-term IDR of bank $b$ at the end of quarter $t$ and Assets$_{b,t}$ is the natural logarithm of total assets in U.S. dollars, normalized using the consumer price index.$^{11}$ OtherRisk$_{b,t}$ is a vector of our remaining risk measures as bank controls. In the baseline specification, this vector includes net charge-offs/total assets, RoA (net income/total assets), tier-1 capital/total assets and trading assets/total assets. $\varepsilon_{b,t}$ is the error term. All specifications control for country-fixed effects $Z_{t}$ and quarter-year fixed effects $X_{t}$. We also consider specifications in which we control for bank-fixed effects instead of country-fixed effects. We refer to this alternative specification as Model 2. The standard errors are robust and adjusted to control for clustering at the bank level.

Table 2: Summary statistics

Notes: the table presents summary statistics on total assets and our risk variable ratios by bins of government support. We rely on the following variables from Bankscope (series in parentheses): total assets (DATA2025), impaired loans (DATA2170), net charge-offs (DATA2150), net income (DATA2115), tier-1 capital (DATA2140) and trading assets (DATA29190). We normalize each risk measure by total assets, converted to 2012. U.S. dollars are presented in millions.

Source: Authors' calculations, based on data from Fitch Ratings and Bureau van Dijk's Bankscope.

---

11 We use 2012 dollars as the baseline. We pull the “All urban consumers, all items, not seasonally adjusted” series from Federal Reserve Economic Data.
Finally, since a bank’s creditworthiness will likely play a role in the effect of government support on its risk-taking activities, we also consider a version of our model that includes the interaction between the SRF and the long-term IDR, \( \varphi \times SRF_{at} \times IDR_{kt} \).

5. Results

5.1 Impaired loans

Impaired loans are those that are either in default or close to default. These loans are typically behind in payments or restructured from a previous loan. They constitute a good measure of the amount of bad debt currently in the loan portfolio of a bank. Regulatory agencies require banks to write down loans as impaired under specific delinquency criteria, which may vary by country. Typically, regulators classify loans that are delinquent for 90 days (one quarter) as impaired.

In our analysis, we use impaired loans (from Bankscope) as our baseline measure of a bank’s riskiness. The main hypothesis that we intend to test is that banks with higher government support engage in riskier (lending) activities. Specifically, if the level of government support affects bank preferences for risk, we would expect that banks with stronger SRFs would engage in riskier lending activity. This, in turn, implies that more loans would become delinquent, resulting in an increase in impaired loans in the following quarters.

Table 3 summarizes our results. It presents the value of the coefficient \( \beta \) on the SRF in our models of risk for different lags (1 to 11 quarters) of sovereign support. The top rows of panel A show the effect of government support on the level of impaired loans. The main finding is that stronger sovereign support is...
associated with an increase in the ratio of impaired loans to total assets. In the model that includes country-fixed effects but no bank-fixed effects (Model 1), this result is statistically significant at the 1% level and the effect is economically meaningful; each notch increase in the SRF increases the impaired loan ratio by just under 0.2, which is an approximately 8% increase for the average bank. The effect is persistent and roughly constant through the following 10 quarters.

Results are similar but weaker in the analysis that includes bank-fixed effects instead of country-fixed effects (Model 2). In particular, we find a statistically and economically significant effect of sovereign support on the proportion of a bank’s impaired loans approximately seven quarters ahead. The lack of within-bank variation in government support may drive this weakness, as suggested by the lower t-statistics.

Figure 5 presents the relevant coefficients for both models. The squares and triangles correspond, respectively, to the values and significance at the 10% level of the SRF coefficient through time. This graphing of our results illustrates the importance of timing after a change in the SRF. The grey line in Figure 5 shows that an increase in sovereign support leads to a rise in the ratio of impaired loans as early as a quarter after the change in support in the model with country-fixed effects. We also see that this result is persistent and statistically significant through the following 10 quarters. The yellow line presents the results of the specifications that control for bank-fixed effects (but no country-fixed effects). An increase in government support to a bank also leads to a higher impaired loan ratio, but the effect is only significant seven quarters after the change.

5.2 Net charge-offs
For robustness, we also look at alternative measures of a bank’s riskiness. Net charge-offs are often used as a proxy for bank risk because they tend to increase with riskier lending activities. They are defined as the difference between charge-offs and recoveries, where charge-offs are debts that a bank declares
likely uncollectible and recoveries are collections on debts that a bank had previously written down as charge-offs. As with impaired loans, we scale net charge-offs by the total assets of the bank. Similar to our test based on impaired loans, if changes in sovereign support affect bank preferences for risk, then we expect that increases in SRFs would lead to riskier lending activity, resulting in an increase in net charge-offs.

The second set of rows in panel A of Table 3 presents the results of the analysis where the dependent variable is net charge-offs, with country-fixed (Model 1) and bank-fixed (Model 2) effects. Our findings support and complement our previous result that stronger sovereign support is associated with an increase in riskier lending activity. When we control for bank-fixed effects (Model 2), we find that the effect is statistically and economically meaningful, comprising a change in net charge-offs of approximately 0.04 per SRF notch, or 7% of an average bank’s net charge-off level. Figure 6 shows these results. The coefficients on sovereign support are positive, but not statistically significant in the model with country-fixed effects.

The dynamics and timing of debt charge-offs are complex. On the one hand, there is guidance from governments and regulators to encourage early charge-offs through tax exemptions and regulatory enforcement. On the other hand, banks still retain some discretion and may prefer to delay charging off debt within the timing established by the regulatory guidelines. Consistent with this pattern in the timing of charge-offs, we find that the effect is strongly significant for the two quarters following a change in support; it becomes weaker for the third to sixth quarters and then strongly significant after seven quarters.

Table 4: Impaired loan response, interaction
Notes: the table presents results on the relationship between government support, credit quality and impaired loans. We report the value of the estimated coefficient on the SRF, IDR and their interaction for different lags (1 to 11 quarters). Model 1 in panel A corresponds to the analysis with country-fixed effects and without bank-fixed effects. Model 2 in panel B includes bank-fixed effects, but no country-fixed effects. Each specification uses robust standard errors clustered by bank. a, b and c signify statistical significance at the 1%, 5% and 10% levels, respectively.
Source: Authors’ calculations, based on data from Fitch Ratings and Bureau van Dijk’s Bankscope.
Do “too-big-to-fail” banks take on more risk?

The results that we have reported thus far suggest that government support influences bank preference for risk. Given that finding, a natural question to ask is whether the link between government support and bank risk-taking varies with a bank’s creditworthiness. We are particularly interested in finding out whether the link is stronger for riskier banks because, all else equal, we would expect these banks to be more prone to taking on additional risks. To test this hypothesis, we extend our impaired-loans regression analysis and include a term for the interaction of the SRF and the issuer default rating. The size of the interaction captures the marginal effect of government support for safer banks relative to risky banks. As before, we estimate two models: one with country-fixed effects, Model 1, and the other with bank-fixed effects, Model 2. We include the same controls for bank size and risk, that is, (the natural logarithm of) assets and our remaining risk ratios [net charge-offs/total assets, RoA (net income/total assets), tier-1 capital/total assets and trading assets/total assets]. In each model, we estimate the different specifications for 1- to 11-quarter lags.

Table 4 summarizes our results. Our main variables of interest are SRF and SRF * IDR. For completeness, we also present the coefficient on the IDR. Panel A shows Model 1, which includes country-fixed effects, while panel B presents Model 2, which includes bank-fixed effects. Each column indicates a different quarter lag specification. Figure 7 illustrates the timing of the SRF and SRF * IDR coefficients in the left and right panels, respectively.

5.3 Does government support have a bigger effect on riskier banks?

Looking across the 11 specifications in Model 1, each with a different lag, we find a persistent, statistically significant relationship for all three coefficients. As before, the level of impaired loans in a bank loan portfolio increases directly with the level of government support. Reflecting the timing of impairment, this effect increases with higher lags. Similarly, the interaction of the SRF and the IDR grows increasingly negative and significant, indicating that riskier banks are more likely to take advantage of potential sovereign support. In other words, though all banks increase impaired loans proportionately to their SRF, riskier banks do so even more. For each one-notch level of the IDR, a one-notch change in the SRF increases
the impaired loan ratio by approximately 2% for the average bank. When we control for bank-fixed effects in Model 2, the interaction effect is still present, but it is significant only if we examine lags four to seven.

6. Robustness

6.1 Other measures of risk

For completeness of our analysis, we consider three additional measures of bank risk: the tier-1 capital ratio (tier-1 capital/total assets), RoA (net income/total assets) and trading assets (trading assets/total assets). The traditional rationale behind capital requirements is that capital acts as a buffer for protection against unexpected losses. In that sense, a higher capital ratio implies a safer bank. However, capital can also act as a measure of bank risk: the amount of capital a bank needs for protection against losses is closely related to the risk profile of the bank that will ultimately lead to those losses. From this perspective, a higher capital ratio is indicative of a riskier bank because of the requirement of a higher buffer against losses. RoA captures the profitability of a bank’s assets. Banks with higher RoA typically have riskier asset portfolios and, as such, RoA can be considered a proxy for the risk preference of a bank. In a related spirit, trading assets can also act as an indirect measure of bank risk. Trading assets are securities that banks hold for reselling at a profit (as opposed to investment purposes). As a result, we could expect that banks with a higher ratio of trading assets to total assets would engage in riskier activities. We do not discuss composite measures of bank risk, such as z-scores, because of data availability limitations.

As shown in panel B of Table 3, banks with stronger government support have a higher tier-1 capital ratio, RoA and trading asset ratio in the specifications with country-fixed effects. The effect is statistically significant only for the tier-1 capital ratio. As an additional robustness test to this interesting result, we consider an alternative definition of the capital ratio, calculated as the ratio of tier-1 capital to risk-weighted assets. This analysis takes into account the riskiness of bank asset portfolios. Results are similar (statistically significant at 5% level in the model with country-fixed effects) and consistent with the second interpretation of bank capital, in which riskier banks hold higher capital.12

6.2 Domestic banks

In our analysis, we derive all of our results with country-fixed effects (Model 1) or bank-fixed effects (Model 2). Nonetheless, one may still worry about the large diversity of countries included in our sample. To address this concern, we limit our sample to include only banks headquartered in the U.S., which is the country with the largest number of banks in the sample. We are interested in understanding if the relationship between sovereign support and risk-taking documented in sections 5.1–5.3 is also present in the U.S. Table 5 summarizes our main results.

We see in panel A of Table 5, consistent with our previous findings, that an increase in government support leads to a higher ratio of impaired loans and to higher net charge-offs. Similar to our results for the global sample, the effect on impaired loans is stronger for riskier banks, reflecting the fact that they are more likely to exploit potential sovereign support by engaging in even riskier activities than their safer counterparts do (panel B of Table 5).

6.3 Alternative hypothesis

In this paper, we find evidence that suggests that banks with stronger sovereign support engage in riskier lending activities, which translate into a higher ratio of impaired loans. One alternative hypothesis could be that financial conditions were already deteriorating, which would lead to a higher ratio of impaired loans. Although we cannot completely rule out this premise, all of our specifications control for bank credit quality. Specifically, as shown in section 4, we control for the long-term IDR of each bank at the end of each quarter to take into account variation in bank financial strength.

In addition, our results in Table 3 and Figure 5 show that the effect becomes stronger, rather than weaker, over time (that is, the value of the coefficient on government support is increasing with the number of lags). This finding is inconsistent with a story in which the deterioration was already taking place and the change in sovereign support is a response to worsening conditions.

Also inconsistent with the alternative hypothesis are our findings on the tier-1 capital ratio. If stronger government support was the response to a bank’s weaker conditions, we would expect the tier-1 capital ratio to decrease rather than increase (panel B of Table 3).

12 Analysis not included, available upon request from the authors.
As an additional robustness test, we also consider a variation of our sample in which we exclude banks that experience a simultaneous (within quarter) change in both sovereign support and credit quality. The idea behind this analysis is to consider a sample without potential contamination of the identification. After dropping such banks from our sample (23% of SRF changes), we find qualitatively similar results. Overall, all these findings support our initial hypothesis that banks with stronger government support take on more risk.

7. Final remarks

This study offers new and relevant evidence on a long-debated question: does the too-big-to-fail status increase bank risk-taking incentives? Our evidence is novel because it focuses on Fitch’s new SRFs, which aim at isolating the likelihood of governmental support from other sources of external support. Of course, SRFs only reflect Fitch’s views on potential government support and of the government’s ability to support a bank. As is the case in all studies based on ratings, our results hinge on this assessment’s reliability. The key advantage of our approach is that SRFs only include (Fitch’s views on) sovereign support, and exclude parent corporations’ support.

Our findings are also innovative in that we focus on impaired loans to measure bank risk-taking incentives. This analysis is important because impaired loans, in contrast to other, more general measures of risk, are more directly under bank control. Our results account for the governmental interventions during the financial crisis, but do not reflect the long-term effects that may arise from the regulatory changes introduced in its aftermath. An interesting area for future research would be to investigate to what extent the new regulations, in particular those dealing with the too-big-to-fail banks, affect the link we unveiled between the likelihood of governmental support and bank risk-taking policies.
References


Li, Z., S. Qu, and J. Zhang, 2011. “Quantifying the value of implicit government guarantees for large financial institutions,” Moody’s Analytics Quantitative Research Group, January


Morgan, D., and K. Stiroh, 2005. “Too big to fail after all these years,” Federal Reserve Bank of New York Staff Reports, no. 220, September


An overview of the risk-neutral valuation of bank loans

Danilo Tilloca
Vice President, Group Special Entities Risk and Evaluation, Unicredit Holding SpA

Luciano Tuzzi
Senior Vice President, Head of Group Special Entities Risk and Evaluation, Unicredit Holding SpA

Abstract
This paper provides an overview of a new methodology that allows banks to evaluate loans using the risk-neutral approach. In specifically, it illustrates the methodological framework behind the definition of the risk-neutral default probabilities used to estimate the loans credit spreads. These risk-neutral probabilities are calculated using a contingent-claims approach conceptually similar to the Black-Scholes and Merton framework for modeling corporate liabilities. The proposed risk-neutral approach is suitable for producing estimates, in a fair value computation context, that are as close as possible to the “exit price,” as mandated by IFRS 13, with a lower dependency on internal parameters.
1. Theoretical framework

1.1 Risk-neutral pricing overview
The introduction of IFRS 13, at the beginning of 2013, required, inter alia, that fair value measurements made for disclosure purposes should maximize observable market data by minimizing the usage of unobservable and entity specific inputs. Given these requirements, a new approach is proposed in this article, based on the contingent-claims framework for the pricing of bank loans. The problem of evaluating risky debt using contingent-claims approaches has been addressed by many researchers in literature, as a result of which we have had a proliferation of complex models, most of which have paid little attention to the practical applicability of the models [Bohn (2000)]. The proposed approach tries to balance between simplicity and usefulness by providing a tractable, but realistic, approach to loan fair value estimation. We hope that the model proposed here would contribute toward defining a new standard in the banking industry for IFRS 13 compliant valuation of loans.

Prior to the introduction of IFRS 13, bank loans were usually evaluated using the real-world internal default probabilities and a certain number of other internal parameters, such as the bank’s cost of equity, credit VaR, administrative costs and internal transfer rate. After the introduction of IFRS 13, banks were required to replace the old internally focused approach with a new model that used parameters linked to observable market inputs, moving to the so-called risk-neutral approach.

The problem, of course, is that very few bank loan transactions are observable on the market. Consequently, even if the discount methodology used to evaluate the loans is similar to the one used for pricing the debentures, it is not usually possible to apply the same spreads, making it necessary to find a link between observed market spreads and bank loans (even if the latter are not traded). To solve this problem, the proposed methodology assumes that bonds and loans have the same market risk premia in common. This assumption, which has its foundation in the Capital Asset Pricing Model (CAPM) theory, allows us to transform, through a correlation factor, the market risk premium into a loan-specific risk premium. In effect, if one accepts that the observed market spreads incorporate the market risk premium, then it is reasonable to assume that the loans prices, if quoted on the market, should incorporate, in some way, the same premium. From an operational standing point, it is necessary to take into account the differences and specificities that arise when comparing observed market spreads with specific loan characteristic and debtor creditworthiness.

The proposed model addresses this issue by using the contingent-claims model for the evaluation of risky debt, based on the pioneering work of Black, Sholes and Merton (BSM). Using this approach, we find that there is a link between the real-world default probabilities, which is used to measure the creditworthiness of a debtor, and the risk-neutral one (observed and embedded in the market spreads), and that this link is a function of the market risk premium, also called “the Sharpe ratio.” The proposed methodology bases its estimates on the calibration of the Sharpe ratio using quotes observed in the most advantageous credit market. The estimated fair value calculated in this way aims to determine the exit price in these markets.

The market price of risk estimated in this way reflects all the factors that a market participant would consider when pricing the asset or liability, such as the default probability, recovery rate and funding costs.

In this paper, it is assumed that, if the loans evaluated using the proposed approach were listed on an active market, they would be subject to the same market price of risk as the ones observed on the public markets. In other words, there is no need for additional illiquidity premiums, just as IFRS 13 had mandated.

1.2 The relationship between real-world probabilities and risk-neutral probabilities
We use the Merton (1974) approach to define the relationship between the real-world default probabilities (PD) and the risk-neutral ones. Following Black and Cox (1976), we assume as given the exogenous insolvency trigger; when this trigger is reached, all the outstanding debt is assumed to default and the investor will recover a given amount “R”. Under these hypotheses, it is possible show that the cumulative risk-neutral default probability CDPN is linked to the real-world CDPm through the Sharpe ratio (λ), which measures the excess return per unit

---

1 Loans and receivables with banks and customers, debt securities held to maturity, liabilities due to banks and customers and issued bonds.

2 Given that there wasn’t any industry-wide accepted methodology, the internal parameters differed from one bank to another.
of risk. Since the Sharpe ratio of the asset under evaluation will usually differ from the Sharpe ratio of the market, we have to use the CAPM model to express the specific λ\textsubscript{i} of the asset under evaluation, in terms of the market risk premium (λ\textsubscript{m}). The asset-specific risk premium is the market-specific risk premium adjusted by the correlation between the asset and the market. In particular, the calibration of risk-neutral PDs is conceived within a sector-specific context that aims to estimate sector-specific market parameters. Under this approach, λ\textsubscript{m} is substitute by λ\textsubscript{s}, and the asset-specific Sharpe ratio is given by the following relationship: λ\textsubscript{i}=ρ\textsubscript{i,s}·λ\textsubscript{s}.

1.3 Default probability estimation from market quotes
1.3.1 Estimation from credit default swaps
The cumulative PDs implied by the market can be derived from the CDS market quotes.\textsuperscript{3} CDSs are contracts that give the right and the obligation to be compensated for a loss, given the default of a reference entity. The assurance premium for this kind of protection is paid in the form of a spread over the notional covered by the CDS contract. We can price the CDS using the difference between the fee-leg paid by the investor, which includes the premiums streams to compensate the protection seller, and the default-leg paid by the latter, which includes the payment of the loss in case of default of the reference entity. The estimation of the risk-neutral cumulative default probabilities can be performed using the full bootstrapping of the term structure of hazard rates.

1.3.2 Estimation from risky bonds
Using the prices of risky zero-coupon bonds, it is possible to estimate the implied market default probabilities."\textsuperscript{4} The market convention is to incorporate all the expectations regarding the default and recovery on the spread over the risk-free rate. Assuming that the expected loan recovery rate R is known, we can write the fair value formula in order to incorporate explicitly the cumulative default probability CDP(t), hence removing the need for using the credit spread. In effect, we know that, if the bond will not default, the investor will receive at the maturity t, 1 with probability S(t)=1-CDP(t), while in case of default the investor will receive the expected recovery R with marginal default probability δS(τ)=S(τ-1)-S(τ) for every τ ≤ t (since it is unknown when the default will occur).

This relationship can be extended to coupon-bearing bonds, since in this case, the fair value can be expressed as a strip of zero coupons. Consequently, the bond-based approach applied to market PD estimation consists of the following two steps:

1. Spread computation: issuer-specific spread is estimated using the bond return and the swap curve of reference for the issuer selected.
2. Market CDP estimation: the CDP at time t implied by the market is estimated through a bootstrapping procedure.

With reference to loss given default (LGD), the issuer-specific parameter is typically estimated by the bank; however, if that is not available it is assumed to equal 60% (CDS market convention).

1.4 Relationship between CDS quotes and bonds spreads
The proposed approach uses CDS quote spreads to estimate the credit risk premia, though it is also possible to estimate this premia using the credit spread paid on bonds.\textsuperscript{5} In the market, the difference between these spreads is referred to as CDS-bond basis and has been the subject of several studies. For example, Hull et al. (2005) demonstrated that, if the LIBOR curve is used as the risk-free curve when estimating the bond spread, the two spreads are almost equivalent. The reason for this equivalence is explained in Figure 1. The figure illustrates how, for investors that funds themselves at LIBOR, a combined position of buying protection in a CDS and entering into an asset swap in which the fixed-coupon payments of a bond that trades at par are swapped against a stream of floating rates, is fully hedged in any state of the world. Before maturity, there are two possibilities: no-default (left-hand side) and default (right-hand side, assuming physical delivery and unwinding of the IRS). In both cases, the combined position is credit risk-free.

Consequently, the CDS premium should match the asset swap spread. If the difference between the CDS premium and the asset swap spread were to diverge from zero, it would result in a risk-free curve when estimating the bond spread, the two spreads are almost equivalent. The reason for this equivalence is explained in Figure 1. The figure illustrates how, for investors that funds themselves at LIBOR, a combined position of buying protection in a CDS and entering into an asset swap in which the fixed-coupon payments of a bond that trades at par are swapped against a stream of floating rates, is fully hedged in any state of the world. Before maturity, there are two possibilities: no-default (left-hand side) and default (right-hand side, assuming physical delivery and unwinding of the IRS). In both cases, the combined position is credit risk-free.

\textsuperscript{3} To ensure that we avoid the impact of market-maker spread, which is not directly related to the creditworthiness of the entity, on our estimation of credit risk we will use the par-spread estimated from CDS mid-quotes.

\textsuperscript{4} We are following an approach similar to the one proposed by Fons (1994), which can be considered as one of the first reduced-form models.

\textsuperscript{5} From a practical point of view, we refer to the asset swap (ASW) spread since it is the simplest way of transforming a fixed-rate bond into a strip of floating cashflows plus spread.

\textsuperscript{6} We define this basis as: ParSpread\textsubscript{CDS} – Spread\textsubscript{ASW}
An overview of the risk-neutral valuation of bank loans

Theoretical arbitrage opportunity. If the CDS-bond basis becomes positive (negative), an arbitrageur will find it profitable to sell (to buy) the cash bond in an asset swap and to sell (to buy) CDS protection at the same time. In effect, the arbitrage is generally not perfect and there are a few technical reasons (difficulty in short-selling bonds, cheapest-to-deliver option) that tend to push the basis into the positive territory. In summary, the pre-crisis literature suggests that the arbitrage relationship between CDS and cash-bond spreads holds fairly well and that, if anything, the basis should be slightly positive. As a consequence of the 2007–2008 financial crisis, this basis has become persistently negative.

In order to exploit the negative basis, an arbitrageur would have to finance the purchase of the underlying bond and buy protection. During the credit crisis, because of funding liquidity shortage and increased risk in the financial institutions, closing this gap became more difficult, risky and expensive. During the financial crisis, the negative basis was stable and persistent; in effect, the basis trading was facing liquidity and counterparty risk, and hence was not risk-free.7

We conclude by stating that, in normal market conditions, we can consider the CDS quotes and bond spreads interchangeable for the estimation of credit risk premia. However, since during stress times, asset prices depart from frictionless ideals (due to funding liquidity risk faced by the financial institutions and investors), the CDS quotes should be preferred in order to avoid embedding in the estimation factors that are not strictly related to the credit risk.8

2. Calibration methodology

As suggested, the definition of risk-neutral cumulative default probabilities lies on the correct and consistent calibration of the market parameters. To this extent, the present calibration methodology illustrates the procedure to consistently estimate the market-specific risk premium ($\lambda$). In specifically, the calibration methodology uses historical cumulative PDs, issuer-specific correlations ($\rho$) and market PDs as input data to ultimately estimate market-specific risk premium ($\lambda$). By minimizing the estimation error on the market quotes, the proposed methodology maximizes the market information embedded in these values. The proposed calibration methodology also provides stable estimates, since the final values are obtained using an average of the last

---

7 The analysis of this basis and of 2007/08 financial crisis is beyond the scope of this paper. More details and a complete bibliography can be found in Augustin (2012).

8 In situations where there is a persistent (not temporary) negative CDS-bond basis, it might be necessary to consider it as an additional liquidity premium for the funded loans market and incorporate it in the loans discount spread.
250 estimates and hence is able to absorb the daily volatility of quotations.

2.1 Sample portfolio
It is necessary to define a sample portfolio in order to estimate the market risk parameter \( \lambda_i \) in a sector- and region-specific context. This means that, for every country where this methodology is applied, it will be necessary to estimate the country-specific lambdas sectors.

2.2 Input data
2.2.1 Market PDs
As described in section 1.3, the market cumulative PDs can be derived from the CDS market quotes or from the bonds prices observed on the market.

2.2.2 Internal PDs
The issuer-specific internal cumulative probabilities of default are the physical default probabilities usually adopted by banks to perform internal evaluation of loans. It is important to stress the concept that, due to the way we calibrate and transform the real-world probabilities, the proposed approach is robust to misspecification of this parameter. Berg and Kaserer (2009) demonstrated that the market risk premia estimation is particularly robust when applied to the five-year bucket.9

2.2.3 Rho
The correlation parameter can be defined on either a specific-issuer or sector-specific basis. In the CAPM, this parameter represents the correlation between the issuer’s assets and market returns. Instead of the CAPM, a more sophisticated factor model can be used to determine the sensitivity of the default probability of the issuer to certain risk factors. In particular, it is possible to model the single issuer creditworthiness as a function of both the cluster the issuer belongs to and obligor-specific credit risk.

2.3 Market price of risk calibration
As suggested, market-specific risk premium represents the expected excess return demanded by investors per unit of risk. This parameter is estimated through a calibration where \( \lambda_i \) is calculated as the average of daily \( \lambda(t) \) estimations. According to this approach, \( \lambda_i \) is calculated through the following process:

- Estimation of the daily \( (t) \) individual market CDP\(_i\) for every issuer \( i \) in the sample.
- Estimation of daily individual expected CDP for the issuer \( i \).
- \( \lambda_i \) is, ultimately, calculated as the arithmetic average of the daily observed \( \lambda(t) \).

In order to make the estimate as specific as possible, a different parameter is estimated for every sector – financial, large corporate and government. The estimates are run over a time horizon suitable at guaranteeing a consistent calibration in terms of both asset liquidity and number of observations.10

3. Formula behavior
In this section, we will provide more information about the “fair value” loan formula and, in particular, on its behavior and applicability in the credit market. As the source for the real-world default probabilities, we will use “Global corporate average cumulative default rates” (CDR), published by Standard & Poor’s [Vazza et al. (2012)], and for historical CDS quotes, we will use the Markit sector curves composite, calculated using the methodology applied by Markit (2012).

3.1 The fair value loan formula
It is possible to extend the pricing of risky zero-coupon bonds to loans pricing by substituting the bond coupons with the loan cashflows. However, it is important to note that the LGD for loans is calculated only on the outstanding debt; hence, in the case of default, it is assumed that the recovery of remaining interest coupons is zero. This implies that, for pricing purposes, the loan cashflow needs to be split into two components: the repaid notional and the paid interests.

3.2 Negative risk-neutral PDs
Theoretically, it is possible that, during calibration, we estimate negative market risk premiums. This fact could pose problems during the transformation from real-world PDs to risk-neutral PDs, since we could obtain risk-neutral PDs that decrease over the time. We know that this is not possible in the real world, and we expect that this would also be the case in a risk-neutral world.

---

9 The authors showed that, in the CDS market, the five-year maturity is the most liquid and that trades during periods of turmoil tend to concentrate to this maturity.

10 A time horizon that could be used as reference is around 250 observations.
From a theoretical point of view, the risk-neutral world is one wherein the expected return required by all investors on all investments is the risk-free interest rate. It is then theoretically possible that, if the investors’ assumptions on the riskiness are wrong and too optimistic, to move from the real world to this world, the market risk premium could be constrained to become negative.

Looking at historical CDS quotes, we have observed that, in certain circumstances, the market risk premium is negative. We have observed this behavior from the CDS quotes related to issues rated B from the period January 2013 to December 2013 (see the left-hand panel of Figure 2). To determine the accuracy of our findings, we also tested our hypothesis using the spread of Bloomberg’s composite of U.S. issues rated B (see the right-hand panel of Figure 2). Both analyses provide evidence in support of the proposition that B-rated short-term debt (years 1–3) has an implied negative market risk premium.

Since the information regarding the historical defaults, for this class of loans, is publicly available to all the investors, the question that arises is why investors ignore this information. Our conclusion is that the observed behavior is probably due to the fact that, in a low interest rate regime, investors who are looking for higher yields are willing to invest in short-term sub-grade loans. This has the effect of lowering the market-implied default probability. Naturally, over the long term, the market risk premium turns into the positive territory. This example provides support for the proposition that real-world probabilities are different from the risk-neutral probabilities, since the latter are an expression of the forces driving the markets and could show implied default probabilities that could never be observed in the real world.

Our opinion is that, even if a negative market risk premium could theoretically be observed, this event can practically happen very rarely, and only in certain conditions (such as very low risk-free returns and short-term low investment grade debt). To understand the reasoning, it is important to note that the calibration process for the lambda estimation is performed only on market quotations of observable obligations quotes or CDS par spreads. This implies that all the reference entities used for the calibration should have an agency rating. Since the internal models used by banks, for the estimation of default probability, is based on the shadow rating approach, the internal estimate of the PD should be near the agency one. Given that this information is publicly available, the market quotations should always imply a market PD higher than the internal PD estimate, since it should at least incorporate a premium for liquidity and recovery expectations [Turnbull (2003)]. In summary, an estimation...
of negative lambda should never happen, and if it does, it is probably a sign that the internal PD has been wrongly assigned. Before accepting it, the assigned rating should be verified. The positivity of the marginal PD should always be checked and confirmed to mitigate the risk of incorrectly assigning internal PDs. In case the credit spread of the loan counterpart can be inferred directly using market quotes of liquid instruments, it is possible to avoid the use of lambda and directly apply the default probabilities implied by the market spread.

3.3 Constancy of the sector lambda

The model assumes that we can keep constant the sector lambda $\lambda_s$ for each counterpart belonging to a given sector. This assumption is a direct consequence of the CAPM, which indicates that the market risk premium is constant and equal for all the market participants. This means that the $\lambda$ estimated from an AA counterpart can be applied to BBB or AAA counterparts. To verify this hypothesis, we have calculated the lambda from Markit generic sector curves to see if they are interchangeable. The CDPW has been taken from Vazza et al. (2013), while the term of the CDS quotes considered was five years ($T=5$). The results are shown in Figure 3.

We can observe that the estimated $\lambda_s$ is not constant over the time, but that the differences are not very large, with the exception of those with a low rating (below BB rating). The results corroborate those of Berg (2010), who found that the term structure of the risk premia was flat before the 2007–08 crisis, and subsequently sloped downward during the crisis. Sadly, since that paper was published in 2010, it does not contain information about the period after the credit crisis. Our analysis, which covers a period of one to two years, shows that the term structure of risk premia is again becoming flat.

Since the $\lambda_s$ used in the model is an average of 250 daily estimates, it is useful to analyze the behavior of the model over this period. We have found that these averages behave in a similar fashion to the daily observations but are smoother and allow us to better highlight market trends (see right panel of Figure 3). All estimates fall following the reduction in market risk premia requested by investors, which started in mid-2012. But even in this case, the lambda estimated from the CDS quotes related to the low-rating class appear to be consistently lower than the other ratings.

CAPM indicates that the market risk premium is constant, not only for all the market participants, but also over time. We investigated whether this is a realistic assumption. For this test,
we have used the same data as before, but this time we have estimated the term-structure of the lambda, where $\lambda$ has been calculated for the available CDS maturity.\footnote{For $T = 1, 2, 3, 4, 5, 7$ and $10$.} We used averages in order to smooth out daily volatilities and because the $\lambda$ used in the model is an average of $N$ estimations. The results of this test are shown Figure 4.

We find that the model hypothesis that lambda is constant over time is a good approximation for what happens in the market. We observed only few outliers, the more significant one is the AAA rating class at one year, the observed difference of which is mainly due to the fact that the real-world PD is so low that even CDS quotes with low price imply high lambdas.

To conclude this section, we want to stress that the default probabilities implied in the bond prices and CDS spreads (risk-neutral PDs) should always be higher than the real-world PDs. This fact was first highlighted by Altman (1989), who described the discrepancy between bond prices and historical default data. Elton et al. (2001) provided an in-depth analysis of the spread between rates on corporate and government bonds. The main reasons for such differences are the following:

1. The investors raise the expected returns they require on corporate bonds to compensate for their relatively low liquidity.
2. Bond traders are allowing, in their pricing, for the possibility of depressed scenarios much worse than anything seen during the time period covered by historical data. This occurs because a large part of the risk on corporate bonds is systematic rather than diversifiable.

Hence, we cannot expect risk-neutral PDs to converge toward internal PDs. In addition, the fact that this difference grows with the square root of time is a consequence of the assumption that the assets that could trigger a default, follow a log-normal geometric Brownian motion. This kind of assumption is used in many credit models, such as KMV [Agrawal et al. (2004)], and has been recently proposed by Berg and Kaserer (2009).

4. Conclusion
To mark credit assets using a valuation model, it is necessary to estimate the risk-neutral default probabilities from the quotes observable in the credit market (such as bond prices and credit default swap spreads). Currently, two approaches to mark a credit asset to model are proposed in the literature: 1) the reduced-form approach, where it is assumed that the price of any security is the expected value of its future cashflows and 2) the structural
approach, where the default occurs when the value of a firm’s assets declines below a given threshold. We believe that the two approaches are not mutually exclusive, but are complementary. In this article, the two approaches were combined in order to allow for the evaluation of a great variety and type of commercial loans. Both internal and external data are used for this purpose. The first approach has been used to build the discount methodology, while the second approach allowed us to extend the methodology, to those loans whose prices are not observable in the market.

We summarize the steps proposed in this paper for calculating the credit spread for loans pricing as follows (Figure 5):

1. Estimate of the market risk premia using the observed credit-market quotes.
2. Using the right correlation factor, calculate the specific risk premia of the loan under evaluation.
3. Knowing the physical default probability and the specific risk premia of the loan, it is possible to calculate the proper risk-neutral default probability.
4. Using a non-arbitrage approach that takes into account the loan collateral, it is possible to transform the risk-neutral PD into a proper credit spread suitable to be applied to discount the loan cashflows.

It is important to note that we have substituted most of the internal parameters with information obtained from the credit markets, and only left the LGD, correlation factor and real-world default probability to be derived from internal information. It is also important to highlight the fact that, due to the way in which the calibration and transformation of the real-world probabilities work, the proposed approach is robust to misspecification of this parameter.13

References

Vazza, D., N. Kraemer, and N. Richhariya, “2012 annual global corporate default study and rating transitions,” Standard & Poor’s Financial Services LLC

13 This fact was first highlighted by Berg and Kaserer (2009).
Smart beta: too good to be true?

Bruce I. Jacobs
Principal, Jacobs Levy Equity Management
Kenneth N. Levy
Principal, Jacobs Levy Equity Management

Abstract
Smart beta strategies promise to deliver market-beating returns with simplicity and low cost, but the reality is more complicated. Contrary to popular perception, smart beta strategies are neither passive nor well diversified. Nor can they be expected to perform consistently in all market environments. Perhaps most importantly, because of their focus on only a limited number of factors, smart beta strategies fail to exploit numerous potential profit opportunities.
Smart beta or alternative indexing – whatever its label (we will stick with smart beta) – is a relatively new investment approach that has attracted considerable attention and investment from pension funds and individuals. Its popularity is hardly surprising, as smart beta promises to deliver market-beating returns in a convenient, low-cost, easy-to-understand manner.

Smart beta promoters emphasize the simplicity of the strategy’s portfolio construction and trading rules. They often compare it with passive index investing, which delivers market returns at low cost and with high transparency. Yet the goal of smart beta is the same as that of active investing – to outperform the market. Unlike active strategies, however, smart beta eschews security research. Instead, it seeks to beat the market by replacing the security weighting scheme used by passive management (capitalization weighting) with a weighting scheme that emphasizes certain security characteristics, or factors – value, size and momentum, among them – that have performed well historically.

It sounds simple enough. However, the reality of smart beta is more complicated, and its promise of higher return with lower risk is less certain. Below, we debunk some common misconceptions associated with smart beta strategies.

**Smart beta portfolios are passive**

Smart beta is often compared with passive investing because, like index funds, it does not require the portfolio manager to forecast security returns and risks. It is essentially a rules-based approach, with preset criteria dictating the weighting of securities in the portfolio.

But a truly passive portfolio buys and holds the capitalization-weighted market; that is, the stocks are weighted according to the ratios of their market values (or capitalizations) to the total market value of all stocks in the index. It requires little trading because the portfolio and the benchmark index adjust simultaneously as security prices change. The result is a portfolio that delivers the underlying market’s return, along with the underlying market’s risk.

**Smart beta targets the most significant return-generating factors**

Smart beta equity portfolios in general target only one or a limited number of factors – value, small size, price momentum and/or low volatility. Smart beta providers would have you believe that these factors have the greatest impact on security returns. Some factors that have performed as well as, or better than, the chosen few are left off the smart beta menu.

In our own research, first published in 1988, we looked at 25 security characteristics, including most of the factors currently used in smart beta strategies (Jacobs and Levy 1988). We identified as statistically significant many more than the few factors pursued today by smart beta strategies. More recently, researchers have found dozens of factors to be significantly related to stock returns (Green et al. 2014). Interestingly, some popular smart beta factors, such as book-to-price, small size and price momentum, were not among the most significant. Portfolios restricted to the handful of factors targeted by smart beta are overlooking many potential opportunities.

**Smart beta portfolios are well diversified**

Most smart beta portfolios hold a large number of stocks, but numbers may not translate into diversification. Smart beta’s focus on a particular factor can lead to incidental bets and sector

---

1 Assets under management globally total U.S.$544 billion (see Evans (2015)).
2 Trading may be required to reinvest dividends and to adjust for corporate actions (such as mergers, acquisitions and spin-offs) and changes in index membership.
biases, which may introduce unintended risks. A focus on value, for example, can result in exposure to distressed firms. A focus on price momentum would have loaded up on the technology sector in 1999, prior to the tech wreck. More recently, low-volatility portfolios had a large bet against the financial sector at the market bottom in 2009, which contributed to their subsequent underperformance.

**Smart beta factors perform consistently**

Smart beta factors are selected and security weights determined based on historical data rather than on forecasts. These choices represent an expectation that targeted factors will continue to perform as they have in the past.

As economic or market conditions change, however, factor returns can vary significantly. It is well established that small stocks have periods of outperformance followed by periods of underperformance; the same holds true for value stocks. Price momentum, which performed well prior to the financial crisis, suffered in 2009, as the market reversed direction. A constant exposure to a factor regardless of underlying conditions leaves a portfolio vulnerable when that factor underperforms, as it inevitably will.

**Smart beta portfolios benefit from mean-reversion in prices**

One argument in favor of some smart beta strategies is that the systematic portfolio rebalancing required is a significant contributor to excess returns [see, e.g., Arnott et al. (2013) and Steward (2014)]. It has the effect of forcing sales of appreciated securities and purchases of securities that have declined in price. In theory, the portfolio will benefit as the prices of both types of securities revert to “normal” levels. However, empirical tests of smart beta exchange-traded funds (ETFs) find no consistent evidence of a mean-reversion benefit [Glushkov (2015)]. Furthermore, any active portfolio can choose to implement a rebalancing scheme that takes advantage of mean-reversion and, what’s more, can do so using proprietary trading rules less susceptible to front running than those of smart beta, and more responsive to changing market conditions.

**Smart beta portfolios can be efficiently combined**

Smart beta promoters often recommend investing in multiple factors to protect against the underperformance of any single factor. Value and momentum is one recommended combination. Returns to the momentum factor have a negative relationship (or correlation) with returns to value factors. Momentum strategies buy past winners and sell losers, whereas value strategies typically buy past losers and sell winners. When the momentum factor produced large losses in 2009, value factors such as book-to-price performed well.

Combining two separate smart beta portfolios can be problematic, however. There is no unambiguously correct method to determine the relative weightings of the two portfolios. Some of the holdings of the two portfolios may overlap, increasing security risk. Or the focus on different factors may lead to one portfolio buying a security even as the other is selling the same security, increasing transaction costs.

Some smart beta providers target multiple factors in a single portfolio. But this may complicate factor selection. For example, are value and momentum enough? What about small size? After the market trough in 2009, the small-size factor would have boosted the performance of a value-plus-momentum strategy. And this solution still fails to take advantage of the full range of return-generating factors, including those overlooked by smart beta strategies.

**Smart beta benefits from transparency**

Smart beta is typically more transparent than other active strategies. Investors know up front the factor(s) to be targeted, the frequency of rebalancing and the weighting scheme. Transparency can be beneficial for investors, enhancing their understanding of the strategy and allowing them to better gauge investment performance. However, transparency also has inherent disadvantages that can prove costly. In particular, the generic nature of smart beta factors, combined with preset rebalancing rules, can render such strategies vulnerable to both front running and factor crowding.

Front running occurs when others can anticipate the rebalancing needs of smart beta portfolios and buy stocks before they are

3 A recent analysis of the performance of numerous smart beta ETFs indicates that the negative effects of unintended exposures offset in part or in full any return advantages provided by desired factor exposures [Glushkov (2015)].

4 For a discussion of using the full range of return-generating factors, see Jacobs and Levy (2014b).
added to those portfolios or sell stocks before they are dropped.\(^5\)
This type of anticipatory trading can run up the prices of securities before they are purchased by smart beta portfolios and push down the prices of securities before they are sold, eroding portfolio performance.\(^6\)

Factor crowding occurs when large numbers of investors buy or sell the same securities on the basis of similar factors. This can lead to factor overvaluation and factor crashes, just as too many investors chasing any asset can lead to overvaluation followed by abrupt reversals. In the market turmoil of August 2007, for example, some quantitative hedge funds were forced by margin calls to liquidate holdings; they sold off stocks associated with commonly used factors, causing losses for other managers holding the same stocks (Khandani and Lo (2007)).

**Smart beta has nearly unlimited capacity**

Some smart beta promoters assert that, because smart beta represents a small portion of the equity market, there is more than enough capacity to handle growing assets in these strategies [Bell (2015, p.52)]. But this does not mean capacity is unlimited.

Everyone can hold the capitalization-weighted market index because it represents the entire stock market. This is not the case for smart beta strategies. For every smart beta investor who overweights a stock (relative to its market weight), there must be another investor who underweights it. As a factor outperforms over time, more investors will want to buy those securities associated with the factor and fewer will want to sell. That will drive up these securities’ prices and lower their future returns.\(^7\)

Many active managers try to protect the liquidity and profitability of their strategies by imposing limits on the amount of assets they manage.\(^8\) But smart beta factors are publicly available and product offerings are similar; there is no way to control the volume of investment in a factor. Even if one manager closes its strategies to new investors, other managers can continue to invest in that factor.

**Smart beta streamlines the investment decision process for investors**

One of the purported benefits of smart beta is that it streamlines the investment decision, not only for managers, but also for investors. Given its simplicity and transparency, it is said to be easier to implement and to require less due diligence than other active strategies (Hsu et al. (2012, p. 11)).

With other active strategies, however, portfolio managers shoulder the responsibility for determining what investment criteria to emphasize, and for deciding whether and when to alter them as conditions change. Smart beta shifts those decisions from the portfolio manager to the investor. With smart beta, it is incumbent on the investor to select the right smart beta factor, or factors, and to decide if, and when, to get into or out of a particular factor. In doing so, investors are taking on substantial investment responsibility.

**Smart beta costs less than active investing**

Smart beta is generally viewed as less costly than traditional active management, primarily because its management fees and portfolio turnover are usually lower. When evaluating smart beta, however, investors must consider the hidden costs. Smart beta portfolios may incur substantial opportunity costs. By failing to take into account all the factors that research has shown to be significantly related to security return, smart beta portfolios miss out on potentially rewarding opportunities. Also, by ignoring the changing relationships between factor returns and underlying economic and market conditions, they may end up exposed to risks without rewards.

Furthermore, smart beta’s use of preset rebalancing frequencies and generic factors opens the door to front running and factor crowding, which can increase transaction costs and reduce or even eliminate any value added from the factors targeted. Finally, smart beta imposes on the investor responsibilities for factor selection and timing that, if done properly, are likely to entail considerable research expenditure. These costs are not reflected in the fees of smart beta strategies.

---

5 It is well known that the annual rebalancing of the most prominent small-capitalization stock index is affected by front running (see Madhavan (2003)).
6 As assets in smart beta strategies increase, adverse price pressure can be expected to increase accordingly, leading to even larger profit opportunities for front-runners and more erosion of factor returns. Recent evidence has documented adverse price pressure on smart beta strategies that rebalance on the basis of the book-to-price and size factors (see Yost-Bramm (2014)).
7 Smart beta performance in recent years has been disappointing (see Barlyn (2015), Evans (2015) and Malkiel (2015)).
8 On the importance of setting capacity limits for a firm’s assets under management, see Perold and Salomon (1991).
Conclusion
Smart beta strategies may be a useful addition to the range of investment approaches available to investors, but they are not a magic formula for increasing returns while reducing risks. Investors would be better served by a more realistic consideration of the pros and cons of smart beta investing.

References
Evans, J., 2015, “Smart beta is no guarantee you will beat the market,” Financial Times, February 2
Glushkov, D., 2015, “How smart are ‘smart beta’ ETFs? Analysis of relative performance and factor timing,” Wharton Research Data Services, University of Pennsylvania, April
Regulatory herding versus democratic diversity: history and prospects

Nicholas Dorn
Associate Research Fellow, Institute of Advanced Legal Studies, University of London

Abstract
Regulatory convergence – within the E.U., across the Atlantic and internationally – is conventionally represented as not only benign but also as essential in crisis prevention. This paper articulates a different frame of reference: one in which regulators “crowd,” “herd” and sometimes merge, so mimicking and exacerbating financial market tendencies toward similarity and contagion, and drawing regulators and markets into the same vortex. The paper looks at some of the historical and contemporary circumstances in the U.K., wider E.U. and the U.S. that have given reign to these tendencies and also at some aspects of regulatory architecture and governance that reduce such tendencies. It mentions pre-crisis tendencies to regulatory subservience to financial markets, with such subservience having a deep history in the U.K. and a shorter one in the U.S.; so-called command regulation, which has the potential to either deepen subservience or transcend it; and the institutional preconditions for permanent regulatory vigilance, such as democratic appointment of heads of agencies. The paper concludes by pondering the prospects for the democratic direction of financial market regulation, in terms of its distributional logics and extraterritoriality.
1. Introduction

A financial trade is said to be crowded when many market participants are trying to transact similarly—for example, to get into the euro, out of bank equity or into bonds. Traders may, in principle, hold to a variety of trading purposes, strategies and intentions; however, if they have come to rely upon the same, linked or similar information sources, analytical approaches, calculative tools and data pools, then they start to make similar decisions (in parallel rather than in consort). Crowding is widely considered to cause difficulties, since not everyone can move in the same direction without causing what eventually becomes recognized as bubbly and sometimes chaotic turning points. Economists, sociologists and ethnographers have various explanations for market crowding—such as information spillovers [Acharya and Yorulmazer (2006)], the performativity of calculative tools [MacKenzie (2003)] and common constraints on investment strategies [Pitluck (2014)]. Most commentators agree that crowding is not essentially a willful copying of others (that is herding, to be discussed below). Indeed, some crowders may hate that they feel obliged to do so, following common information sources and/or logics that seem to restrict their options.

In herding, by contrast, market participants consciously, willingly and strategically emulate the broad trading patterns of those around them. Needless to say that despite the fact that they try to differentiate and spice up their market offerings—through corporate and product branding, complexification of derivatives, combinations of strategies and routing through multiple venues—the underlying aims and tactics remain similar. This deliberate and “sophisticated” herding goes beyond reliance on the information environment; it results from the explicit “me too [but better]” management claims, occupational cultures, product development and sales pitches.

Do crowding and herding also occur in financial market regulation? Yes, they do. Crowding comes to the fore when regulators become reliant on similar mindsets, information sources and regulator rationales and objectives. Such was the case in the long period of complacency prior to the crises of 2007 and onward [Engelen et al. (2011)]. It was not just that the regulatory mind-set models were wrong but rather that they were so similar; hence, allowing the same basic business model to proliferate so widely. More diversity would have been safer. There was a temporary disruption in the convergence in regulatory thinking, as the response to crisis precipitated some radical rethinking. “Post crisis” rethinking then quite rapidly recoalesced around the new notion of macroprudential regulation, which recognizes that problems arise not only with specific firms and trades but also systemically, in the interstices, connections and similarities. A similar recognition is needed in relation to regulation itself, this paper argues.

How solidified and homogeneous will macroprudential thinking turn out to be internationally? Empirically, the answer to that question is not clear. A range of answers might be suggested, depending on the perspective taken. Some commentators, taking a wide-ranging global perspective, suggest that crises have introduced greater heterogeneity into financial markets regulation, as many rising powers, such as the BRICS, deepen their critique of the failure of western, neoliberal, hitherto dominant approaches [Helleiner and Pagliari (2011)] and encourage construction of pivots for regulatory and interregional regulation [Leahy and Harding (2014)].

Other commentators, more focused on the U.S. and/or the E.U., have noted the rapidity with which regulators picked themselves up from the epistemic floor and have rebuilt and extended their organizations, networking and influence [Blyth (2013)]. Even so, European regulators in particular seem now to be in a more explicitly close relationship with ministers and heads of state than hitherto. That may sometimes push regulators in the same direction (for example, when heads of states are in accord), while at other times allowing for some national regulatory leeway (for example, when countries’ national elites seek to retain influence over regulation of “their” banks, see Spendzharova (2014)).

Jurisdictionally specific work, to be deployed here, suggests a long-term historical continuity: separation of financial market regulation from democratic politics. This separation leaves regulators with each other as their primary reference points, with herding as one consequence. Of course, the market retains some of its influence, even after the continuing scandals over fraud, conflict of interest, benchmark manipulation and trading venues’ favoritism of some market participants over others. Such conduction seems to have been very widespread, possibly more the norm than the exception.
Given recent developments, we seem to be in a historical period in which the international (Western led) regulatory-market nexus is under more pressure than was the case before the events of 2007 onward, opening up some possibilities for non-trivial changes in governance of financial markets. Whether such opportunities can be grasped must depend, in part, on understanding the conditions that encourage regulatory crowding and herding.

2. Private regulation: aristocratic beginnings

In other policy areas, political parties choreograph and sometimes polarize questions of broad strategy, which, in relation to financial market policy, would involve asking questions like: what sort of banking do we want in coming decades in our country or region? Not so in relation to financial market regulation, and for good historical reasons. In exploring the historical issues, our principal scholarly debt must be to economic historians Cain and Hopkins for their account of the development of the City in relation to world trade and politics. Whilst the City’s side-stepping of the expanding sphere of democratic politics is not the main focus of Cain and Hopkins’ work, the latter helps us to contextualize (what became known as) financial market self-regulation.

With reference to the period from the mid-nineteenth century up until the end of the Second World War, Cain and Hopkins (1987) refer to an “aristo-financial elite” of “gentlemanly capitalists,” which preferred to invest internationally rather than to support domestic industry. The success of this footloose strategy made the City more dynamic than international rivals and more politically important domestically than industrial capital.

“The City’s overt political influence remained limited. But since City, government, and administration were so closely entwined, it is not surprising that many policy questions were regarded as being beyond the realm of party politics. It was assumed that matters of high finance could safely be left to the small circle of institutions which were thought to have an intuitive understanding of the ‘national interest.’ There is, of course, no denying that as landed wealth declined in importance, industrial as well as financial and commercial wealth grew rapidly. Nonetheless, the influence of industry on central government and on economic policy continued to be limited by its relative lack of access to the major sources of power and influence. Manufacturers were still largely outside the circle of gentlemanly culture and did not ‘speak the same language’ as the aristo-financial elite. And, insofar as they did gain political influence, any residual Cobdenite radicalism was likely to be muted by the need to join gentlemanly interests in defending property against the threats posed by trades unions and the spread of democracy.” [Cain and Hopkins (1987, 6)]

Thus, potential opposition to the aristo-financial elite was fragmented. “Militant manufacturers” – as Cain and Hopkins at one point describe British industrialists, who sometimes found themselves disadvantaged by the City’s investments in foreign industry – tried to push back politically; however, they were inhibited from taking an overtly oppositional position, because of their concerns over working class pressures, against which manufacturers made common cause with London and the shires. Meanwhile, working class political claims were also somewhat circumscribed by pride over Great Britain’s standing in the world and by political identification with the industrial, financial and political arrangements that underpinned this (international trade, sterling, Britain’s standing in the world, empire, etc.). An important consequence was that issues concerning the City and its governance were conceptualized as being “above politics” [Cain and Hopkins (2013, 340)] or, as one could also say, beyond politics. The City did not need to influence government, because financiers and political elites were interwoven.

The adoption of Cain and Hopkins’ work as a broad framework for understanding the history of financial market regulation might be thought too risky because, although their work has attracted praise for its scope, depth and originality, it also has been criticized. The grounds of criticism include that it is allegedly “excessively monocausal;” finance-centric (paying insufficient attention to other sectors and their strategies); Eurocentric (paying insufficient attention to “autonomous impulses emanating from the periphery”); neglectful of cultural and religious aspects; and in some constructivist critiques, fails to recognize that “the Empire was always an imaginative construct” [see Cannadine (1995, 194)]. It seems to the present writer, however, that a “gentlemanly capitalist” framing of finance usefully captures cultural as well as economic aspects of class alliances and antagonisms – both internationally and within Britain. The international aspects are by no means restricted to the Empire, since much of the City’s investment historically went much wider. These authors’ placing of the City in terms both of
foreign policy and domestic policy is valuable and some of the potential implications for us today will be taken up later.

In short, Cain and Hopkins’ work, alongside that of other economic historians, such as Daunton (1989), and political scholars such as Michael Moran (1988), usefully illuminates the historical, institutional and cultural setting within which financial market regulation arose in club/private form, outside the party political sphere. Indeed it has become common ground amongst social, economic and political historians of the City that, despite the advent of the general franchise in 1918, such private regulation in the City was hardly touched by democracy. As Johal et al. (2012) put it: “The Bank of England, which had hitherto been fairly marginal in the regulation of the City, now emerged as a critical institution. It used its authority to reshape the government of markets. The war had destroyed the kind of open international economy of which the City had been a centrepiece. After 1918 City markets were organised into a series of cartels policed by trade associations. The cartelisation of the markets, coupled with the authority of the Bank of England, was sufficient to sustain what the City called self regulation: accepting the disciplines of self regulation was the price firms paid for being allowed into the privileged cartels. The stability of the self regulatory system in the decades after 1918 allowed the City further to elaborate its regulatory ideology. This pictured the City as a special part of the economy, claiming exemption from one of the main features of 20th century economic government in Britain — the apparently inexorable rise of the state as a regulator of economic life” (ibid., 69).

Turning now to the period from the end of the Second World War up until the 1980s – a period in which Britain struggled in the face of recession, war and competition from the U.S. – Cain and Hopkins point to a development that is in no way prefigured by their analysis of earlier centuries. British finance found itself rehabilitated by the geopolitical response of the U.S. to communism. “In the Great War, Britain had needed American capital to secure victory: after 1939 her very survival depended on American aid. [...] The initial quid pro quo sought by the United States for her extensive aid was the abolition of imperial preference, the destruction of the sterling system, and – ultimately – decolonization. However, with the emergence of Russo-American antagonism after the [Second World] War, all these elements of Britain’s world power survived as valued assets in the Cold War, so much so that sterling was launched on a new international career in the 1950s as a junior partner of the dollar. Gentlemanly capitalists who had once provided the framework for the Pax Britannica now survived to fight another day under the protection of the Pax Americana.” [Cain and Hopkins (1987, 17)]

In these circumstances, fortunate for the U.K. and especially for the City, the latter not only survived but thrived, and its economic expansion helped to underpin its political autonomy, at least up until the 1980s. What remains somewhat controversial is to what extent that situation changed as a result of neoliberal reforms in the 1980s.

### 2. Mother of regulators: history of pseudo-public regulation in the U.K.

An important question is whether the U.K. policy reforms of the 1980s can be understood in terms of political disruption of the preceding historical arrangements of private or club regulation. The orthodox position – which dominated the academic literature as well as policy thinking before the financial crisis – concedes the history of club regulation, whilst claiming that modernizing reforms under both Conservative and Labour administrations displaced historical arrangements. According to the orthodox narrative, the opening up of the City by “Big Bang,” the creation first of a Securities Investment Board (SIB) and then of an “independent” regulator, the Financial Services Authority (FSA), swept away the old order. So says the conventional history; but on what grounds?

This question can be taken in stages, starting with the SIB, the forerunner to the FSA.

Writing in the 1980s about the SIB, Moran refers to the City having “lost the battle to keep the politicians and the civil servants at bay.” On the other hand, the SIB “is conventionally described as self regulation within a statutory framework” [Moran (1988, 22)]. And: “It [the SIB] is, in essence, a franchising operation: semi-private and private bodies (the SIB and the various self-regulatory organizations) will be awarded a franchise to exercise legally backed powers of regulation” (ibid., 24).

If the transition arrangements represented by the SIB did not really signal a decisive break with club/private regulation – but rather institutionalized those arrangements within public policy –
then the next question is, to what extent did the advent of the FSA as such change things? Or, to put the question slightly differently, did the FSA outgrow and overcome the historical, club/private regulation heritage?

With benefit of hindsight – in the circumstances observable from 2007 onward – revisionist work (including my own) suggests that the FSA represented continuation of private/club regulation within a façade of public regulation [Dorn (2014)]. In a focused study of the FSA, McPhilemy (2013) suggests that the FSA deployed a massively technical rulebook alongside continuation of cosy “club” regulation. The older tradition was hidden within the newer one: so says the revisionist story. Gilligan’s (1997) deployment of the notion of the “relative autonomy” of financial market regulation drives in a broadly similar direction.

In the angry atmosphere following political recognition of the risks taken by the financial markets, the U.K. Parliamentary Commission on Banking Standards castigated senior management within banks for evading and gaming regulation by thus setting up compliance regimes as “Potemkin villages to give the appearance of effective control and oversight, without the reality” [Parliamentary Commission on Banking Standards (2013a, 43)]. However, the parliamentarians stopped short of applying such language to regulation itself. That distinction may be merited on empirical ground (and not simply as a matter of politesse). Potemkin villages are deliberately constructed from the start in order to mislead, a process that may be fairly discerned in financial market participants’ overt and frank gaming of regulation. Can we really say the same of the historical evolution of regulation? Until we find clear evidence of conspiracy, it might be safer to think of regulation in terms of a series of sins of omission.

Bellringer and Michie (2014) have put forward the view that the opening up of the City was “an accident,” in the sense that it occurred in response to a contingent series of events. Bellringer and Michie’s work focuses on the liberalization of the London Stock Exchange, although they extend their conclusions to banking. There is, they say: “[A]n absence of evidence to connect the sequence of events that led to Big Bang, and even less to its consequences, with any conscious decision by the Conservative government whether through the actions of its politicians or officials. Nor was Big Bang the product of City influence with the deliberate intention of making London into a global financial centre. […] There is no evidence that the Conservative government under Mrs Thatcher intended to transform British banks into the dynamic sector of the British economy they had become prior to the Global Financial Crisis of 2007/6. Neither the Conservative government that won the election of 1979 nor, in all probability, the Labour government that emerged from the landslide victory in 1997 had a detailed plan of how to transform the UK financial system into a globally competitive sector that would replace manufacturing as the engine of the British economy.” [Bellringer and Michie (2014, 22)]

Some refinement of that account is in order. No doubt it is true that neither Labour, in creating the SIB, nor Conservatives, in replacing the SIB with the FSA, had “a detailed plan of how to transform the U.K. financial system.” Incoming governments quite often lack detailed plans. However, both governments held a broadly “modernizing” agenda, rather putting into doubt Bellringer and Michie’s description of the reforms as accidental. Rather, the broad direction of change was foreseen, within which negotiations and step-by-step decision making filled in the detail. As the authors observe, the Conservatives came into office in 1979 on the basis of a radical commitment to competition and the break-up of all forms of cartels. It would have been politically difficult to retreat from that position, especially since a Restrictive Trading Practices Act (RTPA) had been enacted under the receding Labour government, after which the Office for Fair Trading had referred the rule-book of the Stock Exchange to the Court of the RTPA (ibid., 12). As for market context, Bellringer and Michie astutely observe that interests in the financial markets were diverse and shifting, there being both support for the old City ways and for innovation – as represented by upstart and unclubable banks, notably S.G. Warburg, which had introduced the Eurobond market and championed hostile takeovers in the industry.

On the evidence presented by Bellringer and Michie and by many other commentators – and unsurprisingly – there were schisms that ran across politics and the markets, rather than any hegemonic bloc. For example, these authors refer to widespread City hostility to a new and upstart bank, SG Warburg. On the basis of having worked at SG Warburg, the then Secretary of State for Trade, John Nott, denounced its culture as “vigorous, unsentimental [and] meritocratic” [Nott (2002)] – a description
that would have been found apt by many at the time, be they detractors or supporters of such a manner of doing business.

Given the mingling of and antagonisms between old, clubby, aristocratic financial networks and newer, “unsentimental” meritocrats, it is unsurprising that the outcomes, in terms of regulation, were not unitary. Taking the SIB and the FSA in turn and applying to them the historical analysis mentioned above, we might characterize the SIB in terms of a technocratic but somewhat flimsy superstructure (the board itself) erected over – almost floating over – the foundations of club regulation (diverse self-regulatory bodies). As Moran (1988, 27) has it, with the establishment of the SIB and its self-regulatory bodies, “semi-private and private bodies [were] awarded a franchise to exercise legally backed powers of regulation.” Of course, each self-regulatory body more or less corresponded to an existing pocket of private regulation, already holding a franchise, so perhaps it was rather more a case of the SIB acknowledging than awarding franchises. At that stage of regulatory reform, the private nature of regulation was retained, being overtly acknowledged, although being somewhat bureaucratized: the proliferation of rulebooks began.

The FSA presented a further development of these private-public arrangements, installing club regulation within a unitary public institution. The FSA was a victory for modernity in terms of its appearance – diverse bodies were replaced by one – and in terms of its formal routines at the operational level. However, it retained clubby relations and functioning at management and policy levels (as McPhilemy and other revisionist authors cited above suggest). The FSA was not a Potemkin village – the construction of such a conspiracy would have required a level of unity and organization that was beyond the capacities available at the time – but neither was it an accident. It was more of a regroupment of historical forces, re-settling private mentality within a public façade.

3. Global subservience, national command, local activism: three regulatory playbooks

Widening our focus to Europe and the U.S., we find three regulatory playbooks: (i) regulators’ subservience to markets, which in the pre-crisis period had to some extent spilled over from the U.K. to other jurisdictions; (ii) so-called “command regulation,” in which regulators are strongly supported by or directed by federal politicians, meaning that they may be directed to apply either a heavy hand or a lighter and “recalibrating” one, depending on the political mood (a heavier hand emerging with recognition of the seriousness of the crisis, then receding); and (iii) regulators and prosecutors (acting as conduct supervisors) are sometimes locally appointed or elected, making them structurally independent of federal governments and also of regulates (and a thorn in the side for both). Here we briefly discuss these three, before going on to assess future prospects within the contemporary European scene.

3.1 Pre-crisis regulatory subservience, as revealed by the crisis

Once the depth of the crisis had been recognized, public regulation of financial markets became seen as something of an imposter. In the U.S., federal agencies became seen as being rather too close to the industry [Miller and Dinan (2009) and, for a pre-crisis review of evidence, see Bó (2006)]. Keynesian economist James Crotty put things as follows: “The design and implementation of the changes needed in financial markets is a political as much as an economic challenge. Unfortunately, most elected officials responsible for overseeing US financial markets have been strongly influenced by efficient market ideology and corrupted by campaign contributions and other emoluments lavished on them by financial corporations. […] Moreover, powerful appointed officials in the Treasury Department, the SEC, the Federal Reserve System and other agencies responsible for financial market oversight are often former employees of large financial institutions who return to their firms or lobby for them after their time in office ends. Their material interests are best served by letting financial corporations do as they please in a lightly regulated environment. We have, in the main, appointed foxes to guard our financial chickens” [Crotty (2009, 577)].

On reflection, perhaps that last line should have been written as appointing “fox cubs to guard foxes?” Leaving that aside, acerbic commentary – which would have been dismissed as oddball or (in the U.S. context) un-American before the emergence of financial market crisis – became mainstream in the years following 2008 [see for example Parliamentary Commission on Banking Standards (2013a)].

It is however fair to mention that the above judgment about ideology, emoluments, revolving doors and institutional corruption in regulatory agencies may be rather too sweeping.
Not all those who cross over from markets to regulatory agencies, or vice versa, take all of their intellectual baggage, buddy obligations or emotional attachments with them.

Indeed, some research suggests that the workings of regulatory agencies at national and at international levels are more nuanced than that. Looking at some case studies involving the U.S. Securities and Exchange Commission, Tai and Carpenter (2014) summarize that “the evidence for capture is variable [and] the precise role of the revolving door is not yet clear. These authors suggest that, rather than motivating regulators to act favorably toward industry or causing them to adopt industry arguments uncritically, the revolving door seems to act as one type of a general influence that past experience has on regulators’ perspectives” [Tai and Carpenter (2014, 227)] [regarding international regulatory committees, see Young (2012)]. In one sense, a person who goes through a revolving door may be better placed (and sometimes may even be motivated) to take action: whether they do so or not depends on circumstances (see following paragraphs). On the other hand, a person who has never worked in the industry, and who comes with a different occupational history, may struggle to make sense of aspects of the industry [especially since some of its famous complexity has been designed to “game” regulation, see McBarnet (2010) and Gerding (2013)]. So, it is not so easy to say that separating regulatory careers from industry careers would automatically improve the quality of regulation. We have to say: it would depend.

3.2 Command regulation: not necessarily restrictive

Political leadership matters: it can exacerbate collusional dynamics within the regulator-industry nexus – as U.K. regulator Lord Turner plausibly alleged in relation to U.K. Chancellor of the Exchequer Gordon Brown in the run-up to the crisis [Waugh (2009)] – or it can work against the grain of regulator cooption. In this paper, we refer to both these possibilities as command regulation, although the second sense is the more familiar use of the term. Indeed history offers us examples of regulators who gain the support of political leaders to carry through sustained regulatory reform – such as James Landis, who gained the support of several U.S. Presidents for firm regulation from the 1930s onward [O’Brien (2014), Scott/National Public Radio (2004)].

Something similar happened (albeit briefly) in the U.S. from 2008 onward, with a return to forms of command regulation that had not been seen since the New Deal. And of course the U.K. – long considered within Europe to be a bastion of neoliberalism – became equally, if not more, interventionist, with parliamentarians feeling badly let down by financial market elites, particularly banking elites [Parliamentary Commission on Banking Standards (2013b), Kerr and Robinson (2011)]. For some years, the U.K. found itself broadly in accord with other European Union member states on the need for action across the board in the years immediately following 2008, resulting in a massive re-regulatory effort, causing dizziness for regulators [Tsingou (2010)] as well as for market participants and their compliance departments.

Nevertheless, there are already reminders that command regulation is the historical exception to the rule, both in the U.S. and in Europe. In the U.S., the foxes are returning [Gandel (2015)] and there has been a sustained push back against the Dodd-Frank Act [United States (2010)], with sections of the industry putting sand in the wheels of the laborious process of moving from primary legislation to the regulatory rulebook [Coffee (2011)]. Moreover, at the end of 2014, Congress cut back aspects of the legislation [Przybyla and Wasson (2014), Johnson (2015)]. And in Europe, the longer drawn-out and deeper crisis and depression is currently being interpreted in terms of a need to encourage capital markets in particular, through the creation of a Capital Markets Union (CMU): thus command regulation may turn out to be a quite brief historical moment, which is now moderating in favor of regulatory “recalibration” [Van Steenis (2014)]. CMU seeks to overcome “barriers to a well-functioning securitization markets in the E.U.” [Bank of England and European Central Bank (2014, 15–18)]. These barriers include “too big to fail” banks, aspects of Basel capital rules and rating agencies’ caution toward derivatives in the shadow of the crisis.

Interestingly, CMU and the recalibration opportunity that it offers may be seen more as an opportunity than a threat within the banking sector, if recalibration involves some loosening of bank regulation [Milne (2014, 16)]. European Commission President Jean-Claude Juncker’s 2014 appointment letter to Lord Hill, the then Commissioner-designate for Financial Stability, Financial Services and Capital Markets Union, tasked him with finding “appropriate ways to revive sustainable and
high quality securitization markets, to reduce the cost of raising capital in the Union and to develop alternatives to our companies’ dependence on bank funding” [Juncker (2014, 4)]. Considering what this might mean in terms of regulation, Hill’s statement to the European Parliament included the following: “Now [after the banking crisis] we are entering a new phase. Although we must continue to be alert to the emergence of new risks in our system and stand ready to take appropriate action, we are unlikely, over the next 5 years, to need to pass the same amount of new legislation again. The work done by the last Commission therefore provides a clear framework for the next Commission: regulation needs to be stable as well as rigorous. The priority will thus be implementation, enforcement, and evaluation. If, during this process, evidence appears that we have not got it quite right, we should not be afraid to make quick and effective adjustments” [Hill (2014, 4)].

All of the above historical and contemporary points show that regulatory command over financial markets is possible when policy makers lend their support to it – the other side of that coin being that “policy fatigue” and/or a change in political atmosphere allows regulatory recalibration and roll-back.

3.3 Exceptionally: local and democratic mandates
Our third and last illustration of the contingent nature of the relation between markets and regulators concerns local pivots of power within federalized systems. Within the U.S., state Attorney Generals have long led the charge against market misconduct, and for good reason. In most U.S. states, Attorney Generals are elected and, as such, are subject to a governance structure and a set of motivations different from those of federal regulatory agencies. State Attorney Generals can be socially distant from federal ministries and closer to popular sentiment. Moreover, in the important case of New York, the Martin Act gives extra leverage to prosecutors [O’Brien (2005)].

The combination of these factors – electoral selection, specific powers and distance from ministerial restraint – gives market (mis)conduct regulation more “bite” in New York than elsewhere in the U.S., and also much more than in European countries. Before the emergence of the financial crisis, New York State Attorney General Elliot Spitzer castigated investment banks as being “at the vortex” of wrongdoing and he opined that “the solution is not more regulation but more innovative application of existing enforcement strategies” [Eliot Spitzer, New York Attorney General, speaking in 2004, cited in O’Brien (2005)]. Understandably, Mr Spitzer was unpopular in the industry – though not in his electorate – and in time the sort of investigative powers he had deployed against financial market misconduct in New York were turned against his person, personally disgracing him and causing his resignation [Gross (2010)]. Whether or not that was a honey-trap operation has never been ascertained. However, New York State Attorneys – and the New York Department of Financial Services (DFS) – continue to weigh in, frequently embarrassing federal agencies [Schneiderman (2014), O’Brien (2004)]. All of this suggests that, even when federal agencies lack high political backing for robust action, local circumstances can provide a basis for it.

We can now summarize the above three points about regulation and its varied capacities for capture and for activism. (i) An industry-regulation revolving door certainly exists but what goes through it may be a set of occupational experiences and mentalities, rather than sure-fire corruption. (ii) The revolving door can be locked shut by political power – for example, if one key senior policy maker gives his or her support to a regulator committed to command regulation; this tends to happen for a while after periods of crisis, during which time light touch regulation becomes less politically untenable than in boom times. (iii) Local circumstances, electoral politics and legal powers can make a difference, distancing agencies and individuals within them from federal government influences as well as from the industry. Which of these comes to the fore? We have to say, “it depends.”

4. Europe and the single market: mixed prospects for politicization
Following the Second World War, the London approach to financial market regulation was projected upwards and outwards, to the European and also up to international levels. This was important for the development of the European Community (EC), its single market and the modes of governance thereof. Elsewhere I have summarized the forces at work in the following terms: “[The Bank of England] was strategically pro-EC, acting in accord with the City’s historical international orientation. For example, to the considerable annoyance of Prime Minister Margaret Thatcher, Governor of the Bank Leigh Pemberton later went so far as to sign the 1989 report of
Jacques Delors on monetary union [...] Understanding and shaping the emerging EC agenda on financial service regulation facilitated the bank’s domestic agenda-shaping, underlining that Parliament was the agenda-taker” [Dorn (2014, 16)].

Why the Bank of England favored the development of the EC, Britain’s eventual entry into it, and the construction of the euro, is no mystery. The Community and its single market represented an opportunity for the City. Some commentators go further, seeing British entry into the EC as a part of the politics of opening up the British economy, including financial services, to the rigors of a global economy: “membership of the EC was trumpeted as a liberal measure opening up British markets to competitive pressures that would force [UK] business to rationalise” [Gifford (2007, 467–468)]. Some see British entry as transforming Europe: through entry, Britain became “a vehicle and protagonist for the globalisation of Europe” (ibid., 465). Indeed, that had been a fear of President De Gaulle, in opposing British entry. In the event, eventual entry in 1973 not only helped to open up the EC (as it then was) to international trade, it also provided a meta-reform at home: financial market regulation, once tied into the single market process, was further distanced from the domestic political process. Since then, there has been much water under the bridge, with the creation of the multi-level Lamfalussy architecture [Posner (2010)], which was then replaced by three European Supervisory Authorities and, for the Eurozone, the European Central Bank – the influence of which has become much enhanced in the context of the Eurozone crisis [Moloney (2012)].

What has not changed is the extent to which parliaments at either the regional or national levels have struggled to exert influence. The European Parliament has been rather in the position of dotting the i’s and crossing the t’s of proposals brought forward by the European Commission: a task of considerable interest to regulatees – who lobby the Parliament passionately – but not clearly connected to debates between MEPs and their publics. National parliaments have rather acted more as semi-expert arenas for governments to test out and fine-tune proposals, and as arenas to let off political steam as displays of “banker bashing” indicate, rather than stirring strong inter-party debates amongst European citizens. These are honorable roles but they also represent the historically-bequeathed, rather restricted notion of what parliaments are for, when it comes to financial market regulation.

What then might be the prospects for a broader and more adversarial public politics of finance? According to political scientists Swen Hutter and Edgar Grande, politicization means issues becoming salient not just for elites, but also for wider sections of society and for political parties, who then take a variety of positions on those issues, amounting to polarization within the political arena [Hutter and Grande (2014)].

Polarization, expressed in the form of institutional conflict in open fora, is taken as evidence of politicization, with salience being the necessary precondition. Politicization may sometimes be triggered by so-called “fringe” parties (which were quite successful in European elections in 2014). Yet, it could also result from the mobilization of “mainstream” parties.

For present purposes, the question is to how much salience, mobilizing power and polarizing potential currently attach to issues around financial markets, politicizing their governance. Such politicization would demote econocrats down to the level of operationalization of policies, rather than formulation thereof. Empirically, the answer seems to be that politicization of such issues has not strongly occurred yet, in the sense that, whilst there has been much political pressure on regulators, this pressure has come more from elite political circles to “do something,” than from political parties’ mobilization of citizens. Admittedly, in the European elections in 2014, an anti-neoliberal party in Greece, Syriza (Coalition of the Radical Left) did very well against incumbent political parties – suggesting that under conditions, such as a combination of bank bailouts plus austerity under outside direction, domestic political opinion can become polarized. Yet, in other countries, such as a Spain (admittedly not so badly hit as Greece), consensus politics has held up. Indeed, most mainstream political parties across the E.U. did not strongly articulate issues around financial markets per se – they focused more on the euro, postulating a need to save it. Nor was financial market reform a central rallying cry for the various Eurosceptic parties that have come to prominence in countries such as France and the U.K. (the latter’s UKIP is headed by an ex-commodity trader, who however talks about other issues).

So, scanning and summarizing across the EU, it seems that while fiscal issues may have been contested in some member states, financial market issues have not been strongly politicized. Why so? The context is a series of interlinked crises around banks and bailouts, the euro and austerity measures. The “highly unequal
distributional effects” of all this cause a “dramatic decline of citizens” trust in the European polity in recent years [which] is a consequence of the higher political salience and lower factual plausibility of arguments adduced to legitimate the exercise of governing powers at the European level” [Scharpf (2014)]. However, the potential political mobilizing consequences have to some extent been contained by European elites’ top level power bargaining (more between themselves and with banking interests than with their citizens), and by heads of states’ common strategy of building new arrangements of economic governance on an intergovernmental basis, outside the E.U. and escaping many of its legal and political constraints [Schimmelfennig (2014)]. Associated with this, “mainstream” political parties have sought not to rock the political boat, so do not publicly contest with each other the fundamental questions about financial markets. Elite bargaining, at a distance from party politics, is one reason that such questions have not been more politicized.

A second and rather ironic reason that financial markets have been spared politicization is the euro crisis. The euro and the various positions taken vis-à-vis it — save it, stay in it, make sacrifices for it, leave it, etc. — have definitely occupied much of the public debating space in the euro area since 2009. Lacking such a political preoccupation, banking per se might otherwise have become a bit more politicized in Europe. Such a “currency distraction” does not occur in the U.S., which does not suffer from structural conditions and contradictions that produced the euro crisis (currency union without political union).

A third reason that public debate on financial markets and their governance has remained somewhat shallow, and not politicized in the sense under discussion here, is that the financial markets have become a focus for moral outrage, as elites, media and populations react to recurrent scandals (frauds, mis-selling, market manipulation of Libor and many other benchmarks, “front running” by high frequency traders in dark pools, etc. see Schneiderman (2014)]. Under differing political circumstances, moral/ethical considerations might have acted as a stimulus to the asking of fundamental questions about what financial markets are for, and to political parties’ contestations on such questions. In the event, however, given the “issue competition” provided by the euro crisis and austerity measures, and the reluctance of mainstream political parties in most E.U. member states to open up fundamental questions about financial markets, such questions have tended to decompose into debates about punishment of individuals and the reform of “banking culture.” In parliaments, the press and scholarly work, the latter issues still tend to be posed in terms of professional ethics rather than democratic direction [see, for example, Law and Financial Markets Review, 8(2), whole issue].

For all these reasons – mainstream political parties’ reluctance to engage with big issues around financial markets, issue competition from the euro crisis and the seduction of responding to financial market participants’ bad behavior – politicization of the governance of financial markets has been limited to date. The technocratic terms of debate seem as firmly seated as ever [Blyth (2013)]. Experts argue amongst themselves in a language that is as assertive as it is impenetrable to citizens. Outside the regulatory gates, there has been public anger, yet this has taken the forms of cynicism over elites, with some displacement of anger unto foreigners and marginal groups. Banking and other financial services have indeed moved into the center of political debate, yet the discourse still escapes party political polarization (with the exception of a minority of hard-hit Eurozone countries). This is both a normative problem and a functional one.

5. Articulating public policy preferences
Normatively, it must be unacceptable that such a core aspect of governance should continue to be convened outside the arena of common politics. Tactily, political parties have adopted emergency powers, crisis resolution mentality, giving carte blanche to leaders and technocrats. Thus, whilst financial market regulation was previously implicitly off the party political agenda, it is now explicitly off. Centrist political parties of both Left and Right hold much responsibility here, for their failure to take up fundamental questions about banking and other aspects of financial markets and to stimulate debate and deliberation. Shamefully, it has been left to so-called “fringe” parties to engage with these issues, starting from the political orientations of their constituencies, then from there articulating visions of the broad purposes that should be served by financial markets.

What might such debates look like? We start with some functional questions, which on the face of things might appear to some citizens to be only lightly politicized, then move on to some
politically frank but functionally vague positions. Functionally, what kinds of financial markets citizens might want and for what purposes. For example, just personal banking or also some or all of the following. If “all,” then in what priority? Housing purchase (much more strongly emphasized in the UK than in some other countries); retirement savings (collective or individual?); domestic industrial investment (where the U.K. tradition has been to direct investment externally); and/or infrastructural development (by public or private actors?).

Cutting across preferences regarding the end-purposes of financial markets are some geopolitical questions. Political parties might articulate value positions and preferences concerning the local, regional and international foci of financial markets and their relation to the real economy. Should financial markets be open to the international winds, in a universal model that allows any firm structure in any country and local context – this being more or less the current arrangement for the U.S. and the E.U. Alternatively, should particular types of financial markets be encouraged by particular countries, and other types by other countries, on the basis of complementarity with countries’ economic and social policies? Such differentiation is currently discernable on the wider global level – taking account of China, Russia and other countries [Tett and Farchy (2015)] – and is possibly being consolidated as the post-crisis norm [Helleiner and Pagliari (2011)]. And/or, should financial markets, or aspects of such markets, be subordinated to regional and local policies and economies, on the model of old-style local and regional banks [Knafo (2006, 183), and for prospective comments see Leaver and Williams (2014, 221)]?

Also cutting across questions about the purposes of financial markets is the question of who pays for them when they blow up. Are citizens still prepared to bail out banks and other vehicles at national level? That is the model perfected in the U.S. [Dorn (2012)] but made notorious in Ireland and many other countries from 2008 onward. Do citizens – and the political parties that compete to gain their votes – prefer to mutualize such risks at regional level? That is not an approach that currently gains much support in the E.U., although some actions of the European Central Bank allegedly arrive at such a result [which may survive legal challenge, see Cruz Villalón (2015)]. If citizens and political parties do not advocate any public support, then they need other means through which difficulties could be dealt with. National and European legislation has put in place resolution policies, according to which public bailout of failing banks is only one option, and not the first one. However, the long timeframe envisaged for the building up of banks’ own resolution funds makes clear that, for the next decade at least and possibly for the foreseeable future, public funds remain vulnerable. These distributional questions are inherently political.

Let us briefly mention three broad political profiles. Unreconstructed free marketers, such as neoliberal Republicans and Tea Party people in the U.S., disfavor state “interference.” They favor leaving the industry – and its customers – to their own devices. That would mean allowing failing banks, for example, to do just that: to fail. That policy option has so far been held at arm’s length, due to concerns about possible knock-on effects, especially through bank bond-holdings. By contrast, and with relevance in some parts of Europe, some political constituencies articulate the issues in terms of economic justice, the direction of financial recourses to the real economy, and possibly public ownership of financial entities and infrastructures (on a continuing basis, not just as an emergency intervention). For green parties, the purposes of the financial industry, as with other industries, include transformations aiming to safeguard the planet and its ecosystems. That could mean special attention to modes of financing (and hence governance) of commodity producers and traders, and control of financial instruments referencing foodstuffs, for example.

And so on, the general approach being to start not with bankers’ or regulators’ agendas, but rather with political parties’ convictions, seeking to shape finance accordingly: chacun à son gout. The principle may be extended from domestic to international policy preferences. Consider, for example, the dissatisfaction that arose in French political circles in 2014 over the large level of fines, business curtailments and reputational damage that BNP Paribas suffered at the hand of the U.S. authorities in 2014, arising from the bank’s sustained breaches of U.S. sanctions law in respect of Iran and Sudan (the latter’s President having been indicted by the International Criminal Court, see Protess and Silver-Greenberg (2014)). As an article in the Economist noted: “The case has left people on both sides of the Atlantic unhappy. One reason is that the individuals responsible seem to be getting
Regulatory herding versus democratic diversity: history and prospects

of the extraterritorial reach of U.S. justice [Financial Times (2014)]. Such jurisdiction remains applicable as long as banks clear in dollars, which at some stage have to pass back to the U.S., triggering the latter’s jurisdictional claim. Could the euro or the yuan provide an alternative vehicle for clearing? Potentially, however for the present the dollar remains the principal currency of international trade and it would take some time and effort to put another currency into a comparable position. Maybe in time the BRICS will develop a new channel. If so, bank regulation might not only diverge from U.S. norms in respect of international sanctions, it might also diverge on a regulatory wider front.

In some regions and countries, national political elites’ historical concern over U.S. economic dominance coalesces with dislike of the extraterritorial reach of U.S. justice [Financial Times (2014)]. Such jurisdiction remains applicable as long as banks clear in dollars, which at some stage have to pass back to the U.S., triggering the latter’s jurisdictional claim. Could the euro or the yuan provide an alternative vehicle for clearing? Potentially, however for the present the dollar remains the principal currency of international trade and it would take some time and effort to put another currency into a comparable position. Maybe in time the BRICS will develop a new channel. If so, bank regulation might not only diverge from U.S. norms in respect of international sanctions, it might also diverge on a regulatory wider front.

Summarizing, we can say that as for the future, there is potential for financial market regulation to be shaped – and differentiated – by the diverse foreign policies of countries and blocs, as well as by their domestic policy preferences. Against that, regulatory elites to some extent transcend domestic and regional differences, so reinforcing preferences within transglobal financial service firms for similarity in regulatory regimes. Reading this, most readers will know which side they are on.

6. Conclusion

Regulators crowd when they form a common culture, which, although it is partially detached from their diverse origins as individuals, does allow them to understand each other and to negotiate the meaning of events and what should be done about them. Going a step further, regulators herd when they intentionally and strategically set out, negotiate and agree global principles and rules (which nevertheless may be fine-tuned operationally and reputationally with an eye to jurisdictional competitiveness). Arguably, regulators were crowding before the crisis and, since then, have been engaged in various (sometimes fraught) arguments over the extent to which they should herd, at the European level and across the Atlantic [Akhtar and Jones (2013)].

Both crowding and herding are distinct from the final stage in convergence, which is merger. In the market, mergers occur through takeovers or through agreed mergers of equals, producing “financial institutions” that may then be regarded as too big to fail. Such was the situation that became recognized from 2008 onward, which is not solved today, as many regulators have pointed out [Haldane (2013)]. It is difficult and arguably impossible to discipline such entities.

By analogy, what are we to think about global and regional regulators when they are in various stages of merger, such as in the E.U. (and particularly in the Eurozone)? Do they then become too big to fail, in the sense that they become politically untouchable? That question will come more into focus as European arrangements bed down: the Eurozone has a degree of centralized and technocratic rule that will be much studied and debated over the coming years. Whilst it is welcome that the European Parliament and national parliaments scrutinize financial market legislation proposals that have been proposed by other actors, such scrutiny remains agenda-taking.

Meanwhile, the wider international regulatory networks, which before the crisis were dominated by U.S. and European ideas and personnel, have lost prestige and leadership capacity. As one transatlantic network has lamented: “the future of coherent global financial regulation is unclear [...] the United States and European Union must act expeditiously and collaboratively if they are to continue as leaders of financial reform on the global stage” [Bowles et al. (2013)]. Those authors – the first named
of whom stood down as a British Liberal Democrat MEP at the May 2014 elections, having up until then chaired the European Parliament’s Economic and Monetary Committee – advocate the global regulatory herding that has been critically dissected in this paper. Our concern is not about whether or not regulation should (continue to) be under U.S. and E.U. leadership. Rather, we draw attention to the adverse consequences inherent in any international regulatory herding.

As regulators internationally converge in their thinking – as regulatory crowding (thinking and acting similarly because sharing the same social and epistemic space) evolves into regulatory herding (thinking and acting similarly because we are strategically emulating each other) – so the questions of “too similar” and “too big” become as relevant for financial market regulations as for markets. We are only at the start of recognizing these phenomena and their consequences.

Acknowledgement
This paper draws upon the author’s 2014 book, Democracy and Diversity in Financial Market Regulation, Abingdon: Routledge. During the period in which the book was developed, the author was employed in the Erasmus School of Law, Rotterdam.

References
Atlantic Council, 1
Dorn, N., 2014, Democracy and diversity in financial market regulation, Abingdon: Routledge
Gerding, E., 2013, Law, bubbles, and financial regulation, Abingdon: Routledge
Gross, D., 2008, “Spitzer gets Spitzered: how Spitzer was brought down by the same kind of investigation he pioneered,” Slate, 10 March
Haldane, A., 2013, “Have we solved ‘too big to fail?’” VoxEU, 17 January
Regulatory herding versus democratic diversity: history and prospects


Johnson, S., 2015, “The Republican strategy to repeal Dodd-Frank,” The Baseline Scenario (blog), 7 January


Law and Financial Markets Review, 8(2), whole issue


McPhilemy, S., 2013, “Formal rules versus informal relationships: prudential banking supervision at the FSA before the crash,” New Political Economy, 18:3, 748-767


Squandering home field advantage? Financial institutions’ investing in their own industries

Aneel Keswani
Reader in Finance and Director of the Centre for Asset Management Research, Cass Business School, City University

David Stolin
Professor of Finance, Toulouse Business School, University of Toulouse

Abstract
In the extensive debate about investment professionals’ ability to add value, there has been scant evidence on the role played by industry expertise. To shed light on this issue, we study own-industry investing. Specifically, we analyze how well individual mutual funds as well as mutual fund companies, banks and insurance companies invest in the shares of listed companies in the mutual fund, banking and insurance industries, respectively. We find little evidence that such inherent industry knowledge enhances these institutions’ ability either to time the investment in their industry as a whole or to select individual stocks within it.

1 We thank Harjoat Bhamra, Chris Brooks, Giovanni Cespa, Giacinta Cestone, Pasquale Della Corte, Stefan Jacewitz, Robert Kosowski, Art Kraft, Ian Marsh, Kjell Nyborg, Marco Pagano, Grzegorz Pawlina, Richard Payne, Ken Peasnell, Alan Pendleton, Tarun Ramadorai, Christel Rendu de Lint, Diego Ronchetti, Stefano Rossi, Sergey Sanzh, Lucio Sarno, Henri Servaes, Clemens Schild, Raman Uppal, Steven Young, participants at the Leading Lights in Fund Management Research II Conference, Midwest and Southwestern Finance Association annual meetings in New Orleans, and seminar participants at the Universities of Lancaster, Munich, Reading, Imperial College and at Cass Business School. All errors are ours.
1. Introductions

Do investment professionals add value? This question has been the subject of intense debate since the seminal paper of Jensen (1968), who found no evidence that mutual funds beat passive benchmarks. More recently, Barras et al. (2010), Busse et al. (2010) and Fama and French (2010) all report that, on average, active mutual funds do not justify their fees, while Lewellen (2011) finds little evidence of stock-picking skill for the entire institutional investor universe, including banks and insurance companies.

In spite of the above, academics have been able to identify a variety of circumstances that are correlated with better stock selection decisions by managers [see Cuthbertson et al. (2010) for a survey]. Notably, these include having an informational advantage over other investors as expressed by geographic proximity to one’s investments (Coval and Moskowitz (2001), Baik et al. (2010)), having greater industry focus (Kacperczyk et al. (2005)) and being part of the same social networks [Cohen et al. (2008)]. These papers, therefore, suggest that managerial ability is present by demonstrating that better information leads to higher investment returns. However, these are vague on how this occurs and the exact nature of the information concerned.

In this paper, we examine fund managers’ ability to generate abnormal returns by leveraging a concrete informational advantage they have, namely industry expertise. A recent book by two leading mutual fund industry practitioners defines obtaining “insights from industry sources” as one category of fundamental research used by financial analysts and emphasizes the value added in being an industry insider: “To get information that is not available to the public at large, analysts must turn to other sources within the industry. They’ll read the trade publications, and they’ll consult with industry experts. In fact, some analysts have once worked in the industry they now cover, so that they have an insider’s view of its trends and prospects.” (Pozen and Hamacher (2011, p. 108))

Several academic papers document intra-industry transfer of value-relevant information, while industry specialization by financial analysts suggests that industry insight is perceived to be of crucial importance by the investing world. We expect industry insiders to be better informed because they are more adept at processing public information about their industry (e.g., by reading specialized publications), and because they have greater access to information about the industry that may not be readily available to industry outsiders, whether it is inferred from their own firms’ operations or absorbed through formal and informal professional networks. Projected sales growth, cost structures and trends in product development, as well as “soft” information, for example, about different firms’ human capital are all value-relevant and more readily available to industry insiders. One may, therefore, expect investors with firsthand experience in the same industry as their investments to reap the rewards from being well informed about their own industry’s future profitability and the specific firms that are likely to outperform going forward.

Indeed, investment managers’ faith in the value of industry expertise is demonstrated by the popularity of so-called expert networks, which are meant to provide superior interpretation of non-privileged information while circumventing legal pitfalls. This faith is particularly striking given that agency problems inherent in dealing with outside experts, as well as the treacherous legal terrain on which such dealings take place (as highlighted recently by the Galleon insider trading scandal) add considerably to the costs of relying on external expertise of this sort. Seen from this perspective, investing in one’s own industry brings the advantages of an expert network without any of the costs.

Berk and Green (2004) present a theoretical model where managerial skill does not manifest itself in the form of positive abnormal returns due to a combination of decreasing returns to scale in investment management and inflows to the best performing portfolios. Their work suggests that to reveal managerial skill one needs to move away from fund-level performance tests. Alexander et al. (2007) do this and use trade-based data instead to test for skill. They find evidence

---

2 See Koo et al. (2010) for recent evidence and references to earlier work.

3 Hong et al. (2005) report evidence of word-of-mouth effects among mutual fund managers.

4 According to BusinessWeek, “Gerson, Vista, and Nitron [three prominent expert network firms] say they would never set up an investor interested in a company with someone who worked at that company – the discussions are supposed to focus only on industry trends or technical questions. […] Vista Managing Director Stanton Green points out that stock analysts have always called on industry experts, usually without any signed agreements, adding that he believes ‘it is much safer to be having these conversations within our service than not.’”

of managerial skill but only when trades are information-motivated rather than being just a passive response to fund flows. Their paper highlights the importance of controlling for contemporaneous fund-level flows when measuring trade performance.

Together, these insights set up a new experiment to shed light on investment professionals’ skill. Mutual fund companies, banks and insurance companies all manage pools of assets and have to make security selection decisions. We examine whether they do so better in their own industry (respectively, mutual fund management and distribution, banking and insurance). If investment professionals are unable to leverage their informational advantage over other investors by performing better in the industry they know most intimately, this would give novel and powerful ammunition to those who are skeptical about traditional money managers’ stock-picking skills. If, conversely, industry-insider investments do outperform outsider ones, this would raise interesting questions as to what specific features of industry information allow them to do so.

Our analysis below is conducted using the following steps. We first focus on the mutual fund industry. We do so because most of the analysis of institutional investors’ skill to date has been done in the mutual fund context. We begin by constructing a sample of trades that are likely to be information-based. To filter out trades that are made in response to investor flows, we focus on trades where the percentage change in the number of shares held is 20% greater in absolute terms than the contemporaneous percentage fund-level money flow. We also impose a minimum size of trade of U.S.$200,000 to focus on trades that are material in absolute terms. Thus, our sample consists of trades that are substantial both in relative and absolute terms.

Using these information-based trades, we then examine whether mutual fund managers anticipate their industry’s market performance better than they are able to anticipate the market performance of other industries. While we find some evidence of industry timing ability as a whole, we do not find that fund managers are able to time their own industry significantly better. Furthermore, when we perform our timing tests industry by industry, we find that fund managers are able to predict the performance of the majority of control industries better than their own.

Second, we examine whether mutual fund managers can pick individual stocks in their own industry better than in other industries. To do this, we measure each period where a fund manager increases or decreases their position in a given stock and then test whether this correlates more strongly with subsequent stock-level performance for own-industry stocks or not. As variations in the information environment across industries might explain trade performance differences, we include variables to control for these differences in our regressions. In addition, as nearby and smaller companies are arguably easier to value, we control for these aspects as well. Still, our stock selection tests by mutual fund managers reveal no evidence of superior stock selection ability in their own industry.

We then conduct similar tests at the mutual fund company level, as well as for banks and insurance companies. Overall, we find little evidence of superior industry timing or stock selection ability when financial institutions invest in their own backyard. Seen against the backdrop of existing trade-based studies that have shown evidence of managerial ability, a lack of evidence of skill in own-industry trading is a surprising and thought-provoking result.

The rest of the paper is organized as follows: section 2 describes our data and sample formation, section 3 presents and discusses the results, and section 4 concludes.

2. Data
Our data on mutual fund and financial institution stockholdings originate from U.S. Securities and Exchange Commission (SEC) filings. U.S. institutional investment managers who exercise investment discretion over U.S.$100 million or more must report their holdings on Form 13F with the SEC each quarter. We obtain our data from Thomson-Reuters S12 and S34 datasets. The S12 data contain 13F holdings for individual mutual funds, while S34 data contain 13F holdings for institutions, including banks, insurers and mutual fund companies (in the latter case, aggregated across all individual funds). Since there are no data on mutual fund holdings on Thomson-Reuters prior to 1980, this is the year our sample period starts, and it runs through December 2009.

We begin by identifying U.S. exchange-listed ordinary shares issued by banks, insurers and mutual fund companies (henceforth MFCs, to distinguish fund complexes from individual mutual
Squandering home field advantage? Financial institutions’ investing in their own industries

Determining whether the target security is a bank or an insurer is easy as banks have a North American Industry Classification System (NAICS) code beginning with 5221 and insurers have a NAICS code beginning with 5241. In the average quarter during our sample period, there existed 346 different listed banks and 106 different listed insurers.

Defining whether a company is in the mutual fund business is more difficult, as MFCs do not map cleanly into NAICS codes. We begin our MFC sample construction by examining company earnings on Compustat by business segment, and retaining only companies where asset management (excluding hedge funds) averages over 70% of net sales for the two preceding years. A number of large financial conglomerates that undertake a wide range of financial services also sell mutual funds. By focusing on companies whose main line of business is mutual funds, we exclude these firms from our MFC sample. This is consistent with the purpose of our study: unless firms have the majority of their business stemming from mutual funds, we would not expect mutual fund managers to have an industry-insider advantage in valuing such companies.

Our paper is concerned with investments by mutual funds, banks and insurers in their own industry. Panel A, in Table 1, presents statistics compiled from the S34 database on the extent to which these financial institutions’ shares are held by their own industry. These statistics largely reflect the prominence of the investing industry. On average, during our study period, banks held 16.36% of listed banks’ shares, as contrasted with the 15.47% they held in listed insurers and only 10.04% in MFCs. Likewise, insurance companies held a greater percentage of shares in their own industry (5.62%) than they did either in banks (4.24%) or in MFCs (4.02%). MFCs, however, held only 3.47% of shares in their own industry, more than their holdings of banks (2.22%) but less than their holdings of insurers (4.55%), controverting the notion that institutional investors are biased toward their own industry.

Panel B shows what proportion of the institutions’ equity portfolios are invested in the three industries. These statistics largely reflect the weight of these industries in the U.S. stock market. While the banking industry accounts for a greater share (4.01%) of banks’ equity portfolios than does either the insurance industry (3.85%) or the mutual fund industry (2.69%), insurers do not display an analogous “home bias”: insurance companies represent only 3.07% of insurers’ equity portfolios, less than they do of banks’ (3.30%). Lastly, as the market capitalization of the mutual fund industry is much smaller, it represents only a fraction of a percentage point of institutional investors’ portfolios.

<table>
<thead>
<tr>
<th>Panel A: Percent of listed financial firms’ equity held by 13F institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listed banks</td>
</tr>
<tr>
<td>Percent held by banks</td>
</tr>
<tr>
<td>Percent held by insurance companies</td>
</tr>
<tr>
<td>Percent held by MFCs</td>
</tr>
<tr>
<td>Average number of listed firms</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Listed financial firms as percent of 13F institutions’ equity portfolios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listed banks</td>
</tr>
<tr>
<td>Percent of banks’ portfolios represented by</td>
</tr>
<tr>
<td>Percent of insurance company portfolios represented by</td>
</tr>
<tr>
<td>Percent of MFC portfolios represented by</td>
</tr>
<tr>
<td>Percent of U.S. stocks’ market value represented by</td>
</tr>
</tbody>
</table>

Table 1: Investment by banks, insurers and MFCs in their sectors

---

5 Specifically, some come from code 5231 (“securities and commodity contracts intermediation and brokerage”), some from code 5239 (“other financial investment activities”) and some from code 5259 (“other investment pools & funds”). Further, these codes also contain companies that are not primarily involved in mutual fund management and distribution.

6 Details as well as the resulting sample of 38 companies representing 378 company years are available from authors.
fraction is greater, however, for portfolios of MFCs (0.21%) than those of banks or insurers.

Panel B also contains U.S. market cap weights for listed banks, insurers and MFCs. By comparing these market cap weights with the actual portfolio fractions held in banks, insurers and MFCs it is possible to understand whether these institutions overweight or underweight their own industry. Only MFCs appear to be (slightly) overweight their own industry.

In order to evaluate institutions’ ability to exploit their industry knowledge, we will compare the performance of their buys and sells in their own industry with the performance of their buys and sells in other industries. Accordingly, Table 2 presents data on trades by mutual funds of MFC shares, as derived from the S12 dataset.

To ensure that the trades in our sample are not due to the need to (dis)invest money due to investor (out)flows [Alexander et al. (2007)], we adjust each position change for the contemporaneous money flow to the equity portfolio to which the position belongs. That is, our quarterly flow-adjusted position change for stock i, portfolio j and quarter t is calculated as

$$\text{Change}_{i,j,t} = \frac{\text{Shares}_{i,j,t} \cdot \text{Shares}_{i,j,t+1} \cdot \text{Value}_{i,j,t} \cdot \text{Value}_{i,j,t+1} (1 + r_{j,t})}{\text{Shares}_{i,j,t+1} \cdot \text{Value}_{i,j,t+1}}$$

where Shares$_{i,j,t}$ is the number of shares (adjusted for splits) held in the portfolio, Value$_{i,j}$ is the market value of the portfolio and $r_{j,t}$ is the portfolio’s investment return during the quarter.

We impose several constraints to ensure we capture only meaningful and substantive position changes. First, we only consider portfolios-quarters for which the change in portfolio size is not explained by investment returns,

$$\frac{\text{Value}_{i,j,t} \cdot \text{Value}_{i,j,t+1} (1 + r_{j,t})}{\text{Value}_{i,j,t+1}}$$

does not exceed 20% in absolute value. This filters out cases where a change in the portfolio size from one quarter to the next is so large as to be likely the result of data errors or unusual events such as mergers or extreme money flows. Further, in order to ensure that we capture only substantive changes in shareholdings, we only retain those trades for which Change$_{i,j,t}$ is greater than 20% in absolute value.

In other words, for a portfolio that did not experience any money flows in a given quarter, we focus on changes in stockholdings that are below −20% or above 20% of the original position. If a portfolio experienced an inflow of 10% during a quarter, we focus on changes in stockholdings that are below −10% or above 30%. Our method ensures that, say, a 15% increase in the holding of a stock would not be considered as an informative trade if it is merely the result of allocating a 15% money inflow to existing positions. At the same time, by choosing the same 20% cutoff value for both portfolio inflow and for the change in stockholdings, we avoid the possibility of

<table>
<thead>
<tr>
<th>Trade type</th>
<th>Number of trades</th>
<th>Mean</th>
<th>1st quartile</th>
<th>Median</th>
<th>3rd quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sells</td>
<td>2,439</td>
<td>3,043</td>
<td>2,355</td>
<td>778</td>
<td>290</td>
</tr>
<tr>
<td>Buys</td>
<td>2,580</td>
<td>2,600</td>
<td>2,094</td>
<td>703</td>
<td>272</td>
</tr>
</tbody>
</table>

Table 2: Mutual fund buys and sells of MFC shares
considering unchanged positions as trades even if the portfolio experienced a substantial money flow. Lastly, we require the position change to reflect an (unadjusted) trade size of at least U.S.$200,000 to focus on trades that are significant in absolute terms, as well. While the 20% / $U.S.200,000 cutoff is arbitrary, we also rerun our results with a 10% / U.S.$100,000 cutoff, and our conclusions remain qualitatively unchanged.

Table 2 presents these trades for two levels of cutoff size. In the first case, the cutoff is set at 10% and both the upper bound on absolute flow and the lower bound on change are set at this level (with a minimum trade size of at least U.S.$100,000). In the second case, this cutoff is set at 20% (and the minimum trade size in this case is at least U.S.$200,000). With a 10% cutoff, there are around 2,500 eligible buys and sells each and the average (median) trade size is around U.S.$3 million (U.S.$0.75 million). With a 20% cutoff, the number of eligible trades is lowered by about a fifth while the average (median) trade size is around a third (a half) higher, at over U.S.$3 million (U.S.$1 million). In either case, the trades we focus on are large both in relative and absolute terms, and are, therefore, presumably the result of careful decision-making.

Analogous to Table 2, Table 3 shows bank, insurer and MFC trades of shares in the same sector as derived from the S34 dataset. Focusing on trades selected with the 20% cutoff (which is the basis for results reported in the remainder of the paper), the number of MFC trades (803 buys and 719 sells) is a fraction of the number of individual fund trades reported in the previous table. This reflects both aggregation of individual fund trades at the fund company level and the fact that the earlier table included all mutual funds, not just those run by listed companies with little or no activity outside the mutual fund business. This is also reflected in the much larger average (median) trade size — U.S.$16.6 (U.S.$4.3) million for buys and U.S.$16.7 (U.S.$4.1) million for sells.

The number of trades by insurers (banks) in their own industry is more than 10 (40) times greater than the corresponding number of MFC trades: this reflects their greater number both as investors and as investments. The average trade size, however, is substantially smaller when looking at own-industry trading by banks and insurers as compared to that of MFCs, likely reflecting the fact that the average bank or insurer in our sample is smaller than the average MFC, both as investor and as investment.

Even in the absence of transaction costs, not all tradeworthy information would lead to changes in quarterly stockholdings that we use to identify trades. Some round-trip trades may take place within a single quarter. Actual trading may not materialize due to regulatory constraints on short selling, or on the maximum

<table>
<thead>
<tr>
<th>Trade type</th>
<th>Trade size (U.S.$000s)</th>
<th>Number of trades</th>
<th>Trade size (U.S.$000s)</th>
<th>Number of trades</th>
<th>Trade size (U.S.$000s)</th>
<th>Number of trades</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>Mean</td>
<td>Median</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td><strong>Using a 10% cutoff</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sells</td>
<td>28,411</td>
<td>8,263</td>
<td>1,233</td>
<td>1,183</td>
<td>8,801</td>
<td>9,703</td>
</tr>
<tr>
<td>Buys</td>
<td>37,294</td>
<td>8,517</td>
<td>1,183</td>
<td>1,183</td>
<td>8,501</td>
<td>9,703</td>
</tr>
<tr>
<td><strong>Using a 20% cutoff</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sells</td>
<td>28,411</td>
<td>8,263</td>
<td>1,233</td>
<td>1,183</td>
<td>8,801</td>
<td>9,703</td>
</tr>
<tr>
<td>Buys</td>
<td>37,294</td>
<td>8,517</td>
<td>1,183</td>
<td>1,183</td>
<td>8,501</td>
<td>9,703</td>
</tr>
</tbody>
</table>

Table 3: Bank, insurer and MFC trades of shares in the same sector
amount of stock that can be held. However, this caveat applies to all research using 13F holdings data, and we have no reason to believe that our study is particularly sensitive to this issue.

3. Results
We begin by analyzing the trades of mutual fund managers using the S12 dataset and then examining the performance of financial institutions' trades using the S34 data. We consider in each case whether industry timing and stock selection ability are stronger when trades are in own-industry securities.

We measure security level quarterly abnormal returns using a two-step approach. We first calculate the Daniel et al. (1997) characteristic-adjusted returns for each stock, where stocks are matched to 1 of 125 benchmark portfolios on the basis of their market capitalization, book-to-market ratio and their returns over the previous year. Then we four-factor adjust these characteristic-adjusted returns using the Carhart (1997) model.

Using both characteristic adjustment and then four-factor adjustment is the approach taken by Cohen et al. (2010), and we follow it to ensure that we control as fully as possible for market, size, value and momentum factors. Further, in order to control for industry effects, our stock-level analyses examine deviations of the Daniel- and Carhart-adjusted measure from the corresponding industry members' equal-weighted or market value-weighted average.

In order to compare the performance of MFC trades in their own industry with the performance of their trades in other industries, we form a control sample of industries using NAICS industry definitions. We wish to select industries of comparable size to the mutual fund industry, so that fund managers have a comparable incentive to acquire industry-specific information, and so that industry-level performance measures for the control industries have similar statistical properties. Since we have close to 400 firm-year observations for the mutual fund industry, we select as our controls 6-digit NAICS industries that have between 200 and 600 firm-year observations on Compustat between 1979 and 2009. This results in 167 NAICS codes, from which we exclude those not on the official 2007 list of NAICS codes, and those which include any of the stocks in our MFC sample.

Table 4: Mutual fund trades and subsequent industry-level performance

Panel A: Pooled regression across industries

<table>
<thead>
<tr>
<th>Risk-adjusted industry portfolio performance</th>
<th>Equally weighted</th>
<th>Value weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.00043</td>
<td>-0.00607</td>
</tr>
<tr>
<td>PROP_BUY</td>
<td>0.00447</td>
<td>0.01290</td>
</tr>
<tr>
<td>BUY_OWN</td>
<td>0.00972</td>
<td>0.00709</td>
</tr>
<tr>
<td>Number of observations</td>
<td>4,120</td>
<td>4,120</td>
</tr>
<tr>
<td>Clustering by time</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

R-squared

|          | 0.0037 | 0.00179 |

Panel B: Summary of regressions by industry

<table>
<thead>
<tr>
<th>Value of PROP_BUY coefficient</th>
<th>Mutual fund industry</th>
<th>Mean for control industries</th>
<th>Median for control industries</th>
<th>Proportion of control industries with higher value of PROP_BUY coefficient than the mutual fund industry</th>
<th>Number of control industries with enough observations to estimate the regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutual fund industry</td>
<td>-0.00480</td>
<td>0.01602</td>
<td></td>
<td></td>
<td>127</td>
</tr>
<tr>
<td>Mean for control industries</td>
<td>0.00448</td>
<td></td>
<td>0.02180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median for control industries</td>
<td>0.00443</td>
<td></td>
<td>0.01718</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of control industries with higher</td>
<td>0.620</td>
<td></td>
<td>0.516</td>
<td></td>
<td></td>
</tr>
<tr>
<td>value of PROP_BUY coefficient than the mutual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fund industry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of control industries with enough</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>observations to estimate the regression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Mutual fund trades and subsequent industry-level performance

7 The Investment Company Act of 1940 Rule 12d-3 specifies certain investment constraints for diversified and non-diversified mutual funds. U.S. mutual funds are prohibited from owning more than 5% of other investment companies, defined as firms that derive more than 15% of revenue from securities-related activity. In addition, there are restrictions on the asset concentration of diversified funds as they may hold no more than 5% of their assets in any one particular company. Limits for banks are set by The Bank Holding Companies Act of 1956, which states that a bank cannot acquire more than 5% of the outstanding voting shares of any company or bank. As U.S. insurer asset-holding regulation is at the state level, rules vary according to locality. Babbel and Fabozzi (1999) present cross-state asset allocation guidelines, which state that generally “an insurer may not purchase common stock which would result in more than 5% of the value of the common stock portfolio being invested in the securities of one issuer.” In practice, hardly any of the holdings we observe are close to the limits discussed above.

8 While the value-weighted average is the more meaningful economically, representing the industry’s aggregate market performance, this can be overly influenced by a single firm. Consequently, as an alternative, we also use equally weighted averages.


10 The list of the resulting 148 control industries is available from the authors.
only company years where the company existed as a listed entity until fiscal year-end.)

We start our performance analysis by comparing mutual fund managers’ ability to time their aggregate own-industry investments to their ability to time other industries. To this end, for each industry in each quarter we calculate the proportion of all substantive mutual fund trades (i.e., those that clear the 20% hurdle defined in the preceding section) that are represented by buys. This gives us a measure of mutual funds’ bullishness about the industry’s prospects. We then regress the value-weighted or equally weighted characteristic and four-factor-adjusted monthly abnormal performance of stocks in each industry over a quarter on the proportion of buys in that industry in the prior quarter (PROP_BUY) as well as a dummy variable (BUY_OWN), which is the product of PROP_BUY and a dummy variable equal to 1 if the industry is the mutual fund industry, and 0 otherwise. Our results are presented in Table 4 Panel A (where t-statistics are adjusted for time clustering). The positive loading on the proportion of buys suggests that fund managers buy more heavily in industries that do well in the subsequent quarter, however, the t-statistics of 0.74 (in the case of equally weighted industry portfolios) and 1.79 (for value-weighted portfolios) do not reach statistical significance. Most importantly for our purposes, the insignificant coefficient on BUY_OWN (t-statistics of 1.42 and 0.61) mean that that there is no significant evidence that fund managers time their own industry better than they do other industries.

Panel B of Table 4 presents results for the same type of timing analysis performed at the individual industry level. Specifically, for each industry, we regress the value-weighted or equally weighted quarterly abnormal performance on the proportion of mutual fund buys in that industry in the prior quarter. If managers have the ability to predict the performance of a given industry, the coefficient on PROP_BUY should be positive, and if they have a greater ability to predict returns in their own industry then the coefficient on PROP_BUY should be greater in the case of the mutual fund industry. Table 4 shows that the coefficient on PROP_BUY is on average positive across industries, consistent with our pooled industry analysis. However, the coefficient on PROP_BUY is below the corresponding average performance of control industries. Further, mutual fund managers are able to predict the future performance of 62% (52%) of control industries (if performance is measured on a value-weighted basis) better than of their own industry when performance is measured on a value-weighted (equally weighted) basis.

We next turn to mutual fund managers’ ability to select individual stocks, comparing it to other-industry stock selection. To do so, we start by regressing the monthly abnormal performance of individual fund trades over the subsequent quarter on a buy dummy (BUY, equal to 1 for buys, and to 0 for sells) together with a BUY_OWN dummy that is equal to one if the transaction is a buy, and if the company whose stock is being traded belongs to the mutual fund industry. If stocks that are purchased subsequently outperform those that are sold, then the coefficient on BUY should be positive and significant. When we look at our results in Table 5 Panel A, regression (1), we find that the coefficient on our BUY dummy variable is a statistically significant 0.00123 per month, which translates into a spread of more than 1% per year between mutual fund buys and sells. These results echo the findings of previous studies that look at the performance of mutual fund trades, such as Chen et al. (2000), who also find significant evidence of stock selection ability by mutual funds. Regression (3), where the industry adjustment is value-weighted, also has a significant coefficient on BUY, though smaller in magnitude. Most importantly for our purposes, however, the coefficient on BUY_OWN in both regression (1) and regression (3), which captures whether managers are able to trade stocks better in their own industry than they do those in other industries, is not statistically distinguishable from zero. This suggests that mutual fund managers do not enjoy an informational advantage when selecting stocks in their own industry.

In the same table, we also show results of regressions that follow the setup described above, but additionally include a number of control variables suggested by the literature. As geographic distance has been shown to aid manager stock selection ability [Coval and Moskowitz (2001)], we measure the distance between the investing party’s zip code and the target company’s ZIP code. If this distance is less than 100 miles, we set the dummy variable CLOSE to one, and interact it with the BUY dummy to form BUY_CLOSE, which tells us how much mutual fund buys outperform their sells when the headquarters of the company whose stock is being traded are close to the headquarters of the trading institution. In addition, we also include proxies of the extent to which there is stock-
level private information. Differences in private information might explain why the alpha available from trading varies in the cross-section of securities. Aslan et al. (2011) show that analyst following correlates negatively with private information. We obtain monthly data on the number of analysts providing a one-fiscal-year-ahead forecast of earnings for each stock in our sample where available from the I/B/E/S database. Our FOLLOWING variable equals the logarithm of one plus the number of different one-year forecasts, and we interact it with the BUY dummy to form BUY_FOLLOWING, a control variable that captures the effect of analyst following on the spread between the performance of buys and sells. Ang et al. (2009) suggest that idiosyncratic volatility may also proxy for private information. We, therefore, note the residual risk estimated from our Carhart (1997) four-factor regressions and interact it with the BUY dummy to form BUY_RESRISK as a further measure of private information in our regressions. Lastly, as larger firms may be more difficult to value due to their greater complexity or number of business lines, we interact the logarithm of the market value of the company being traded with BUY to form BUY_SIZE. Equations (2) and (4) in Table 5 show that the addition of these control variables does not change our key finding that own-industry buys do not outperform.

Panel B of Table 5 reports results from running regressions analogous to those reported in Panel A, but on an industry-by-industry basis, and hence without the BUY_OWN control variable. Column (1) of the table shows that mutual fund buys of stocks within their own industry perform 0.171% per month better than corresponding sells over the subsequent quarter, on a risk- and industry-adjusted basis. By contrast, the average of the corresponding quantity across the 140 control industries with enough data to estimate the regressions is 0.271% per month, and it is greater than for the mutual fund industry for 45% of the control industries. For models (2) through (4), mutual fund trades have a higher estimated value of the BUY coefficient for trades in their own industry than either the average or the median of their estimated BUY coefficient for trades in control industries. However, none of these estimated values are statistically significant at conventional levels. Further, mutual fund trading of shares in their own industry is worse than their trading of shares in 14% to 40% of control industries. Thus, there is no significant evidence that mutual funds have superior stock-picking ability in their own backyard than they
Squandering home field advantage? Financial institutions’ investing in their own industries

This complements our earlier finding that mutual funds do not time the performance of their own industry better than they do that of other industries.

Next, we present the results of own-industry timing and stock selection tests conducted for insurers, banks and MFCs using the S34 dataset. In lieu of a control sample, we examine banks, insurer and MFC trades in the two industries other than their own.

Table 6 examines the industry timing ability of financial institutions’ trades in the three financial industries they represent. Specifically, we pool industry-quarter observations and relate the abnormal performance of a given industry’s stocks (equally weighted or value weighted) in a given quarter to the proportion of buy trades in the previous quarter (PROP_BUY) made by all three types of investing institutions, as well as by each type separately. We also include in the same regression a dummy variable equal to the product of PROP_BUY and a dummy variable equal to one when the industry being invested in is the same as the industry doing the investing. Panel A reports results for equally weighted industry portfolios, and Panel B, for value-weighted ones. As the panels show, the coefficient on BUY_OWN is never significant, with the exception of MFCs investing in their own industry, when the t-statistic for the case where the industry adjustment is equally weighted reaches 2.00, so that the coefficient estimate is significant at the 5% level. However, this result is not robust to value-weighting, for which case the t-statistic drops to a very low 0.24. For banks and insurers, as well as for all three financial institution types together, the coefficient on BUY_OWN is negative. In short, there is no strong evidence of financial institutions being able to predict subsequent industry performance better for their own industry than for other financial industries.

Lastly, Panels A and B of Table 7, similarly to Panel A of Table 5, examine whether financial institutions have stock selection ability by regressing the returns of the securities they trade on a BUY dummy and a BUY_OWN dummy, as well as (in even-numbered regression) on the four control variables, BUY_CLOSE, BUY_FOLLOWING, BUY_RESRISK, and BUY_SIZE. The insignificant coefficient on the BUY dummy across specifications tells us that whether we look at bank, insurance company or MFC trades or at the trades of all three institutions together, there is little evidence of stock selection ability (but recall that we restrict ourselves here to the financial services sector). Most crucially, BUY_OWN never reaches significance in any of the 16 regressions in the two panels. Thus, there is no evidence that being an industry insider results in better stock picking.

Our interpretation of the overall results is as follows. We find that the average fund trade – here, across control industries – is not profitable, nor is there significant evidence of successful industry timing by mutual funds in aggregate. Most important for our purposes, though, there is no evidence at all that mutual fund managers are able to make better investment decisions in the industry they know most intimately – their own.

When we look at entire financial institutions investing in shares
### Panel A: Industry adjustment factor is equally weighted

<table>
<thead>
<tr>
<th></th>
<th>All institutions</th>
<th>Investing institution type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.00006</td>
<td>0.00035</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.29</td>
</tr>
<tr>
<td>BUY</td>
<td>-0.00062</td>
<td>0.00228</td>
</tr>
<tr>
<td></td>
<td>-1.15</td>
<td>0.56</td>
</tr>
<tr>
<td>BUY_OWN</td>
<td>0.00055</td>
<td>0.00060</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>0.98</td>
</tr>
<tr>
<td>BUY_CLOSE</td>
<td>-0.00074</td>
<td>-0.00125</td>
</tr>
<tr>
<td></td>
<td>-0.78</td>
<td>-0.94</td>
</tr>
<tr>
<td>BUY_FOLLOWING</td>
<td>-0.00080</td>
<td>-0.00057</td>
</tr>
<tr>
<td></td>
<td>-0.64</td>
<td>-0.46</td>
</tr>
<tr>
<td>BUY_RESRISK</td>
<td>-0.01890</td>
<td>-0.00890</td>
</tr>
<tr>
<td></td>
<td>-0.51</td>
<td>-0.22</td>
</tr>
<tr>
<td>BUY_SIZE</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td></td>
<td>0.33</td>
<td>0.25</td>
</tr>
<tr>
<td>Number of observations</td>
<td>148,039</td>
<td>112,727</td>
</tr>
<tr>
<td>Clustering by time</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.00003</td>
<td>0.00019</td>
</tr>
</tbody>
</table>

### Panel B: Industry adjustment factor is value weighted

<table>
<thead>
<tr>
<th></th>
<th>All institutions</th>
<th>Investing institution type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.00061</td>
<td>0.00972</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>1.20</td>
</tr>
<tr>
<td>BUY</td>
<td>-0.00007</td>
<td>0.00371</td>
</tr>
<tr>
<td></td>
<td>-0.12</td>
<td>1.00</td>
</tr>
<tr>
<td>BUY_OWN</td>
<td>-0.00057</td>
<td>-0.00030</td>
</tr>
<tr>
<td></td>
<td>-0.96</td>
<td>-0.57</td>
</tr>
<tr>
<td>BUY_CLOSE</td>
<td>-0.00047</td>
<td>-0.00124</td>
</tr>
<tr>
<td></td>
<td>-0.52</td>
<td>-0.98</td>
</tr>
<tr>
<td>BUY_FOLLOWING</td>
<td>-0.00081</td>
<td>-0.00053</td>
</tr>
<tr>
<td></td>
<td>-0.75</td>
<td>0.51</td>
</tr>
<tr>
<td>BUY_RESRISK</td>
<td>-0.03469</td>
<td>-0.03588</td>
</tr>
<tr>
<td></td>
<td>-0.90</td>
<td>-0.86</td>
</tr>
<tr>
<td>BUY_SIZE</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td></td>
<td>0.44</td>
<td>0.27</td>
</tr>
<tr>
<td>Number of observations</td>
<td>148,039</td>
<td>112,727</td>
</tr>
<tr>
<td>Clustering by time</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.00003</td>
<td>0.00031</td>
</tr>
</tbody>
</table>

Table 7: Financial institution trades and subsequent stock performance
of other financial institutions, there is likewise no evidence of investment ability either at the individual stock, or at the industry level. Most damningly, however, we find that each type of financial institution is unable to convert its “home field advantage” – buying and selling shares in its own industry – into actual investment returns.

In unreported analyses, we have also examined the possibility that own-industry trades generate superior returns at longer horizons, by studying two-, three-, and four-quarter abnormal returns following the trading quarter and using the Newey-West method to allow for the overlap in observations that the longer return horizon induces. We find no evidence of any longer-horizon superior performance. We stress that none of our analyses take into consideration the effect of trading costs. The fact that we do not find robust evidence of outperformance by three major types of institutional investor in their own backyard, despite studying the world’s largest stock market over a three decade span and despite ignoring trading costs, is striking.

The lack of evidence of own-industry timing or stock selection ability could come about if these industries (mutual funds, banking or insurance) are inherently more difficult to time or select stocks in than other industries. As a result, it would be difficult to ascertain whether it is lack of ability or the information environment that is to blame for the weak own-industry investment performance we observe.

One way to separate out these two aspects and appraise the value of being an industry insider is to hold the trading environment constant and examine how variation in own-industry ability impact investment performance. Our “all institutions” tests in Tables 6 and 7 do just this, as the trade performance of the institutions with own-industry knowledge is benchmarked against the performance of the institutions that do not when trading the same security. In this case, the BUY_ OWN dummy in these “all institutions” tests informs us whether there is any significant difference in trading ability between institutions that have own-industry knowledge and those that do not. The fact that this dummy is not significant in Tables 6 and 7 suggests that differences in the inherent predictability of security returns across industries do not explain the lack of an own-industry effect.

Kacperczyk et al. (2005) find greater investment ability among funds whose portfolios are more concentrated at the industry level. They attribute this finding to the value of industry-level information. In contrast, our paper finds the reverse. How can we reconcile the two sets of results?

While the Kacperczyk et al. argument is plausible, their evidence on the value of industry information is indirect. In particular, it is possible that although funds specialize based on their informational advantage, this information is concentrated across industries, but is not industry information as such. For example, funds may have better information about more proximal investments [Coval and Moskowitz (2001)], and geographic and industrial distribution may be correlated. While Kacperczyk et al. check that investment styles and company size do not subsume their result, there may be other industry-correlated variables that do. Second, it is possible that more industry-concentrated funds have better stock picking not because they have better industry information, but because being industry concentrated correlates to other fund characteristics that are liable to produce better performance. For example, funds that are disciplined enough to stay focused on a small number of industries may also have a more disciplined approach to investment analysis in general, and such discipline may allow them to reap greater returns regardless of any industry-level informational advantage. Third, even if one interprets the Kacperczyk et al. analysis as saying that industry information helps one make better investment decisions, our paper shows that having industry information is not a sufficient condition for outperforming the market in that industry.

Indeed, in line with our findings, a recent paper examining the performance of individual investors in Norway [Døskeland and Hvide (2011)] finds that they get neutral or negative abnormal returns when investing in the same industry as the one they work in.

4. Conclusion

If financial institutions exhibit any investment skill at all, one could expect it to be particularly evident when they invest in stocks of companies that are involved in the same business as themselves. Yet we find that institutions do not display any superior investing ability in their own industry, either at the individual stock or at the aggregate industry level. This is in spite of our focus on trades that are large in absolute and relative terms.
amounts to an indictment of some gravity. trades, their failure to beat the competition on their home turf and are systematically unaware of this fact.” Although we do managers trade, on average, against better informed insiders, in this manner, it must be the case that professional money the fund management industry: for our findings to be explained theory would amount to a profoundly negative indictment of this possibility given the available data, acceptance of this knowledge of future cash flows. While it is difficult to rule out association between institutional trading and stock returns et al. (2011) discuss the possibility that “the negative In an important recent study of institutional trading, Dasgupta et al. (2011) discuss the possibility that “the negative association between institutional trading and stock returns arises because institutions trade against insiders with superior knowledge of future cash flows. While it is difficult to rule out this possibility given the available data, acceptance of this theory would amount to a profoundly negative indictment of the fund management industry: for our findings to be explained in this manner, it must be the case that professional money managers trade, on average, against better informed insiders, and are systematically unaware of this fact.” Although we do not observe systematically negative returns to asset managers’ trades, their failure to beat the competition on their home turf amounts to an indictment of some gravity.

Our paper also offers a novel perspective on stock market efficiency. The idea that in expert hands public information can be used to beat the market is intuitively appealing, and is indeed the reason behind the proliferation of expert network firms. On the other hand, unless the industry expertise furnished is truly unique, one can expect the rewards to such expertise to be competed away in a market setting. While the secretive nature of the expert network business makes it difficult to examine this issue across a wide range of industries, we show that for several important industries in the financial services domain, industry expertise does not translate into greater investment returns.

In References


Pozen, R., and T. Hamacher, 2011, The fund industry: how your money is managed, Wiley Finance

Trond, M. D., and H. K. Hvid, 2011, “Do individual investors have asymmetric information based on work experience;” Journal of Finance 66, 1011-1031

The new Banking Union landscape in Europe: consolidation ahead?

Dirk Schoenmaker
Dean, Duisenberg School of Finance, and Professor of Finance, Banking and Insurance, VU University Amsterdam

Abstract
The establishment of the Banking Union (B.U.) creates a large banking market comparable to that of the U.S. This paper calculates the market share of the top 20 banks in the new B.U. France appears to take a prominent place with 5 banks in the top 10, followed by Germany, the Netherlands and Italy. Earlier integration episodes did not lead to cross-border consolidation in Europe. In contrast, the lifting of interstate banking restrictions caused a cross-state merger wave cumulating in large U.S.-wide banks. This paper investigates whether cross-border consolidation can be expected within the B.U. The answer is yes over time, but not yet as subdued growth, lingering influence of national supervisors and cultural differences may hamper cross-border mergers in the short run. Over time, the B.U. will become an integrated market, where banks can manage their balance sheet at the aggregate B.U. level and the European Central Bank (ECB) conducts supervision with a European perspective.

1 I am grateful to Floris van Ham for research assistance on mergers and acquisitions, to Tim Drost for comments and to Louis Pauly for valuable discussions on the future of the Banking Union.
The new Banking Union landscape in Europe: consolidation ahead?

1. Introduction
The Economic and Monetary Union (EMU) in Europe got its banking equivalent in 2014. The move to B.U. can be seen as a first, major step toward a Financial Union. Next milestones may be a Capital Markets Union and an Insurance Union. The establishment of the B.U. creates a large banking market within the European Union (E.U.), comparable to that of the U.S. Over time, the B.U. may become an integrated market, where banks can manage their balance sheet at the aggregate B.U. level and the ECB conducts supervision with a European perspective. But in the short run national supervisors may still prevent European banks from operating on a European scale, as they informally request banks to lend or invest in the same country as where deposits are collected.

The aim of this paper is to examine the long-term impact of the B.U.. We review previous integration steps, such as the Single Market in 1993 and the EMU in 1999. While cross-border consolidation was predicted at the time of these integration events, it did not happen. By contrast, the lifting of the ban on cross-state mergers in the 1990s led to major cross-state consolidation in the U.S. [Stiroh and Strahan (2003)]. The question is whether the B.U. may lead to a similar consolidation wave in Europe. We argue that consolidation does not only depend on the dynamics at banks, but also on developments at the policy front. The B.U. has still some national features, such as national deposit insurance schemes, and is thus not yet complete. A key determinant for further integration is the political acceptance of risk sharing within the monetary and banking union. That in turn depends on the likelihood of political union.

For banks, cross-border banking delivers diversification benefits [Dermine and Schoenmaker (2010)]. The business cycle is not synchronized across euro-area countries leading to divergent country patterns of credit risk. Moving to the demand side, it may take some time before consumers regard a bank from elsewhere in the B.U. as a “domestic” bank to which they can entrust their money. When that happens, a truly integrated retail banking market will emerge. Corporates, especially the larger ones, are expected to adapt faster and select their main banks from across the B.U.

This paper is organized as follows. Section 2 analyses the new European banking landscape. Next, section 3 examines earlier integration episodes. Section 4 reviews the major drivers of future consolidation in the B.U. Section 5 sketches the new international landscape. Finally, section 6 concludes.

2. The European banking landscape
The advance to the B.U. in November 2014 created a new European banking landscape. We first review the B.U. landscape at the country level. That enables us to compare and contrast the new B.U. of Europe with one of the oldest banking unions, that of the U.S. Next, we take a more detailed look at the largest banks in the B.U.

2.1 Banking systems at country level
The B.U. is built on the EMU. Euro area countries are automatically members of the B.U., while non-euro area countries have the option to join. So far, none of the outs has made use of this option. With the adoption of the euro in January 2015, Latvia has also joined the B.U. at that date. The E.M.U. and the B.U. now consist of 19 member countries. The nine E.U. members that do not use the euro are: Denmark, Sweden, the U.K., Hungary, Poland, Czech Republic, Bulgaria, Romania and Croatia.

The E.U. banking system can be split into the B.U. and the non-B.U. countries. Table 1 indicates that the B.U. covers about 75% of total E.U. banking assets. The E.U. countries have relatively closed banking markets, with assets of banks from other E.U. countries at 14% and from third countries at 3%. The overall cross-border penetration for the E.U. is higher, due to the U.K. with business from third countries at 28% (see Table A.1 in the Annex). This highlights the current status of London as the international financial center. Will London continue to service the B.U., or will Frankfurt emerge as the financial center of the B.U.?

The B.U. banking system is comparable to that of the U.S. in several ways. The number of banks is about 6,000 (see Table 1). The system has a strong domestic orientation with 83% of...
all assets, while 17% of assets come from other E.U. and third countries. Foreign bank affiliates account for 14% of the U.S. banking system. But there is an important difference. The U.S. banking system has fewer assets, €12 trillion (amounting to 94% of U.S. GDP) compared with the B.U., which has €30 trillion (amounting to 302% of euro area GDP). The U.S. financial system depends less on bank intermediation and more on capital markets and nonbank financial institutions [Schularick and Taylor (2012)]. The European financial system is excessively bank-based (also characterized as “overbanking”) and expected to move toward a more balanced bank- and market-based system [ASC (2014)].

2.2 Major banks

The major banks in the B.U. and the U.S. are also comparable. For our analysis, we use the new home base, which has become the entire B.U. area for B.U. banks. Figure 1 shows the geographical segmentation of the top 20 banks in the three regions (E.U., B.U. and U.S.). The large B.U. and U.S. banks have more than 70% of their assets at home (i.e., the B.U. and the U.S., respectively). The rest of the region (i.e., the rest of Europe – the non-B.U. part – and the rest of North and South America) counts for 11%, while the rest of the world amounts to about 18%. The large E.U. banks are more international. They not only have a smaller home base (one country), but they also have more business in the rest of the world. Examples of major global banks outside the B.U. are HSBC, Barclays and Standard Chartered from the U.K.

The picture emerging from our analysis at country and bank level is that the B.U., just like the U.S., is a relatively closed banking system with limited inward and outward expansion.

2.2.1 Picture of the largest banks

Table 1: Banking systems across three regions: B.U., E.U. and U.S.; end of 2013

<table>
<thead>
<tr>
<th></th>
<th>Number of banks</th>
<th>Total assets in €b</th>
<th>Of which: home (in %)</th>
<th>Other E.U. (in %)</th>
<th>Third country (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.U.</td>
<td>5,999</td>
<td>30,035</td>
<td>83</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>Non-B.U.</td>
<td>1,724</td>
<td>12,008</td>
<td>60</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>E.U.</td>
<td>7,723</td>
<td>42,043</td>
<td>77</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>U.S.</td>
<td>6,813</td>
<td>11,862</td>
<td>86</td>
<td>-</td>
<td>14</td>
</tr>
</tbody>
</table>

Note: Total banking assets come from the home country, other E.U. countries and third countries (i.e., outside the E.U. or the U.S.). The three components add up to 100%.

Source: Author calculations based on ECB for European banks and Federal Reserve, FDIC and Flow of Funds for U.S. banks.

Zooming in on the European banking landscape, banks can be divided into four categories depending on the international composition of their assets [Schoenmaker (2013)]. Table A2 (in the Annex) shows the biggest 30 banks in Europe before the start of the B.U. A global bank has less than 50% of its assets in the home country and the majority of its international assets in the rest of the world. These banks include HSBC, Barclays and Standard Chartered from the U.K., Deutsche Bank from Germany and Credit Suisse and UBS from Switzerland. The major global players in Europe are thus the U.K. and Swiss.
The new Banking Union landscape in Europe: consolidation ahead?

A European bank has less than 50% of its assets in the home country and the majority of its international assets in the rest of Europe. Some European banks focus on a specific region in the E.U. The Nordea Group, for example, primarily operates in the Nordic and Baltic region. Other European banks operate Europe-wide; examples include BNP Paribas, UniCredit and ING.

A semi-international bank has between 50% and 75% of its assets in the home country. Examples are RBS from the U.K., BBVA from Spain, Commerzbank from Germany and KBC from Belgium. Finally, a domestic bank has more than 75% of its assets in the home country. These banks include Crédit Agricole, Lloyds Banking Group, Rabobank and Intesa Sanpaolo.

2.3 The new landscape

The start of the B.U. entails a paradigm shift for banks and policymakers. The home market expands for banks from their country to the wider B.U. This paper presents new data on the top 20 banks in the emerging B.U. market. Table 2 contains the geographic segmentation, splitting a bank’s business between the B.U., the rest of Europe (i.e., businesses in the non-B.U. member states) and the rest of the world.

Some of the European banks operating on a regional basis (the

<table>
<thead>
<tr>
<th>Banking group</th>
<th>Market share in B.U. (%)</th>
<th>Total assets (€b)</th>
<th>Of which: B.U. (%)</th>
<th>Rest of Europe (%)</th>
<th>Rest of world (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Crédit Agricole</td>
<td>5.0</td>
<td>1,707</td>
<td>89</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>2. BNP Paribas</td>
<td>4.0</td>
<td>1,800</td>
<td>66</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>3. Société Générale</td>
<td>3.4</td>
<td>1,235</td>
<td>82</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>4. Groupe BPCE</td>
<td>3.1</td>
<td>1,124</td>
<td>84</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>5. Deutsche Bank</td>
<td>2.5</td>
<td>1,612</td>
<td>46</td>
<td>12</td>
<td>42</td>
</tr>
<tr>
<td>6. UniCredit</td>
<td>2.3</td>
<td>846</td>
<td>82</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>7. Crédit Mutuel</td>
<td>2.1</td>
<td>659</td>
<td>94</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>8. ING Bank</td>
<td>2.0</td>
<td>788</td>
<td>76</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>9. Intesa Sanpaolo</td>
<td>1.9</td>
<td>626</td>
<td>92</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>10. Rabobank</td>
<td>1.8</td>
<td>674</td>
<td>80</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>11. Banco Santander</td>
<td>1.3</td>
<td>1,116</td>
<td>34</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>12. Commerzbank</td>
<td>1.1</td>
<td>550</td>
<td>63</td>
<td>21</td>
<td>16</td>
</tr>
<tr>
<td>13. DZ Bank</td>
<td>1.1</td>
<td>387</td>
<td>87</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>14. La Caixa Group</td>
<td>1.1</td>
<td>351</td>
<td>96</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>15. ABN AMRO</td>
<td>1.1</td>
<td>372</td>
<td>90</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>16. BBVA</td>
<td>1.1</td>
<td>583</td>
<td>57</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>17. Landesbank Baden-Würt.</td>
<td>0.8</td>
<td>274</td>
<td>85</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>18. Bayerische Landesbank</td>
<td>0.7</td>
<td>256</td>
<td>83</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>19. KBC Group</td>
<td>0.6</td>
<td>241</td>
<td>72</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>20. Erste Group</td>
<td>0.4</td>
<td>200</td>
<td>56</td>
<td>41</td>
<td>3</td>
</tr>
<tr>
<td><strong>Top 20 banks</strong></td>
<td><strong>37.3</strong></td>
<td><strong>15,400</strong></td>
<td><strong>73</strong></td>
<td><strong>11</strong></td>
<td><strong>16</strong></td>
</tr>
</tbody>
</table>

Table 2: Top 20 banks in B.U., end of 2013

Note: A bank’s market share in the B.U. is calculated as a bank’s assets in the B.U. divided by total banking assets in the B.U. (€30,035b from Table 1). A bank’s total assets are divided into assets in the B.U., in the rest of Europe, and the rest of the world. The top 20 is ranked by market share.

Source: Assets are taken from The Banker (July 2014). The segmentation of assets is calculated by the author based on annual reports.
second group in Table A2) have now become pan-B.U. banks. BNP Paribas, UniCredit and ING Bank operate throughout the B.U., with 65% to 80% of their assets in the B.U.. These banks are comparable with the super-regional banks in the U.S., such as Bank of America (see below), with a large presence across the whole region. Also the semi-international and domestic banks (the third and fourth group in Table A2) have become large players in the B.U. It should be added that the B.U. has only one truly global player with Deutsche Bank and one regional player with Banco Santander. All other banks have the majority of their activities within their new home market, the B.U. This confirms the relatively closed nature of the B.U. market, discussed in section 2.1.

The second column in Table 2 illustrates the B.U. market shares. The market share of the biggest banks in the B.U. lingers at about 2% to 5%, which is low. The top 5 banks by market share are four French banks (i.e., Crédit Agricole, BNP Paribas, Société Générale and Groupe BPCE) and a German bank (i.e., Deutsche Bank). The market share of the 20 biggest banks amounts to 37%. The prominent position of the French banks is due to their large presence across the B.U., ranging from 66% to 89%. By contrast, Deutsche Bank is more international with a strong presence in London and the U.S., but only 46% in the B.U.. Furthermore, the major Spanish banks, Banco Santander and BBVA, have a strong presence in London (for Santander), the U.S. and South America.

3. Earlier integration episodes
It may be instructive to examine earlier integration episodes in Europe and the U.S. The question is whether cross-border mergers happened around these integration episodes. To answer that question, bank merger data for the Eurozone and the U.S. are used. The following criteria are used to identify bank mergers:

1. Both target and acquirer are a bank or a bank holding company.
2. Both target and acquirer are from one of the 11 countries that introduced the Euro in 1999 for the Eurozone sample, and both target and acquirer are from the U.S. for the US sample.
3. The acquirer holds less than 50% of the shares of the target before the merger and more than 50% of the shares after the merger.

The value of these mergers is aggregated by year, and type of merger (domestic or cross-border). The data is obtained from Thomson’s SDC Platinum database.
The new Banking Union landscape in Europe: consolidation ahead?

3.1 U.S. integration episodes
The major U.S. banks were formed after the lifting of restrictions on interstate banking by the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994 [Brook et al. (1998)]. Through several mergers and acquisitions, super-regional banks, such as JPMorgan Chase and Bank of America, emerged with a market share of 13% and 11%, respectively. Stiroh and Strahan (2003) provide interesting evidence on consolidation after the deregulation. They show a competitive reallocation of assets to better performers. Better banks did grow, while the poorly performing banks shrank as well, and those with the worst performance shrank the most. Figure 2 illustrates the cross-state merger wave after the Riegle-Neal Act. A large increase in cross-state mergers is observed, while the volume of within state mergers is relatively flat. So, the lifting of interstate restrictions did lead to a cross-border merger wave in the U.S..

3.2 European integration episodes
The Single Market in 1993 and the start of the EMU in 1999 did not lead to the ~ at the time ~ widely expected European consolidation. In contrast, several domestic mergers happened in anticipation of the new setting. Examples are the merger of ABN and AMRO in 1991 and the creation of the BNP Paribas Group from the merger of BNP and Paribas in 1999. To analyze the impact of these events, we use data for the original 11 members of the Eurozone. Figure 3 presents Eurozone merger volumes around the start of the Single Market in 1993. No large cross-border mergers with targets from the Eurozone took place in the entire five-year window around the Single Market. From 1994 onwards, however, increased domestic merger volume is observed. Figure 4 shows Eurozone merger volumes around the introduction of the euro in 1999. Large domestic merger volumes are observed in the year before and the year following the introduction of the euro. Only a small increase in cross-border merger volume is observed.

Around all three events, it seems that domestic banks acquired attractive targets before the major changes were introduced, while foreign and out-of-state acquirers waited until the new changes actually came into being. In part, this can be explained by the simple fact that foreign and out-of-state ownership was complicated or not allowed at all. But the introduction of the euro did not lift any formal restrictions. Rather, it seems, foreign banks waited until the new currency was actually introduced. This begs the question whether the new B.U. will have similar consequences for the economic environment of banks and similar patterns of bank takeovers (domestic instead of cross-border) may be expected.

4. Consolidation ahead?
What are the expectations on the consolidation front for the B.U.? The current level of overbanking suggests further consolidation to eliminate excess capacity [ASC (2014)]. The question is whether this consolidation will be domestic or cross-border. We examine this question from various perspectives: the market structure, the policy setting, banking strategies and, importantly, banking clients.

4.1 Market structure
The global financial crisis has led to consolidation in the banking sector. Takeovers and mergers are a well-known tool for resolving ailing banks. This consolidation was mainly domestic with a few exceptions, such as the takeover of the Belgian and Luxembourg parts of Fortis by the French bank BNP Paribas and the takeover

---

3 Federal law prevents any bank from gaining more than 10% of national deposits in the U.S. through acquisition. So, JPMorgan Chase and Bank of America can only organically grow in the U.S. Lucas (2014) ranks the top 5 U.S. banks by assets.
of some parts of Lehman Brothers by Barclays. As a result of this domestic consolidation wave, concentration ratios have, on average, increased by 3% within the B.U. countries. Table 3 illustrates that several countries have concentration ratios of over 70% (Estonia, Finland, Greece, Lithuania, Malta, Netherlands, Portugal and Slovakia). For these countries, there is not much scope for further domestic consolidation.

Moving from the country level to the B.U. level, the market share of the five largest banks (CR5) in the B.U. is 18% (see Table 2). To compare, the CR5 is on average 47% for individual B.U. countries and 48% for the U.S.. Even in a large country with a dispersed banking system, like Germany, the CR5 is over 30%. There is thus scope for further consolidation across the B.U.

### Table 3: Concentration (CR5) in the B.U. countries, 2005-2013 (in %)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>45%</td>
<td>43%</td>
<td>37%</td>
<td>38%</td>
<td>37%</td>
<td>-8%</td>
</tr>
<tr>
<td>Belgium</td>
<td>85%</td>
<td>83%</td>
<td>77%</td>
<td>71%</td>
<td>64%</td>
<td>-21%</td>
</tr>
<tr>
<td>Cyprus</td>
<td>60%</td>
<td>65%</td>
<td>65%</td>
<td>61%</td>
<td>63%</td>
<td>+3%</td>
</tr>
<tr>
<td>Estonia</td>
<td>98%</td>
<td>96%</td>
<td>93%</td>
<td>91%</td>
<td>90%</td>
<td>-8%</td>
</tr>
<tr>
<td>Finland</td>
<td>83%</td>
<td>81%</td>
<td>83%</td>
<td>81%</td>
<td>84%</td>
<td>+1%</td>
</tr>
<tr>
<td>France</td>
<td>52%</td>
<td>52%</td>
<td>47%</td>
<td>48%</td>
<td>46%</td>
<td>-6%</td>
</tr>
<tr>
<td>Germany</td>
<td>22%</td>
<td>22%</td>
<td>25%</td>
<td>34%</td>
<td>31%</td>
<td>+9%</td>
</tr>
<tr>
<td>Greece</td>
<td>66%</td>
<td>68%</td>
<td>69%</td>
<td>72%</td>
<td>94%</td>
<td>+28%</td>
</tr>
<tr>
<td>Ireland</td>
<td>48%</td>
<td>50%</td>
<td>53%</td>
<td>47%</td>
<td>48%</td>
<td>0%</td>
</tr>
<tr>
<td>Italy</td>
<td>27%</td>
<td>33%</td>
<td>31%</td>
<td>39%</td>
<td>40%</td>
<td>+13%</td>
</tr>
<tr>
<td>Latvia</td>
<td>67%</td>
<td>67%</td>
<td>69%</td>
<td>60%</td>
<td>64%</td>
<td>-3%</td>
</tr>
<tr>
<td>Lithuania</td>
<td>81%</td>
<td>81%</td>
<td>80%</td>
<td>85%</td>
<td>87%</td>
<td>+6%</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>35%</td>
<td>31%</td>
<td>29%</td>
<td>31%</td>
<td>34%</td>
<td>-1%</td>
</tr>
<tr>
<td>Malta</td>
<td>75%</td>
<td>70%</td>
<td>73%</td>
<td>72%</td>
<td>77%</td>
<td>+2%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>84%</td>
<td>86%</td>
<td>85%</td>
<td>84%</td>
<td>84%</td>
<td>0%</td>
</tr>
<tr>
<td>Portugal</td>
<td>69%</td>
<td>68%</td>
<td>70%</td>
<td>71%</td>
<td>71%</td>
<td>+2%</td>
</tr>
<tr>
<td>Slovakia</td>
<td>68%</td>
<td>68%</td>
<td>72%</td>
<td>72%</td>
<td>70%</td>
<td>+2%</td>
</tr>
<tr>
<td>Slovenia</td>
<td>63%</td>
<td>59%</td>
<td>60%</td>
<td>59%</td>
<td>57%</td>
<td>-6%</td>
</tr>
<tr>
<td>Spain</td>
<td>42%</td>
<td>41%</td>
<td>43%</td>
<td>48%</td>
<td>56%</td>
<td>+14%</td>
</tr>
<tr>
<td>Average B.U. countries</td>
<td>44%</td>
<td>44%</td>
<td>44%</td>
<td>48%</td>
<td>47%</td>
<td>+3%</td>
</tr>
</tbody>
</table>

Note: CR5 is the concentration ratio of the top five banks and equals the aggregate size of the five largest banks relative to the size of all banks. The average for the B.U. countries is asset-weighted. The last column reports the difference between 2005 and 2013 (as a percentage).

Source: E.U. structural financial indicators, ECB.

4.2 Policy setting

At the policy level, there are diverging forces at work. On the one hand, the harmonization of banking regulations in the Single Rule Book as well as the centralization of supervision at the ECB may lift the final barriers to a truly integrated banking market and unleash a cross-border merger wave in European banking. On the other hand, company law, insolvency law and taxation are still organized at the national level. At the time of writing, only the Single Resolution Board is centralized, whereby a gradual shift from national resolution funds to a Single Resolution Fund is foreseen [Gros and Schoenmaker (2014)]. By contrast, the lender of last resort
The new Banking Union landscape in Europe: consolidation ahead?

The official condition is that the country would have to be unable to provide financial assistance to the beneficiary bank without very serious effects on its own fiscal sustainability [Question 5, ESM (2014)].

4 The application of risk sharing across the B.U. is thus still limited. Political economy suggests that another banking crisis may be needed before all functions, from supervision to crisis management, are aligned at the B.U. level.

4.3 Banking strategies
Cross-border banking delivers credit risk diversification benefits to banks. These diversification benefits can only be reaped when business cycles are not fully synchronized within the Eurozone. Figure 5, taken from Saiki and Kim (2014), illustrates that the correlation has increased from 0.7 to 0.8 in the aftermath of the global financial crisis in 2008. Nevertheless, the average correlation remains well below the case of full synchronization with a correlation of 1. Dermine and Schoenmaker (2010) observe several cases of banks that have fared much better than others during the global financial crisis thanks to diversification. Examples are Santander, BBVA, HSBC and BNP Paribas. Similarly, Slijkerman (2007) finds that a merger between banks from different countries offers a better opportunity for risk diversification than a merger between two domestic banks. Mergers by banks in the same country increase systemic risk. Credit risk diversification thus remains an important driver for cross-border banking.

4.4 Banking clients
Turning to the demand side, we examine the preferences of banking clients. Corporates, especially the larger multinationals, are already using major international banks. European multinationals do no restrict themselves to European banks operating on an international scale. Citibank, for example, is a bank of choice for multinationals because of its global payment systems coverage (Citi is member of the payment system in most Western countries). Multinationals are thus already used to selecting their main banks from across the B.U. (and beyond).

The picture at the retail level, both households and small businesses, is different. Although the Single Euro Payments Area (SEPA) facilitates euro transfers across Europe, it will take some time before consumers regard a bank from elsewhere in the B.U. as a “domestic” bank to which they can entrust their money. That will, first of all, depend on cultural and language differences across Europe. It may also depend on deposit insurance arrangements. The debacle with the Icelandic banks (which did not live up to their responsibility of home country deposit insurance) has made depositors across Europe wary of relying on deposit insurance from another country. Moreover, Gros and Schoenmaker (2014) show that national deposit insurance funds are less stable than a European deposit insurance fund. A move to European deposit insurance may thus help cross-border banking (see also the earlier discussion on policy setting).

Moving to small businesses, Degryse and Ongena (2004) find that bank lending to small businesses (SMEs) in Europe is characterized both by local pricing and regional and/or national market segmentation. The European Commission also defines the relevant market for assessing bank mergers for banking services to SMEs as national (or even regional). In several
<table>
<thead>
<tr>
<th>Banking groups</th>
<th>Total assets in €b</th>
<th>World assets rank</th>
<th>Capital surcharge in %</th>
<th>Home country % of total assets</th>
<th>Rest of region % of total assets</th>
<th>Rest of world % of total assets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Global banks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. HSBC (U.K.)</td>
<td>1,937</td>
<td>2</td>
<td>2.5</td>
<td>37</td>
<td>11</td>
<td>52</td>
</tr>
<tr>
<td>2. Deutsche Bank (B.U.)</td>
<td>1,612</td>
<td>10</td>
<td>2.0</td>
<td>46</td>
<td>12</td>
<td>42</td>
</tr>
<tr>
<td>3. Barclays (U.K.)</td>
<td>1,568</td>
<td>11</td>
<td>2.0</td>
<td>36</td>
<td>26</td>
<td>38</td>
</tr>
<tr>
<td>4. Citigroup (U.S.)</td>
<td>1,364</td>
<td>13</td>
<td>2.0</td>
<td>43</td>
<td>12</td>
<td>45</td>
</tr>
<tr>
<td>5. UBS (Switzerland)</td>
<td>821</td>
<td>23</td>
<td>1.0</td>
<td>33</td>
<td>25</td>
<td>42</td>
</tr>
<tr>
<td>6. Credit Suisse (Switzerland)</td>
<td>710</td>
<td>25</td>
<td>1.5</td>
<td>23</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>7. Standard Chartered (U.K.)</td>
<td>489</td>
<td>42</td>
<td>1.0</td>
<td>16</td>
<td>5</td>
<td>79</td>
</tr>
<tr>
<td><strong>Regional banks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Banco Santander (B.U.)</td>
<td>1,116</td>
<td>18</td>
<td>1.0</td>
<td>34</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>2. Nordea (Sweden)</td>
<td>631</td>
<td>31</td>
<td>1.0</td>
<td>24</td>
<td>75</td>
<td>1</td>
</tr>
<tr>
<td><strong>Semi-international banks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. BNP Paribas (B.U.)</td>
<td>1,800</td>
<td>4</td>
<td>2.0</td>
<td>66</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>2. Mitsubishi UFJ Financial Group (Japan)</td>
<td>1,778</td>
<td>5</td>
<td>1.5</td>
<td>66</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>3. JPMorgan Chase (U.S.)</td>
<td>1,752</td>
<td>6</td>
<td>2.5</td>
<td>71</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>4. Bank of China (China)</td>
<td>1,649</td>
<td>9</td>
<td>1.0</td>
<td>74</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>5. Royal Bank of Scotland (U.K.)</td>
<td>1,228</td>
<td>15</td>
<td>1.5</td>
<td>61</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>6. Mizuho Financial Group (Japan)</td>
<td>1,211</td>
<td>16</td>
<td>1.0</td>
<td>75</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>7. ING Bank (B.U.)</td>
<td>788</td>
<td>24</td>
<td>1.0</td>
<td>75</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>8. Goldman Sachs (U.S.)</td>
<td>661</td>
<td>29</td>
<td>1.5</td>
<td>53</td>
<td>5</td>
<td>42</td>
</tr>
<tr>
<td>9. Morgan Stanley (U.S.)</td>
<td>604</td>
<td>35</td>
<td>1.5</td>
<td>69</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>10. BBVA (B.U.)</td>
<td>583</td>
<td>36</td>
<td>1.0</td>
<td>57</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>11. Bank of New York Mellon (U.S.)</td>
<td>271</td>
<td>68</td>
<td>1.0</td>
<td>75</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td><strong>Domestic banks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. ICBC (China)</td>
<td>2,248</td>
<td>1</td>
<td>1.0</td>
<td>94</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2. Agricultural Bank of China (China)</td>
<td>1,744</td>
<td>7</td>
<td>1.0</td>
<td>97</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3. Crédit Agricole (B.U.)</td>
<td>1,707</td>
<td>8</td>
<td>1.0</td>
<td>89</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>4. Bank of America (U.S.)</td>
<td>1,526</td>
<td>12</td>
<td>1.5</td>
<td>86</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>5. Société Générale (B.U.)</td>
<td>1,235</td>
<td>14</td>
<td>1.0</td>
<td>82</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>6. Groupe BPCE (B.U.)</td>
<td>1,124</td>
<td>17</td>
<td>1.0</td>
<td>84</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>7. Sumitomo Mitsui Financial Group (Japan)</td>
<td>1,112</td>
<td>19</td>
<td>1.0</td>
<td>78</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>8. Wells Fargo &amp; Co (U.S.)</td>
<td>1,107</td>
<td>20</td>
<td>1.0</td>
<td>95</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9. UniCredit (B.U.)</td>
<td>846</td>
<td>22</td>
<td>1.0</td>
<td>82</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>10. State Street (U.S.)</td>
<td>176</td>
<td>96</td>
<td>1.0</td>
<td>79</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td><strong>Total G-SIBS</strong></td>
<td>1,180</td>
<td>66</td>
<td>11.0</td>
<td>11.0</td>
<td>9</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 4: Global systemically important banks (G-SIBs), end-2013
Notes: The 2014 update of the list of G-SIBs is used [FSB (2014)]. The second column presents the assets rank on the basis of the Top 1000 World Banks, as published in The Banker (2014). Total assets are measured at end 2013. Segmentation of assets over the home country, the rest of region and the rest of world is calculated by the author based on annual reports. Total of G-SIBs is calculated as a weighted average (weighted according to assets). Source: Updated from Schoenmaker (2013)
cases, the European Commission has raised concerns where market shares for SME financing were 30% or higher [De Haan et al. (2015)]. Next, it is important to know the local business practices and market conditions to assess the credit risk of a small business. Because of these informational asymmetries, Degryse and Ongena (2004) argue that bank mergers and acquisitions seem the optimal route for entering another market for SME financing, long before cross-border servicing or direct entry are economically feasible.

4.5 Cross-border consolidation
Most trends point to cross-border consolidation. But it may take some time before we witness a truly cross-border merger wave, as in the U.S. after the deregulation in the 1990s. The majority of euro area countries still show subdued economic growth in the aftermath of the global financial crisis and the subsequent euro sovereign crisis. Low growth means that banks have not yet been able to strengthen their balance sheets.

Cross-border expansion may happen in different ways. One approach would be a full merger between banks from different countries. Another approach would be a cross-border acquisition. An expanding bank may first acquire a local bank in a neighboring B.U. country and then push more business through this local entity. A case in point is the acquisition of the German Direktbank by the Dutch ING bank. The renamed bank, ING DiBa, is now the third largest retail bank in Germany. A third way would be the cross-border supply of banking services, which can be easily done through Internet. An example is Wells Fargo, a U.S. bank, which entered the Canadian market with small business loans based on credit scoring models. Wells Fargo subsequently established branches in Canada to support its business there.

National authorities may, at least initially, still be reluctant to allow cross-border mergers or acquisitions. With the advance to B.U., the national supervisory authorities assess any proposed acquisition and forward a proposal to oppose, or not to oppose, an acquisition to the ECB. But the ECB has final decision-making power to grant permission for bank mergers within the B.U. from a supervisory point of view.\(^5\)

Next, supervisory authorities have become wary of large banks that are too-big-to-fail (TBTF). This implies that supervisors are reluctant to allow mergers that create mega-banks with over €2 trillion in assets. So, a merger among the top 11 banks in Table 2 is likely to be blocked. Nevertheless, one of the larger banks may take over a relatively smaller bank.

5. The international landscape
What is the impact of the B.U. on the international landscape? The Financial Stability Board [FSB (2014)] has updated the list of global systemically important banks, the so-called G-SIBs. These banks are the large financial players, which can pose a systemic threat to the global financial system [Bertay et al. (2013)]. Table 4 provides an overview of these G-SIBs, which have assets of up to €2 trillion. Remarkably, the B.U. encompasses most of the G-SIBs, with 9 out of the 30 institutions, followed by the U.S. with 8, the U.K. with 4, China and Japan with 3 each, Switzerland with 2 and Sweden with 1.

The ECB, as supervisor of the G-SIBs from the B.U., has thus become a major player together with the Federal Reserve and the Bank of England in international policymaking and supervision. It also confirms our earlier analysis in section 2 that the banking systems of the U.S. and the B.U. are not only similar in size, but are also home to a large number of global systemic banks. Nevertheless, the truly global banks (the first group in Table 4) are from the U.K. and Switzerland, with only one from the B.U. and the U.S. each.

6. Conclusions
This paper examines the new European banking landscape. It appears that the banking market of the B.U. is very similar to the U.S. banking market. In particular, both banking markets have a strong domestic orientation. Foreign banking penetration hovers around 15%. The new B.U. landscape is not very concentrated. The market share of the largest bank is only 5%, leaving scope for further consolidation within the B.U. In contrast, the previous national banking markets are more concentrated and do not leave much scope for domestic consolidation.

The central question in this paper is whether we can expect consolidation going forward. We find that the market structure and credit risk diversification are strong drivers for cross-border consolidation. But banking policies are not yet fully attuned to

---

\(^5\) Article 15 of the Single Supervisory Mechanism Regulation (1024/2013/EU).
the new B.U. setting. National supervisors still have some, albeit waning, influence. Moreover, deposit insurance is still national, which is important for consumers. Cultural barriers may also hamper cross-border consolidation.

Nevertheless, we expect an integrated banking market over time, similar to that of the U.S.. First, sooner or later it will become clear that the ECB is the dominant player in the new B.U., with national supervisors playing the role of local agents. Next, the state of banking is closely linked to that of the economy. A future pickup in economic growth may be the starting shot for a cross-border bank merger wave in the new B.U.

References
ECB, 2014, Banking structures report, Frankfurt
European Stability Mechanism, 2014, FAQ on the ESM direct bank recapitalisation instrument, Luxembourg: ESM, 8 December
Financial Stability Board, 2014, 2014 update of group of global systemically important banks (G-SIBs), FSB, Basel
The Banker, 2014, Top 1000 World Banks, London, July
Annex: before the B.U.
The Annex provides data at country and bank level before the start of the B.U. The geographical segmentation of assets is divided in assets from the home country, other E.U. countries and third countries. The three categories add up to 100%.

### Table A1: Cross-border banking penetration in E.U. member states, end of 2013

<table>
<thead>
<tr>
<th>Country</th>
<th>Number of banks</th>
<th>Total assets (€b)</th>
<th>Of which: home (%)</th>
<th>Other E.U. (%)</th>
<th>Third country (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>731</td>
<td>915</td>
<td>76</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>Belgium</td>
<td>103</td>
<td>1,021</td>
<td>34</td>
<td>51</td>
<td>15</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>30</td>
<td>47</td>
<td>28</td>
<td>69</td>
<td>3</td>
</tr>
<tr>
<td>Croatia</td>
<td>32</td>
<td>57</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Cyprus</td>
<td>101</td>
<td>90</td>
<td>71</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>56</td>
<td>191</td>
<td>1</td>
<td>93</td>
<td>1</td>
</tr>
<tr>
<td>Denmark</td>
<td>161</td>
<td>1,047</td>
<td>81</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>Estonia</td>
<td>31</td>
<td>20</td>
<td>2</td>
<td>92</td>
<td>6</td>
</tr>
<tr>
<td>Finland</td>
<td>303</td>
<td>522</td>
<td>34</td>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td>France</td>
<td>623</td>
<td>7,565</td>
<td>91</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Germany</td>
<td>1,842</td>
<td>7,525</td>
<td>89</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Greece</td>
<td>40</td>
<td>407</td>
<td>97</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Hungary</td>
<td>189</td>
<td>111</td>
<td>48</td>
<td>48</td>
<td>4</td>
</tr>
<tr>
<td>Ireland</td>
<td>458</td>
<td>972</td>
<td>61</td>
<td>29</td>
<td>10</td>
</tr>
<tr>
<td>Italy</td>
<td>694</td>
<td>4,039</td>
<td>87</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Latvia</td>
<td>63</td>
<td>29</td>
<td>40</td>
<td>45</td>
<td>15</td>
</tr>
<tr>
<td>Lithuania</td>
<td>91</td>
<td>24</td>
<td>27</td>
<td>73</td>
<td>0</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>147</td>
<td>843</td>
<td>22</td>
<td>63</td>
<td>15</td>
</tr>
<tr>
<td>Malta</td>
<td>27</td>
<td>50</td>
<td>62</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>Netherlands</td>
<td>253</td>
<td>2,250</td>
<td>92</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Poland</td>
<td>691</td>
<td>362</td>
<td>34</td>
<td>59</td>
<td>7</td>
</tr>
<tr>
<td>Portugal</td>
<td>151</td>
<td>513</td>
<td>80</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>Romania</td>
<td>39</td>
<td>91</td>
<td>30</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>Slovakia</td>
<td>28</td>
<td>61</td>
<td>4</td>
<td>96</td>
<td>0</td>
</tr>
<tr>
<td>Slovenia</td>
<td>23</td>
<td>46</td>
<td>69</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>Spain</td>
<td>290</td>
<td>3,143</td>
<td>92</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Sweden</td>
<td>168</td>
<td>1,212</td>
<td>91</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>U.K.</td>
<td>358</td>
<td>8,889</td>
<td>56</td>
<td>16</td>
<td>28</td>
</tr>
<tr>
<td>Euro area</td>
<td>5,999</td>
<td>30,035</td>
<td>83</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>Non-euro area</td>
<td>1,724</td>
<td>12,008</td>
<td>60</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>E.U.-28</td>
<td>7,723</td>
<td>42,043</td>
<td>77</td>
<td>15</td>
<td>8</td>
</tr>
</tbody>
</table>

Notes: Share of business from domestic banks, share of business of banks from other E.U. countries and share of business of banks from third countries are measured as a percentage of the total banking assets in a country. Figures are for end 2013. Euro area, non-euro area, and E.U.-28 are calculated as a weighted average (weighted according to assets). The new division of euro area (19 member countries) and non-euro area (9 members) as of 1 January 2015 is taken.

Source: Author calculations based on ECB structural financial indicators.
Table A2: Top 30 banks in Europe in 2013

<table>
<thead>
<tr>
<th>Banking groups</th>
<th>Capital (in €b)</th>
<th>Total assets (in €b)</th>
<th>Of which: home (%)</th>
<th>Other EU (%)</th>
<th>Third country (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Global banks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. HSBC (U.K.)</td>
<td>115</td>
<td>1,937</td>
<td>37</td>
<td>11</td>
<td>52</td>
</tr>
<tr>
<td>2. Barclays (U.K.)</td>
<td>67</td>
<td>1,568</td>
<td>36</td>
<td>26</td>
<td>38</td>
</tr>
<tr>
<td>3. Deutsche Bank (Germany)</td>
<td>51</td>
<td>1,612</td>
<td>28</td>
<td>30</td>
<td>42</td>
</tr>
<tr>
<td>4. Credit Suisse (Switzerland)</td>
<td>37</td>
<td>710</td>
<td>23</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>5. UBS (Switzerland)</td>
<td>35</td>
<td>821</td>
<td>33</td>
<td>25</td>
<td>42</td>
</tr>
<tr>
<td>6. Standard Chartered (U.K.)</td>
<td>31</td>
<td>489</td>
<td>16</td>
<td>5</td>
<td>79</td>
</tr>
<tr>
<td><strong>European banks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. BNP Paribas (France)</td>
<td>72</td>
<td>1,800</td>
<td>34</td>
<td>44</td>
<td>22</td>
</tr>
<tr>
<td>2. Santander (Spain)</td>
<td>61</td>
<td>1,116</td>
<td>29</td>
<td>41</td>
<td>30</td>
</tr>
<tr>
<td>3. UniCredit (Italy)</td>
<td>43</td>
<td>846</td>
<td>40</td>
<td>59</td>
<td>1</td>
</tr>
<tr>
<td>4. ING (Netherlands)</td>
<td>38</td>
<td>788</td>
<td>38</td>
<td>48</td>
<td>14</td>
</tr>
<tr>
<td>5. Nordea (Sweden)</td>
<td>24</td>
<td>631</td>
<td>24</td>
<td>75</td>
<td>1</td>
</tr>
<tr>
<td>6. Danske Bank (Denmark)</td>
<td>22</td>
<td>432</td>
<td>48</td>
<td>51</td>
<td>1</td>
</tr>
<tr>
<td><strong>Semi-international banks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Royal Bank of Scotland (U.K.)</td>
<td>60</td>
<td>1,228</td>
<td>61</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>2. BBVA (Spain)</td>
<td>40</td>
<td>583</td>
<td>51</td>
<td>9</td>
<td>40</td>
</tr>
<tr>
<td>3. Commerzbank (Germany)</td>
<td>26</td>
<td>550</td>
<td>51</td>
<td>33</td>
<td>16</td>
</tr>
<tr>
<td>4. DNB Group (Norway)</td>
<td>16</td>
<td>285</td>
<td>74</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>5. KBC (Belgium)</td>
<td>14</td>
<td>241</td>
<td>53</td>
<td>43</td>
<td>3</td>
</tr>
<tr>
<td>6. SEB Bank (Sweden)</td>
<td>12</td>
<td>280</td>
<td>63</td>
<td>34</td>
<td>3</td>
</tr>
<tr>
<td><strong>Domestic banks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Crédit Agricole (France)</td>
<td>63</td>
<td>1,707</td>
<td>81</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>2. Groupe BPCE (France)</td>
<td>47</td>
<td>1,124</td>
<td>77</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>3. Lloyds Banking Group (U.K.)</td>
<td>46</td>
<td>1,012</td>
<td>82</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>4. Société Générale (France)</td>
<td>41</td>
<td>1,235</td>
<td>76</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>5. Rabobank (Netherlands)</td>
<td>35</td>
<td>674</td>
<td>76</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>6. Intesa Sanpaolo (Italy)</td>
<td>34</td>
<td>626</td>
<td>86</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>7. Crédit Mutuel (France)</td>
<td>30</td>
<td>659</td>
<td>84</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>8. La Caixa Group (Spain)</td>
<td>18</td>
<td>351</td>
<td>91</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>9. ABIN AMRO (Netherlands)</td>
<td>17</td>
<td>372</td>
<td>80</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>10. Landesbank Baden-Würt (Germany)</td>
<td>15</td>
<td>274</td>
<td>75</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>11. DZ Bank (Germany)</td>
<td>14</td>
<td>387</td>
<td>75</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>12. Bayerische Landesbank (Germany)</td>
<td>14</td>
<td>256</td>
<td>75</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td><strong>Top 30 European banks</strong></td>
<td>1,135</td>
<td>24,592</td>
<td>53</td>
<td>24</td>
<td>23</td>
</tr>
</tbody>
</table>

Notes: Top 30 banks are selected on the basis of capital strength (Tier-1 capital as published in The Banker). Assets are divided over the home country, the rest of Europe and the rest of the world. Banks are divided in four categories. Global banks: less than 50% of assets in the home country and the majority of their international assets in the rest of the world. European banks: less than 50% of assets in the home country and the majority of their international assets in the rest of Europe. Semi-international banks: between 50% and 75% of assets in the home country. Domestic banks: 75% or more of assets in the home country. Source: Schoenmaker (2013)
About EY
EY is a global leader in assurance, tax, transaction and advisory services. The insights and quality services we deliver help build trust and confidence in the capital markets and in economies the world over. We develop outstanding leaders who team to deliver on our promises to all of our stakeholders. In so doing, we play a critical role in building a better working world for our people, for our clients and for our communities.

EY refers to the global organization, and may refer to one or more, of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. For more information about our organization, please visit ey.com.

© 2015 EYGM Limited.
All Rights Reserved.

EYG no. CQ0244

In line with EY’s commitment to minimize its impact on the environment, this document has been printed on paper with a high recycled content.

This material has been prepared for general informational purposes only and is not intended to be relied upon as accounting, tax, or other professional advice. Please refer to your advisors for specific advice.

ey.com

The articles, information and reports (the articles) contained within The Journal are generic and represent the views and opinions of their authors. The articles produced by authors external to EY do not necessarily represent the views or opinions of EYGM Limited nor any other member of the global EY organization. The articles produced by EY contain general commentary and do not contain tailored specific advice and should not be regarded as comprehensive or sufficient for making decisions, nor should be used in place of professional advice. Accordingly, neither EYGM Limited nor any other member of the global EY organization accepts responsibility for loss arising from any action taken or not taken by those receiving The Journal.

The views of third parties set out in this publication are not necessarily the views of the global EY organization or its member firms. Moreover, they should be seen in the context of the time they were made.

Accredited by the American Economic Association

ISSN 2049-8640