Does managing risk tie your hands, or give you wings?
The Journal of Financial Perspectives aims to be 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.

fsinsights.ey.com/thejournal
06 Ostrom game theory applied to financial services bonuses and policy improvements
   Aidan Walsh, EY Tax Partner, EMEIA Financial Services, Ireland; member of the Irish Prime
   Minister's Banking and Treasury Group; Malcolm Brady, Dublin City University;
   Piotr Magnuszewski, Centre for Systems Solutions, Wroclaw; International Institute for
   Applied Systems Analysis, Laxenburg

21 Risk accountability and risk appetite: enhancing risk culture
   Patricia Jackson, SUERF Council Member; former EY EMEIA Senior Advisor and Risk
   Governance Lead

36 Are banks still special?
   Thomas F. Huertas, EY Global Risk Network Chairman

51 German banks in financial centres: how risky is their business?
   Cornelia Keri, Deutsche Bundesbank

68 Stress testing banks: whence and whither?
   Pavel Kapinos, PhD, Federal Reserve Bank of Dallas; Christopher Martin, Federal Deposit
   Insurance Corporation (FDIC); Oscar Mitnik, Inter-American Development Bank

88 How to capture macro-financial spillover effects in stress tests
   Heiko Hesse, Directorate General for Economic and Financial Affairs (DG ECFIN), Italy desk,
   European Commission; Strategy, Policy and Review Department, International Monetary Fund;
   Ferhan Salman, Middle East and Central Asia Department, International Monetary Fund (IMF);
   Christian Schmieder, Member of Secretariat, Financial Stability Board, Basel

103 Why small portfolios are preferable and how to choose them
   Francesco Cesarone, Department of Business Studies, Roma Tre University;
   Jacopo Moretti, Department of Methods and Models for Economics, Territory, and Finance,
   Sapienza University of Rome; Fabio Tardella, Head of the Master’s course Finance and
   Insurance, Department of Methods and Models for Economics, Territory, and Finance, Sapienza
   University of Rome

117 Ring-fencing cross-border banks: an effective supervisory response?
   Katia D’Hulster, Lead Financial Sector Specialist, East Asia and Pacific Region, World Bank;
   Inci Otker-Robe, Division Chief, Financial Analysis, Monetary and Capital Markets Department,
   International Monetary Fund (IMF)

136 The broken buck stops here: embracing sponsor support in money market fund reform
   Jill E. Fisch, Perry Golkin Professor of Law, University of Pennsylvania Law School

160 Good and bad risk - regulation and loan monitoring
   Norvold Instefjord, Reader in Finance, Essex Business School, University of Essex;
   Hiroyuki Nakata, Senior Lecturer in Finance, School of Management, University of Leicester
Risk taking is the essence of banking. Banks add value because they bear the credit, liquidity, interest rate, foreign-exchange and operating risks that individuals and institutions wish to avoid. This issue of the Journal of Financial Perspectives deals with risk – how banks measure and manage it, as well as how supervisors monitor and, if need be, admonish banks.

The first two articles deal with risk culture. As Patricia Jackson emphasizes in her article, a bank’s risk culture is rooted in its risk management framework and its risk appetite statement. These effectively set the overall script that the bank should follow. To ensure that the bank does so, roles need to be clearly defined, and the senior executives need to be accountable as individuals for the performance of the units for which they are responsible. This goes above all for the line business executives who constitute the first of the bank’s three lines of defense. Together with appropriate incentives and adequate transparency, this approach should foster the bank’s taking the right risks in the right amounts in the right way for the right price.

Although banks’ remuneration practices have been subject to extensive reforms, Walsh et al. argue that further steps are required in order to bring incentives into line with effective risk management. This article brings into play the importance of power within an organisation. If one business makes the rules for the whole firm, it may abuse that power to set transfer prices in a manner that benefits its business and improves the bonus paid to its executives.

Banks take risk in specific legal vehicles, and the risk of a particular liability depends on the risk of the vehicle’s assets as well as on the position of the liability within the creditor hierarchy. Theoretically, it is therefore possible to reduce the risk of a bank’s deposit by ring fencing the bank issuing it from the rest of the group and restricting the ring-fenced bank to less risky activities. In contrast, as the article by Cornelia Kerl explains, conducting risky activities in branches does little to protect the parent bank, as branches are an integral part of the parent bank. The failure of the branch would cause the bank as a whole to fail. Conducting the activities in separate legal vehicles may be more effective, as the article by D’Hulster and Otker-Robe explains. In practice, however, it is difficult to distinguish between “good” risk and “bad” risk, especially where supervisors as well as the bank itself are monitoring the risk of the bank’s assets.

The following articles deal with risk and regulation. Valuation is the foundation on which risk measurement and risk management are built, but valuation depends on how the asset is expected to generate value, and that in turn depends on the business model and the state of the business. This has wide-ranging implications for bank regulation and supervision, especially with respect to capital regulation and resolution reform. As the article by Instefjord and Nakata explains, supervisory monitoring may induce banks to cut back on their own monitoring efforts.
Stress testing attempts to ensure that a bank has enough actual capital in place now to absorb the losses that could arise in the future, if the economic environment were to deteriorate. The stress test has become the cutting edge of capital regulation: it is both forward looking and institution-specific. Moreover, it effectively sets capital requirements at a far higher level than the minimum required under the Basel Accord, for a bank must pass its stress test before it can pay dividends or make distributions to its shareholders. The articles by Kapinos et al. and by Hesse et al. outline how stress testing can be further improved.

Capital surcharges on systemically important institutions help offset the risk to the taxpayer that could arise, if these institutions are in fact “too big to fail (TBTF)”. Such surcharges potentially depress the return on equity that a TBTF institution can achieve. “Small” is therefore beautiful, as the article by Cesarone et al. illustrates with respect to insurance.

Sometimes firms find it in their interest to assume responsibility for certain risks, even though they are not legally required to do so. Over time, market participants may come to expect that they will do so. This can create systemic risk. If a firm does not act in the manner that the market expects, this will cause market participants to revise their expectation of how firms generally will react. That in turn may induce investors to shift assets, creating funding and liquidity pressures on various entities across the financial system. A case in point is the support that sponsors provide to money market mutual funds. The article by Fisch argues that such support should be explicit rather than implicit, and that sponsors should be required to maintain capital so that they can provide support, if the need to do so arises.

Finally, regulators do not only control the risk of banks. They can also pose a risk to banks. A case in point is central bank digital currencies. If central banks were to introduce such an instrument, it would not only replace cash. It could very well also displace deposits and quite possibly the banks that issue them. For details, see “Are banks still special?”

Thomas F. Huertas
EY Global Risk Network Chairman
Ostrom game theory applied to financial services bonuses and policy improvements

Aidan Walsh
EY Tax Partner, EMEIA Financial Services, Ireland; member of the Irish Prime Minister's Banking and Treasury Group

Malcolm Brady
Dublin City University

Piotr Magnuszewski
Centre for Systems Solutions, Wroclaw
International Institute for Applied Systems Analysis, Laxenburg
Ostrom game theory applied to financial services bonuses and policy improvements

Abstract
Elinor Ostrom’s work on sustaining common-pool resources (CPRs) has had a profound impact on the management of natural resources. It has had very little impact on the regulation of the financial services industry. The purpose of this article, which focuses in particular on bonus payments, is to argue that this neglect may result in a number of tools available to regulators, industry and the public being underexploited. Coercively-powered, external rule-setting and monitoring authorities appear to have insufficient ability on their own to sustain a natural resource system.

Using a forest sustainability game as a tool (a common technique in natural resource management flexed to include bonus payments) and undertaken in EY’s offices in Dublin, we argue that business students and finance professionals seem to behave very similarly to exploiters of natural resources. Given that degrees of self-governance, cooperation between exploiters, self-monitoring and self-sanctioning are considered critical to sustaining natural resources, this paper considers whether the same may be true of financial services. We use this framework to comment on recent work on modeling the effect of bonus caps and bonus deferral on bankers’ risk taking.

SPIEGEL: “Maybe sustainability is just against human nature?”
GOODALL: “I don't think so. Even chimps understand the concept. If a troop of chimps enters a fruit tree, they will only pick the fruits that are ripe and leave the others growing. That is sustainability.”

1 Interview with Jane Goodall, Spiegel Online. Bethge & Grolle (2015)
1. Introduction

Elinor Ostrom (1933–2012) won the 2009 Nobel Memorial prize in Economic Sciences for her work on the governance of sustainable resources, what she called common-pool resources (CPRs). CPRs are those artificial or natural limited resources that are not privately owned and are large enough to make exclusion of potential exploiters difficult, but not impossible, for example, fisheries, irrigation systems, forests, commonages and pasture lands. In her classic text, Governing the Commons,¹ she told how she came to study common-pool resources, initially water resources in southern California in the 1950s. She and her colleagues found, completely unexpectedly, that different localities in southern California had self-developed different institutional arrangements to cope with similar problems, but that the different arrangements all worked and continued to work over long periods of time, providing water to local residents, farms and businesses (what she called “appropriators”) in a sustainable way.

Ostrom spent decades expanding these basic insights into studies of thousands of different CPRs, and the strategies and rules adopted by appropriators. She found that appropriators, often without state involvement and without private ownership of the resources, could create rules that allowed for the long-lived sustainability of the CPR. She found examples of resources that were sustained for over a thousand years. The Tragedy of the Commons,² i.e., the depletion of shared resources through the self-interested actions of participants was not inevitable.

It is not too strong to say that her work has transformed our understanding of human cooperation.

Through this work on thousands of sustainable systems, Ostrom and the people that worked with her have discovered certain elements that seem to feature in all successful CPRs and certain elements that seem to result in problems for CPRs. The elements featuring in successful CPRs include cooperation among exploiters, self-governance, self-monitoring and self-sanctioning. Payoff or reward rules are an important element of every CPR.³

As yet, this type of thinking has not been widely applied to financial services, despite the prevalent concern about the sustainability of financial systems.

A significant feature of many financial services systems is the widespread use of bonus payments. The purpose of this article is to apply one of the tools that Ostrom developed – a multi-participant, multi-round sustainability game – to discover if the use of bonuses results in unsustainable exploitation of the natural system. Although the insights from such a game may not necessarily apply directly to financial systems, as different specific mechanisms can be at play, we believe that it is still worth considering and using the analogies that such games can produce.

2. Background to bankers’ bonuses

Bonuses perform a very useful function in the financial services industry. In fact, at one stage, large bonuses were perceived to be essential for a sustainable financial services industry. Philip Augar, an experienced banker, in his history of London's investment banks deplored the shift to fixed salary structures in the 1980s, quoting experienced bankers:⁴

"As a [banking] partnership salaries were kept very low and there was a high-variable element in compensation in form of bonuses and profit share. The mix was about 25% salary and 75% bonus although it varied year to year and that was the beauty of it. In a bad year there was lots of room to manoeuvre on costs. County NatWest reversed the ratio by raising salaries and cutting bonuses. Fixed costs skyrocketed and we were hostage to the downturn in revenues."

"Midland were uncomfortable with low salaries and high bonuses so they switched the mix. It moved the risk from the staff to the bank. It was a clear indication that they didn't understand the business."

However, by the 2000s, the academic focus on game theory⁵ and agency problems⁶ had flipped that perception.⁷ Bonuses were seen as a one-way bet for employees, with the person receiving the bonus getting a percentage of the reward and the

---

2. Hardin (1968)
5. Ostrom E. (2005, 92)
institution (or ultimately, the taxpayer) being stuck with the loss. This phenomenon is succinctly captured by the acronym IBGYBG (I’ll be gone, you’ll be gone [when the loss manifests itself]) as justification for making a financial transaction. More subtly, regulating bonuses is seen as a way of incentivizing individuals within financial institutions to behave sustainably; in other words, to internally motivate them to behave sustainably rather than externally monitoring every transaction — which is considered difficult if not impossible.8

It’s now taken for granted that bankers’ bonuses result inevitably in unsustainable risk taking. The EU Commission amended remuneration rules for bankers in its CRD IV/CRR rules, arguing: “In order to tackle excessive risk taking, the remuneration framework has been further strengthened with regard to the requirements for the relationship between the variable (or bonus) component of remuneration and the fixed component (or salary). For performance from 1 January 2014 onward, the variable component of the total remuneration shall not exceed 100% of the fixed component of the total remuneration of material risk takers. Exceptionally, and under certain conditions, shareholder can increase this maximum ratio to 200%.”9 In the US, under the Dodd-Frank Act, bonuses are to be deferred and can be clawed back in cases of future reversals of profits.

Recent research seems to support the view that bonus caps, despite being a relatively crude one-size-fits-all policy, may be effective in reducing risk-taking by bankers.10 We comment on this paper further below.

3. Some conclusions from Ostrom’s work

It is impossible in this short article to do justice to the extent of Ostrom’s work, and an article like this has leached out all the uplifting drama that flows from her stories of some of the world’s poorest farmers, fishermen and hunters creating and sustaining CPRs from the most unpromising environments and institutional arrangements. But, we can summarize some high-level insights. Ostrom and her colleagues found that successful CPRs had certain common features.11

These “design features” were identified as follows:

- Clearly defined boundaries: Individuals or households who have rights to withdraw resources from the CPR must be clearly defined, as must the boundaries of the CPR itself.

- Congruence between appropriation and provisions, rules and local conditions: Appropriation rules restricting time, place, technology and quantity of resource units are related to local conditions and to provision rules requiring labor, material and money.

- Collective-choice arrangements: Most individuals affected by the operational rules can participate in modifying the operational rules.

- Monitoring: Monitors, who actively audit CPR conditions and appropriator behavior, are accountable to the appropriators or are the appropriators.

- Graduated sanctions: Rules violation as revealed through monitoring is sanctioned, but in a gradual way so that it does not provoke resentment.

- Conflict-resolution mechanisms: In case of disagreement, fast and effective conflict resolution procedures are established so that only more serious conflicts end up in court.

- Recognition of rights to organize: The rights of appropriators to devise their own institutions are not challenged by external governmental authorities. As one fisher said: “We’ve fished here for a long time and we know what’s best for our ground. We know what it can take.”12

- Nest enterprises (for CPRs that are part of larger systems): Appropriation, provision, monitoring, enforcement, conflict resolution and governance activities are organized in multiple layers of nested enterprises.

These rules are still not well understood and well-meaning governmental intervention has often ended in tragedy even in the most sophisticated countries. For example, Ostrom noted “fisheries in Newfoundland that had been organized to cope with

---

8 Goodhart, Hartmann, Llewellyn, Rojas-Suarez & Weisbord (1998, 49)
9 EU Commission (2013)
10 Jokivuolle, Keppo & Yuan (2015)
11 Ostrom E. (1990, 90)
12 Ostrom E. (1990, 177)
many problems [by the fishers themselves] were destroyed by actions taken by the Canadian national Government."\textsuperscript{13} Even where the external rules are well-designed, external coercive rules can “crowd-out” intrinsic motivation for sustainable behavior.\textsuperscript{14}

The central point of all Ostrom’s work was the importance of the appropriators having a real role in governing their own CPRs.\textsuperscript{15} Sometimes, external parties were needed or were useful, but often they were destructive of institutions that were working or might have worked. For instance, appropriators will make rules that are appropriate for their environment whereas governmental authorities will tend to use one set of operational rules across different environments. As Ostrom et al. (1994, p. 237) noted: “The search for one best rule for solving all CPR problems is doomed to failure.” Additionally, the governmental bodies usually recommend quotas for the amount of fish that can be harvested from a CPR in a season. However, in her sample of coastal fisheries, Ostrom found that fishers themselves never used quotas for sustainability purposes, as preordained quotas may take too many fish in lean years and forgo opportunities to take more fish in bumper years.\textsuperscript{16}

Communication between appropriators is the first key step in creating a sustainable institutional framework. For example, appropriators in lab-based games “who use the opportunity to communicate to agree a joint investment strategy, who are able to communicate on a repeated basis or can back their agreements with an endogenously adopted sanctioning mechanism, are able to achieve significantly higher levels of net yield from the CPR” even when the individual incentives conflict.\textsuperscript{17}

Self-sanctioning by appropriators has been found to be an effective mechanism to support sustainability. “[G]roups that did opt to change their rules and impose a sanctioning system upon themselves were able to achieve the highest average net yield (93%) of all experimental groups.”\textsuperscript{18} The ability to self-sanction by the group themselves was particularly effective against those “who do not care what others think.”\textsuperscript{19}

Rule following, rule monitoring and an attitude fostering reciprocity among appropriators are also effective ways of encouraging sustainability. “Where individuals follow rules and engage in mutual monitoring, reinforcing institutional arrangements and individual strategies bolster one another so as to maintain enduring patterns of consistent, but not perfect, rule-following behavior.”\textsuperscript{20} But, on the other hand: “It also makes us pessimistic about the likelihood of self-organized improvements [where] many, but not all, individuals are willing to ... extend reciprocity to others, but lack the authority to create their own self-governing institutions.”\textsuperscript{21}

4. Can CPR thinking be applied to financial services?
While Ostrom's thinking has not been widely applied to financial services, scholars of banking have independently argued that many long-sustained, completely self-regulated financial systems have been created\textsuperscript{22}:

- In Free Banking in Britain, White (2008) explains how the unique Scottish banking system had evolved in the 17th, 18th and 19th centuries. While liability of shareholders for most of the period was unlimited, the ability to establish a bank and then to issue notes was largely unregulated. There was no central bank in this system, no regulator and no deposit insurance. The system was not foolproof; for example, the Ayr Bank crashed in 1772, bringing down 13 (of the 19) banks in Edinburgh with it (albeit these were the smaller private banks). However, the system still created its own, entirely voluntary, sustainability: the Bank of Scotland and the Royal Bank of Scotland, during that crisis, lent to other banks\textsuperscript{23} and even offered to accept the notes of the Ayr Bank. The Scottish banking system was exceptionally innovative: the overdraft in 1728, interest on deposit accounts by 1731 (long before England), branch banking by 1793 and the Union Bank started to publish its annual accounts in 1836. But, competition was also vigorous; profits were squeezed with the spread between deposits and loan interest rates reduced to 1% or below.
No bank or banks became dominant and attempts to create cartels to increase profits and dividends failed.\textsuperscript{24}

- Calomiris (1990) has shown that self-insurance among clubs of US banks in the 19th century was more successful than state-run schemes. Banks in Indiana had a deposit insurance scheme “based on the principles of self-regulation and unlimited mutual liability … The Indiana insurance system relied on bankers themselves to make and enforce laws and regulations.” All the Indiana banks in this club survived the panics of 1854 and 1857 (over half of nonmember Indiana banks failed).

- Also, interesting here is Beck’s review of the German model of deposit insurance.\textsuperscript{25} While there are a number of different deposit protection schemes in Germany, the focus of that paper was the deposit insurance scheme of the German private banks. This was established by the German Bank Association (GBA), a private association, in the 1970s: “privately managed and funded, and stands apart from governmental supervision.” While membership is voluntary, all members of the GBA must be in their scheme or one of the other schemes. Echoing Ostrom, Beck notes: “The almost complete [deposit protection] coverage therefore increases pressure on the member banks to monitor one another.” Beck writes that: “Members may be expelled from the scheme, especially for missing or wrong information … the common goals and professional ethics create a sense of community. Personal contacts facilitate moral suasion and pressure.” In other words, self-regulation can create a virtuous feedback loop for participants where good behavior is respected and rewarded, and bad behavior noticed and sanctioned (however subtly).

- Edward Stringham, in his recent impressive work, \textit{Private Governance}, has emphasized the continuity over time of such self-regulation: “Without the ability to rely on external courts, brokers transformed coffee houses into private clubs that created and enforced rules … The private club known as Jonathan’s Coffee House eventually became the London Stock Exchange, which adopted as its motto ‘My word is my bond’ … In modern times, PayPal and eBay and other payment processors and clearing houses also created order in markets and facilitate exchange … From the world’s first stock markets in the seventeenth century … to millions of credit card transactions governed by private rules today … privately produced rules are more common, effective and promising than most of us believed.”\textsuperscript{26}

And, even in the current environment, some regulators have seen the importance of self-governed behavior and of non-legislative, even tacit, factors. For example, the Dutch National Bank explicitly supervises culture and behavior in their financial institutions. This has become an influential model. Their insight is that: “DNB acts on the premise that all institutions are responsible for their own business operations and the accompanying behavior and culture.”\textsuperscript{27} However, they do not seem to recognize that each individual also has some responsibility toward the sustainability of the system.

\textbf{5. Using games}

Understanding and analyzing complex policy issues is often hampered by the high cost of gathering data about how various members of society actually think and decide about these issues. Similarly, scientists and policymakers must often invest years to gain enough experience to be able to manage systems without risking them. Role-playing games have been successfully used as a way to simulate how people address complex resource decisions, such as sharing water for irrigation in Africa,\textsuperscript{28} farming and subsidies in North America\textsuperscript{29} and land use change around national parks in Poland.\textsuperscript{30} They can address complex concepts, such as the water-food-energy nexus\textsuperscript{31} or ecosystem services,\textsuperscript{32} in an understandable and comprehensive way. Role-playing games are highly flexible and leave room for the demonstration of individual initiative and imagination,\textsuperscript{33} which is an advantage in games involving policymaking. They are already used to test social responses and inspire social policy in relation to challenges like oil or water shortages.\textsuperscript{34}

Games have increasingly become a standard experimental method in economics research.

\textsuperscript{24} White (2008, 22-34) 
\textsuperscript{25} Beck (2002) 
\textsuperscript{26} Stringham (2015, 3, 4) 
\textsuperscript{27} De Nederlandsche Bank (2015) 
\textsuperscript{28} Barreteau, Bousquet, & Attonaty (2001) 
\textsuperscript{29} Taft (1998) 
\textsuperscript{30} Krolikowska et al. (2007) 
\textsuperscript{31} Mochizuki et al. (2017) 
\textsuperscript{32} Costanza et al. (2014) 
\textsuperscript{33} Ladousse (1987); Stefanska et al. (2011) 
\textsuperscript{34} McGonigal (2011)
Most of the existing experiments on CPR governance follow the methodology of experimental economics. Those experiments have been conducted both in the lab and in the field, and have established the use of games as a research method.

Although the method is not ideal because people may behave differently in game settings than in real world settings, it can provide an additional source of insight supplementing the current debate about the reform of financial markets institutions. While games can appear to be completely artificial, it seems that they are not totally divorced from how people behave in real life, as Bowles and Gintis point out: “Of course, experimental results in the laboratory would not be very interesting if they did not reflect real-life behavior. There is some evidence that they do or, to put it behaviorally, the experimental results have external validity.”

Although games and simulations have already been used as a tool for policy improvement, such applications are still rare and have never been considered, as far as we are aware, for financial markets. Their potential has been established to enhance a deeper analysis of the root causes of existing problems and to support policy formulation that could lead to governance innovations.

6. Experiments using simulation games

With the use of gaming, the largely ideological question of the regulation of financial markets and its potential consequences on market participants, their actions and overall equilibrium could become more empirical. Games and simulations could help to uncover the potential unintended consequences of proposed policies and regulations. Therefore, the authors of this paper undertook to organize a game using a dynamic “forest game” – an expanded version of the standard CPR games used for lab and field experiments.

6.1 Description of simulation game

Forest@Risk is a simulation game played by teams of players. Teams may view themselves as being in cooperation or in competition with other teams. The game centers on a forest, which acts as a resource for the community as a whole, providing timber on the one hand and acting as a buffer against flooding on the other.

During the game, players can spend their “action points” (an equivalent to money, time and effort) on cutting trees or planting them – the effects on the state of the forest are immediately visible for all players. When all decisions have been made, the simulation rolls on to the end of the period and presents the outcome to the teams: their income and potential flood damage that also depends on the forest state.

The game exposes participants to the concept of the tragedy of the commons and to the difficult job of preventing excessive exploitation of common resources by self-interested entities in the absence of strong rules. It also exposes players to the difficulties that arise in self-regulation and in the introduction of external governance policies and sanctions.

6.2 Technical specifications

The simulation game – Forest@Risk – has been developed by the Centre for Systems Solutions based in Wroclaw, Poland. It is a web-based multi-player game, but can also be run in workshop format. Each role (forest user) can be represented by an individual player or a team of few players. They make their decisions using a web application with all decisions recorded in a central repository located on a server. The game engine on the server then rolls forward the impact of all decisions as well as simulating regeneration of the forest and natural destructive phenomena like flooding (also influenced by the numbers of trees and other factors). This provides the starting position for the subsequent round. Teams then continue to make rounds of decisions until the end of the game.

The simulation event took place in the Dublin offices of EY during one evening in April 2015. Fifteen teams took part, with each team comprising 4 members: a total of 60 people. Most of the teams comprised students from an MSc in Management program from a local university, partners and employees of EY Dublin, and one team of experienced bankers. Two of the authors acted as independent facilitators for the duration of the game.

36 Cardenas (2003); Rustagi, Engel, & Kosfeld (2010)
37 Bowles & Gintis (2011, 39)
38 Duke & Geurts (2004)
39 Walsh, Magnuszewski, & Słodka-Turner (2012)

---

40 MBA-type students are typical recruits for financial institutions. Michel & Wortham (2009, 109)
The game took two and a half hours in total to play. Ten rounds of decisions were made, each round taking approximately 10 minutes. Two “town hall”-type meetings of approximately 20 minutes each were also held during the course of the game.

6.3 Bonus rounds
To simulate something akin to a bankers’ bonus, two rounds provided for a monetary reward (these were the only monetary rewards in the game). In rounds 2 and 6, a prize of €50 was offered to the team that made the “most profit” in that round.

6.4 Town hall meetings
The teams were encouraged to communicate with each other during two town hall meetings, with one representative from each team at the meeting. These meetings took place immediately after the two “bonus” rounds. The town halls were mostly devoted to discussing policy changes. They also provided an opportunity for teams to make sense of what was happening and see who was “making trouble.” During the first meeting, the teams collectively decided to change policy to a hierarchically driven quota system (forest harvest limit) – for every one tree cut, a team had to plant one tree.

However, this did not achieve the desired result as the forest condition was already too poor for the restriction to have an effect. In the second meeting, they voted in an egalitarian policy (income sharing) with the intention of disincentivizing the free riders, as any extra logging benefits would be shared with everyone. The group representatives worked out a fairly detailed four point plan and voted democratically on each point. They agreed on income redistribution and a plant-and-yield policy, with no harvest, to allow time for the forest to regenerate. They also sought a change in the rules of the game – a typical bargaining action. These decisions resulted in some improvement as the forest started to regrow, but slowly.
7. Results of rounds where bonuses were offered

In some ways, the outcome from the offered monetary prize, the €50 bonus, was predictable: nearly all the trees were cut down. However, a more detailed analysis gives a more nuanced result. The first team performance bonus was allocated in round 2. A number of teams appeared to change their policy in this round. Of the 15 teams, six significantly increased the number of trees that they cut down. Four teams reduced the number of trees that they cut and five teams made little change to their policy. However, in total, the number of trees cut increased by 10% and the number of trees planted reduced by 36%. It would appear that the introduction of a performance-based bonus increased the level of exploitation of the resource, leading the teams collectively down the road to a later tragedy of the commons: denuding of the forest, no timber to sell and ever more disastrous flooding.

The second bonus was awarded in round six. The level of tree-cutting, while much less than in round two, as the forest had been greatly depleted by this stage in the game, increased 27% and the amount of planting reduced by 14%. It is evident from figure 1 that the two spikes in the level of harvesting occurred in the two rounds in which bonuses were offered; the increase in harvesting that occurred in rounds 8, 9 and 10 mirrored an increase in the forest resources. This again implies that the introduction of the performance bonus led to an increase in the exploitation of resources (see figure 2). Note that policy changes may be due to the introduction of the bonus, but could also be due to other forces, for example, a team's dissatisfaction with its results from previous rounds.

The participants individually completed questionnaires during the game. There were 37 questions in the main questionnaire, but only a small number were directly relevant to the bonus issue. Two separate sheets were circulated after each bonus round. These questions were as follows:

- Compared with previous rounds, the possibility of €50 made no difference to the way that we made decisions. (This was probed after both rounds in which money was offered.)

- Question: The team that won the prize went up in my estimation. (This was only asked after the second bonus round.)

Participants rated each question from 1 to 7 rating, where 1 was "strongly agree" and 7 was "strongly disagree."

<table>
<thead>
<tr>
<th>First round bonus question: Compared with previous rounds, the possibility of €50 made no difference to the way that we made decisions.</th>
<th>Strongly agree/ agree/ slightly agree</th>
<th>Neutral</th>
<th>Strongly disagree/ disagree/ mildly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 (i.e. made no difference)</td>
<td>1</td>
<td>33 (i.e. made a difference)</td>
<td></td>
</tr>
</tbody>
</table>

| Second round bonus question: Compared with previous rounds, the possibility of €50 made no difference to the way that we made decisions. | 31 (i.e. made no difference) | 4 | 23 (i.e. made a difference) |

| The team that won the prize went up in my estimation. (i.e. went up in estimation) | 17 | 12 | 28 (i.e. did not go up in estimation) |

Table 1

In addition, in both bonus rounds, participants were asked why the bonus made a difference. We can see the shift toward the bonus having a smaller impact on the second round when we look at the explanations that people gave when they answered the question.

It is no longer a surprise that people do not act in a narrow, self-interested way and, even when there are only virtual trees at stake, people had different motivations in making their decisions. They made different decisions – some of which gave rise to sustainable solutions and some of which did not.
In many ways, both sides of the banking sustainability debate can take comfort from the results. On the one hand, nearly all of the trees were cut down and the environment was unsustainably degraded. It would take the forest many years to recover. This seems to show that individuals will destroy a CPR unless an external party has some power to modify the individuals' natural behavior. It also shows the power of incentives and that badly designed incentives can result in adverse outcomes for all.

However, on the other hand, some of the participants showed themselves capable of resisting the temptation of a short-term reward and looked to a longer-term sustainable future. We can also see that the individuals communicated and tried to self-create a cooperative, sustainable system using rules that they created, and were evolving, to monitor and self-sanction each other. In the end, the number of trees was growing and participants were creating a more sustainable system. As the survey results and group discussions showed, the individuals were grappling with the social problem of how to control themselves and each other to create such a sustainable system. However, they were only in the process of doing this when the game ended.

Both interpretations are entirely consistent with Elinor Ostrom's work. Creating a sustainable system is a complex task and not every sustainable resource is actually sustained. However, Ostrom’s work makes us focus on the heterogeneity of the participants and the heterogeneity of their reactions to the same risks and rewards. We can see that different teams behaved in different ways, and that some were capable of producing sustainable behavior. This is an important point that we will now discuss.

<table>
<thead>
<tr>
<th>First or second bonus</th>
<th>Rating</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round 1</td>
<td>7</td>
<td>We decided against a long-term strategy</td>
</tr>
<tr>
<td>Round 1</td>
<td>7</td>
<td>We aimed to earn the highest income to gain the €50</td>
</tr>
<tr>
<td>Round 1</td>
<td>7</td>
<td>It’s real money … I assume regeneration is factored in</td>
</tr>
<tr>
<td>Round 1</td>
<td>7</td>
<td>We wanted beer money</td>
</tr>
<tr>
<td>Round 1</td>
<td>3</td>
<td>The team became more money focused</td>
</tr>
<tr>
<td>Round 1</td>
<td>6</td>
<td>The lads became more ruthless</td>
</tr>
<tr>
<td>Round 1</td>
<td>7</td>
<td>It pushed me more to make the money as it is real and could be spent in real life not over a computer</td>
</tr>
<tr>
<td>Round 1</td>
<td>7</td>
<td>We blindly harvested ...</td>
</tr>
<tr>
<td>Round 1</td>
<td>7</td>
<td>Incentives drive work ethic and productivity</td>
</tr>
<tr>
<td>Round 1</td>
<td>5</td>
<td>We wanted the money (a number of participants made similar comments)</td>
</tr>
<tr>
<td>Round 1</td>
<td>5</td>
<td>Team cared more about income points than making sustainable decisions</td>
</tr>
<tr>
<td>Round 1</td>
<td>7</td>
<td>More greedy</td>
</tr>
<tr>
<td>Round 1</td>
<td>7</td>
<td>More focused to make more money and not thinking about sustainability</td>
</tr>
<tr>
<td>Round 1</td>
<td>2</td>
<td>There was an emphasis within the group and the room in general toward profit as opposed to sustainability, possibly sacrificing overall chance of success for immediate financial gain</td>
</tr>
</tbody>
</table>

8. Application of the game results to Ostrom's insights

Table 2
9. Application of Ostrom's work to the issue of bonuses

As we have seen, Ostrom and her colleagues found that successful CPRs had certain common features. How well does the financial services system measure up (taking a small number of Ostrom's features and focusing on bonuses)?

- Clearly defined boundaries: Financial services is now a global system, one that has shifted quite quickly from very local, especially in-country, systems and there are many new players at the edges of the system (for example, shadow banking and FinTech). Therefore, we would expect to see a system that has been and remains under sustainability stress.

- Congruence between appropriation and provisions rules and local conditions: Instead, there is a quest for global rules and consistency among larger areas (for example, top-down bonus rules that apply right across all the different banking cultures in the EU).

- Collective-choice arrangements: Does the financial services industry itself still have the power to influence the rules under which it must operate? For example, in relation to bonuses where even participants in the industry who thought that the previous bonus culture had acted as a perverse incentive were not in favor of the top-down rigid solutions from the EU.\(^{41}\)

- Recognition of rights to organize: A poignant feature of the game was the rudimentary attempt by individuals to self-create rules to preserve the forest. As Ostrom would have expected, the individuals were willing to extend reciprocity to each other and to curtail their own behavior. This may be the critical missing feature of the current financial system: the extent to which financial institutions are willing and able come together to discuss and argue out common behaviors that are essential for a sustainable financial system. It is possibly even illegal for industry participants to come together to agree local market-specific bonus rules. This ability to self-govern has been a feature of financial systems in the past, but is now largely gone. This is, potentially, the real tragedy of the commons: “individuals are willing to extend reciprocity to others but lack the authority to create their own self-governing institutions.”\(^{42}\)

---

41 BBA (2015)
42 Ostrom, Gardner, & Walker (1994, 328)
Ostrom's work adds an extra dimension to the debate. To demonstrate this, we can consider a recent Bank of Finland discussion paper on bankers' bonuses. In this detailed and impressive paper, the authors modeled the behavior of bankers' risk-taking as it was affected by bonuses. They concluded: “We calibrate the model to US banking data and show that lengthening the standard one-year bonus payment interval has no material impact, whereas capping the bonus at the level of the base salary substantially reduces the bankers risk-taking. Our results suggest that the EU's bonus cap reduces risk taking, whereas bonus clawbacks as prescribed in the Dodd-Frank Act appear to be ineffective.”

This seems like reasonably definitive support for the one-size-fits-all, prescriptive, externally enforced EU bonus cap. However, Ostrom's work would provide an additional insight – in the paper, the authors model bankers' behavior as homogeneous, that all bankers are “risk neutral,” and that their reaction to risk and to the bonuses was also homogeneous. However, as Ostrom shows individuals are not homogeneous. This heterogeneity is not just in their risk appetite, but, even where their risk appetite may be the same, in the reaction to that risk and their reaction to the potential reward. Our game can be seen as a very limited illustration of that point.

This has been borne out in real life with, for example, Gillian Tett noting that even in the most recent financial crises that some bankers were not swayed by large bonuses and were concerned about the sustainability of the system. On the other hand, in the modern regulated environment they had no ability, incentive or mechanism, no institutional framework, to act on these concerns; no ability to usefully monitor or to sanction. One noted plaintively: “it wasn’t our job to stop other banks being so stupid.” But, implicit in that statement is the fact that the individual did not act stupidly, the bank he or she worked for did not act stupidly. Both forewent bonuses and profits in the short-term, making the bank more robust and sustainable in the long run. We can also see it in the game where 28 (49%) of the questionnaire respondents disagreed with the statement: “the team that won the (€50) prize went up in my estimation.”

10. Conclusion

This was a limited study, which lacked many important design features of larger studies (for example, control groups, multiple runs of the games, diversity of participants and so on). While the results are consistent with the types of results that Ostrom and her colleagues, collaborators and followers have seen over countless such games, much further work will need to be done. Our research can be considered to be a pilot for further study, which will continue.

It is not yet robust, but it may be capable of producing richer insights than purely formal modeling. Our primary conclusion is methodological. We have sought to demonstrate that games may be a useful tool for industry participants, regulators and the public to learn about possible consequences of rules, to test different rules and to teach the benefits of new or different rules. To date, as far as we are aware, tools like this have not been used in the financial services industry in this way.

---

43 Jokivuolle, Keppo, & Yuan (2015)
44 Jokivuolle, Keppo, & Yuan (2015, 10)
45 Tett (2010, 249)
Elinor Ostrom and her husband, the political philosopher Vincent Ostrom, argued for polycentric solutions to problems – layered, different, overlapping, even competitive – and for polycentric ways of investigating and testing these problems. This pilot should be seen in that spirit.

Work on the sustainability of the financial services industry is decades behind work on the sustainability of CPRs. The financial services industry can be compared with the fishing industry at the time of this statement of the Canadian Government (1976): “in an open-access, free-for-all fishery, competing fishermen try to catch all the fish available to them, regardless of the consequences. Unless they are checked, the usual consequence is the collapse of the fishery: that is, resource extinction in the commercial sense, repeating in the fishery context the Tragedy of the Commons.”

Ostrom spent her entire career unpicking this misapprehension, although it is still widely held.

Is relying exclusively on external coercive rules optimal? Stringham has pointed out: “responsible people realize they must constrain their own choices. Yet, of all areas of governance, individual self-governance is the least studied and recognized.”

We are caught, ironically, in an unsustainable dynamic, if we see the regulators of the financial services industry as “wise and ecologically aware altruists” and the participants in the industry as “myopic, self-interested and ecologically unaware hedonists.”

---

46 Ostrom E. (2005, 280)
47 Ostrom E., Governing the Commons (1990, 242)
48 Stringham (2015, 145, 147)
49 Ostrom E. (1990, 218)
References

and Consequences. Retrieved from Regents of the University of Minnesota: http://www.extension.umn.edu/AgLand


Risk accountability and risk appetite: enhancing risk culture

Patricia Jackson
SUERF Council Member, former EY EMEIA Senior Advisor and Risk Governance Lead
Abstract
The magnitude of risk taking that was revealed by the financial crisis has led to increased attention being placed on risk culture and governance of financial institutions. The “three lines of defense” model is often used in risk management, with the second line (independent controls) usually being emphasized. However, it is the first line of defense, the business line, where the emphasis should lie, as it is closest to financial risks and it is where leaders can influence risk taking behaviors, including conduct. A framework needs to be developed within banks, which supports clarity in both the business line and control functions, especially with regards to the risk appetite of the board.

This article looks at the changes that are needed in terms of accountability frameworks, and the role of risk appetite within firms and how it can be used to delineate the accountability of the business line and leaders. There will also be discussion of changes that are needed to risk metrics, and how risk transparency is necessary in order to use the risk management model effectively.
1. Introduction and background

The magnitude of risk taking uncovered by the financial crisis, combined with the post-crisis discovery of a range of conduct issues, from London Interbank Offered Rate (LIBOR) and foreign exchange (FX) to breaches of anti-money laundering and sanctions rules to abetting tax evasion, has turned the spotlight on risk culture in banks and the changes that are needed for governance to go forward. In their 2014 paper on risk culture, the Financial Stability Board (FSB)\(^1\) set out their definition of risk culture as “the norms, attitudes and behaviors related to risk awareness, risk taking and risk management.” This article focuses on the role of risk accountability combined with risk appetite as one essential part in the delivery of a stronger and more effective risk culture, and sets out the expectations from regulators in this area. It builds on the thinking in the economics literature on the use of precommitment to keep risk taking within bounds.\(^2\) It also looks at the practical steps needed to embed risk accountability down through a bank.

The so-called “three lines of defense” model is often implemented in risk management. Under this model, the business line is the first line of defense, the independent control functions (such as risk) are the second line, and the third line is the internal audit function examining how both the first and second lines operate.

Within a bank, those in the business line are closest to both the financial risks being taken and to behavior (which could affect operational costs or reputation risk). The leaders in the business line, both senior and middle management, are also best placed to influence behavior. Although there will be a bank-wide framework for remuneration and promotion, the line management has a strong influence on the outcomes. In addition, and perhaps even more importantly, they control or have a strong influence on many other levers – informal local status and targets inside the business line as well as influencing negative incentives, such as individuals being dismissed for not meeting targets – all of which can have a powerful effect on individual behavior. Being closest to the business activities, the business line management will also be most likely to identify risks quickly. In theory, then, to achieve lasting change in risk culture, it is essential that accountability for these wider risks is firmly placed with this group.

Independent control functions (the second line of defense) also play an essential role. They need to review the way that the front office is managing risks, assess aggregate risk, set limits, and monitor performance against limits and other controls.

In practice, however, how the three lines of defense model has been implemented in almost all financial institutions is very different. The focus has been on the role of the second line in risk management rather than the first, where the prime responsibility is revenue generation. Although the business line may have notional accountability for risk, it has not played a central role in risk governance beyond responsibility to stay within limits set by the second line.

The FSB and different national regulators have now made clear that this shift in accountability to the business lines must happen. The regulatory expectations were set out in the FSB papers in 2013 and 2014.\(^3\) However, for the new approach for accountability to be effective and practical, it is essential that boards and top management go much further than setting tone and values at the top. They must ensure that individuals are clear about how much financial risk the board and top management are willing for the company to take, and what the expectations are regarding nonfinancial risks. To achieve this, the risk appetite framework within a bank must support clarity in both the business line and control functions regarding the amount of risk the board is comfortable with in that specific area. The business line heads can then be held to account for delivery of a risk profile that is not in excess of this level.

This article looks first at the change required in terms of accountability frameworks. It then examines the role of risk appetite in clearly delineating the accountability of the business line and the business line head. It then considers the parallels between this approach to risk appetite and precommitment frameworks in the economics literature. The final two sections look at the change required in risk appetite metrics and risk transparency to make the approach effective.

---

1. FSB (2014)
2. Kupiec and O’Brien (1997); Jackson (2014a)
Risk accountability and risk appetite: enhancing risk culture

The FSB clearly sets out their expectations regarding the ownership of risk in the front line units, as well as the roles of the second line functions and internal audit. There have already been rule changes to implement this approach by national authorities in a number of jurisdictions, including the US with the formal, codified “Heightened Standards” guidance of the Office of the Comptroller of the Currency.

The core elements of the FSB approach are that (1) the front line has responsibility for all the risks attached to the business activities in the areas for which individuals have revenue responsibilities, (2) there must be fully independent second line functions responsible for identifying, controlling, and monitoring aggregate risk, and (3) the third line, embodied by internal audit, must be able to express an opinion on the overall risk governance framework. This is a substantive shift from current risk governance practice in relation to all three elements. The front line currently has responsibilities regarding risks, but those are crystallized in terms of the limits they must adhere to (set by the second line control functions). The reason that this does not achieve full risk accountability in the front line is that limit structures, because of their nature, do not contain all the risks that are attached to front line activities. An example is the trading area. Many trading desks are set value-at-risk (VaR) limits to contain the risks and these must not be exceeded in most banks without prior approval. This provides a forward-looking brake on the amount of risk that can be taken through the aggregate trading positions. However, VaR does not include nonlinear risks that may be covered in part by separate micro-limits. But many other risks are not captured in these limits, in particular all the nonfinancial risks, such as conduct, which relate to the way that the business is transacted.

<table>
<thead>
<tr>
<th>First line</th>
<th>Ownership of risk by the first line. Day-to-day operating model</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Management of risks sits with units originating them</td>
<td></td>
</tr>
<tr>
<td>• The front line has responsibility for assessing and monitoring all risks, financial and nonfinancial, related to activities</td>
<td></td>
</tr>
<tr>
<td>• Frontline activities must be consistent with risk appetite</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second line</th>
<th>All risk types, financial and nonfinancial, covered by independent functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Any units responsible for identifying, measuring, controlling and monitoring aggregate risk must be independent and must not advise the front line</td>
<td></td>
</tr>
<tr>
<td>• They should identify and assess material aggregate risks and ensure that they are controlled consistently with risk appetite</td>
<td></td>
</tr>
<tr>
<td>• Compliance and legal functions should exert controls as well as the risk function</td>
<td></td>
</tr>
<tr>
<td>• Independent risk management should establish concentration limits</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Third line</th>
<th>Internal audit must exercise more judgment regarding risk governance framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Much enhanced role for internal audit</td>
<td></td>
</tr>
<tr>
<td>• Overall opinion on risk governance framework</td>
<td></td>
</tr>
<tr>
<td>• Audit should rate risk presented by each front line unit, product line and function, and evaluate compliance with risk management framework policies</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Responsibilities of the three lines of defense under FSB recommendations (FSB 2013a)
The FSB approach to accountability will also require a change for second line functions. The current approaches to risk governance put the second line control functions firmly in the frame for controlling risks. There has been considerable focus since the crisis on building up the seniority and independence of the risk function responsible for financial risk. Chief risk officers (CROs) now generally report to the CEO or jointly to the CEO and board risk committee.

6 Having a risk function that is fully independent of the business lines has been accepted as a necessity for a number of years. However, other second line functions are not equally independent of business decisions. The compliance functions, for example, usually advise the front line on the application of regulatory rules or laws, compromising full independence when it comes to their role in compliance monitoring and testing.

Similar issues arise with regard to legal risk and IT risk. Legal advises on contracts and even writes them, and IT creates the IT infrastructure. The FSB has asked for clearly independent second line functions. This has created the need to rework responsibilities for controls over nonfinancial risks.

The role of internal audit also needs to be changed. The third line internal audit function is fully independent but looks at governance in a piecemeal fashion. The focus is on whether individual units are meeting the laid-down processes and controls, and if these processes and controls are effective. The FSB wants internal audit to go beyond this remit to produce an opinion on whether the whole risk governance framework seems appropriate, which is much more judgmental.

Table 1 summarizes the expectations of the FSB regarding the different lines of defense.

In the article "Risk management formations – an alternative approach to the three lines of defense model," Martin et al.7 explore the role of different lines of defense using the following terminology to describe the different roles of risk taking and risk management:

- **Take** - the individual or function takes risk
- **Help** - the individual or function helps others to take risk
- **Stop** - the individual or function stops others taking excessive or the wrong risk
- **Keep** - the individual or function keeps score regarding the amount of risk that has been taken

These terms can be used to look at the current and proposed accountability frameworks (see Tables 2 and 3).

This creates the need for a considerable reworking of not only the business lines, but also the control functions. For control functions, such as compliance and legal, this requires a splitting of the functions. The areas that provide advice (and which need to continue to provide advice) are being split from the true control function - giving a line 1A (a function in the first line providing advice) and a line 2 (a separate control oversight function in the second line). Alternatively, some banks are considering splitting the functions into two second line functions - 2A and 2B. The control functions (other than risk) are also not set up to look at forward-looking aggregate risk. For example, they do not examine whether compliance risks are building up because of a change in the organization of the business or business activity, which has increased intrinsic risks or weakened controls. This goes beyond risk and control assessments into identification of risk factors – see section 7, "Risk transparency and risk accountability", below.

Banks are starting to create structures, which ensure that those in the business line are accountable (not just in name but in actuality) for all the risk attached to their revenue stream, but this requires rethinking the capabilities in the front line. The business line cannot be held to account for assessing individual risks if they do not have people who understand the

---

6 EY/IIF (2014)
7 Martin et al. (2015)
Risk accountability and risk appetite: enhancing risk culture

This requires risk training as well as the establishment of control functions in the front line, which report to the head of the business line.

The UK’s Financial Conduct Authority (FCA), as part of their review of the treatment of benchmarks, highlighted the importance of front-office engagement in assessing the risks. They said “we found instances of inadequate or poor monitoring by the first line of defence at a number of firms”.8 They give as examples: desks not having oversight over alterations to submissions by traders, very limited front-office review and inconsistent monitoring of data.9

This change must not result in a weakening in the power of the second line functions. They are still responsible for ensuring that risks in aggregate are consistent with risk appetite, which means that they must set limits and review control structures in the first line, and they must review significant transactions and changes in strategy and products. Questions have been raised whether the front line can be held to account, if the second line can overrule their proposals. However, the business line can still be held responsible for the proposed transaction, even if it is not finally approved.

### Table 3: Future accountability frameworks

<table>
<thead>
<tr>
<th></th>
<th>Take</th>
<th>Help</th>
<th>Stop</th>
<th>Keep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business lines</td>
<td>In taking risks, individuals need to consider if they are excessive or the wrong risks to take</td>
<td></td>
<td>Business line heads/managers will have a responsibility regarding identifying and stopping excessive or wrong risks</td>
<td>Business line heads/managers will have to consider if individual risks are appropriate relative to the board-set risk appetite</td>
</tr>
<tr>
<td>Control functions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk</td>
<td></td>
<td></td>
<td>Although the business line has accountability for the risks taken, the risk function must continue to look at the adequacy of controls, set limits, review sizeable transactions, and stop excessive or wrong risks where they see them</td>
<td>The risk function is responsible for assessing aggregate forward-looking risk relative to risk appetite</td>
</tr>
<tr>
<td>Compliance</td>
<td></td>
<td>The independent control function cannot provide advice</td>
<td></td>
<td>The control function must consider adequate forward-looking risk relative to risk appetite</td>
</tr>
<tr>
<td>Legal</td>
<td>The independent control function cannot take risk</td>
<td>As above</td>
<td></td>
<td>As above</td>
</tr>
</tbody>
</table>

3. How will risk accountability help to improve risk culture?

In the economics literature, it is recognized that there is a cascade of principal agent problems in companies. These affect firms in general, but particularly financial services, because of the potential of individuals to take decisions and translate these into actions that can impact the risk profile and reputation of the organization in a significant way. There is a principal agent problem between the shareholders and the management and also between different layers of management. To quote Kern Alexander: “the opportunity exists for some managers to improve their economic payoffs by engaging in unobserved, socially costly behavior or ‘abuse’.”10 Alexander is focusing on conduct risks, but precrisis there were many examples of decisions taken by individuals, which led to an excessive buildup in financial risk and which could not always be seen at the board or senior management level. A particular example is the removal of the liquidity lines from the structured investment vehicles in London discussed below.11

The authorities and boards have been focusing considerable efforts on redesigning financial remuneration arrangements in order to provide the right incentives for individuals. There

---

8 FCA (2015b, 19)
9 FCA (2015b)
10 Alexander (2004, 6)
11 Jackson (2014b)
is a move towards risk-based compensation, including using balanced scorecards and, as a fail-safe, longer vesting periods for bonuses, with the possibility of clawing back even vested bonuses if poor behavior comes to light later. However, changes to remuneration do not provide the full answer. Research in the nuclear power industry has shown that reward comes in many forms. Paradies shows that financial reward does play a role in influencing behavior as does the desire to save time (cutting corners), but the views of colleagues can outweigh the other influences. He finds that the effect of an immediate boost to standing in the eyes of colleagues may outweigh a future uncertain effect on a bonus. Informal standing of an individual in their business unit is therefore very important, allied to their standing in their cross-firm peer group. The FCA has referred to the tribal loyalties that cross firms. The important message regarding accountability is that line management is much better placed than second line control functions to influence the local standing of an individual. If a trader breaches the controls, but makes a large profit, are they criticized or applauded? If the business lines are to be held to account for risks, including conduct, then negative behavior is more likely to be criticized.

The effect of wider incentive structures, such as targets, set by line management can also have a powerful effect on behavior. Cass Business School and New City Agenda, in their joint report on the culture of retail banking, cite examples from several banks where low-paid staff were encouraged to meet sales targets through negative incentives - in one case, tellers in a branch were given cabbages or cauliflowers, which had to stay on their desk until they opened a new customer account; in another bank, each week, staff exceeding sales targets were given a cash bonus and those falling short were given a cabbage. They conclude that this widespread sales culture “led banks to make risky loans and engage in bad practices.”

These examples demonstrate that it is essential that managers in the business line are accountable for the wider risks, including behavior, that can lead to reputation damage and financial penalties. It is not enough to rely on centrally determined compensation structures to influence behavior. The conduct failures are becoming too expensive for the industry in monetary and reputation terms, and thus need to be dealt with in a fundamental way. In the UK alone, by 2015, costs due to remediation, litigation and fines related to conduct issues had reached £39bn over the period since the global financial crisis (GFC). CCP Research Foundation has calculated that globally, the cost of conduct events (from 2010-14) for a group of 16 international banks reached £200bn (CCP, 2015). This includes regulatory fines, remediation and other costs.

4. Risk accountability and risk appetite

To make accountability practical, it is essential that the business line heads know the acceptable limits on risk taking. This requires banks to go much further in the direction of cascading risk appetite across risk types and business lines.

- A business line, across all activities in the line, needs to be clear about how much risk the board and senior management are willing for them to take - capturing all risk, not just risks covered by explicit limits
- They also need to understand expectations regarding their responsibility for the behavior of those in the line
- The risk, and accountability for the risk, needs to be forward -looking – it can be too late once losses or behavior problems have occurred
- The accountability needs to cascade down the front line to ensure that business leaders and middle management set the right tone for those below

The adoption of new and more stringent individual accountability regimes in some jurisdictions, such as the Senior Managers Regime in the UK, underline the importance of clearly defined limits to the type and amount of risk taking that is acceptable to senior management and the board. Without this, individuals are managing risk in the dark. This needs to be made specific to the business line and needs to be more holistic than simply a set of limits - the onus should be on the business leaders and the second line control functions to demonstrate how the controls and limit structures in place
## Risk Accountability and Risk Appetite: Enhancing Risk Culture

<table>
<thead>
<tr>
<th>Role in Risk Appetite Framework (RAF)</th>
<th>CEO</th>
<th>CRO</th>
<th>CFO</th>
<th>Business Line Leaders and Legal Entity Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Establish a prudent risk appetite for the firm, which is consistent with the firm's short- and long-term strategy, business and capital plans, risk capacity, as well as compensation programs, and aligns with supervisory expectations.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Be accountable, together with the CRO, CFO and business lines, for the integrity of the RAF, including the timely identification and escalation of breaches in risk limits and of material risk exposures.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ensure that the “risk appetite is appropriately translated into risk limits for business lines and legal entities, and that business lines and legal entities incorporate risk appetite into their strategic and financial planning, decision-making processes, and compensation decisions.”</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ensure that the firm-wide risk appetite statement is implemented by senior management through consistent risk appetite statements or specific risk limits for business lines and legal entities.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide leadership in communicating risk appetite to internal and external stakeholders so as to help embed prudent risk taking into the firm's risk culture.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set the proper tone and example by empowering and supporting the CRO and CFO in their responsibilities, and effectively incorporating risk appetite into their decision-making processes.</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Immediately escalate to the board and CEO any material risk limit breach that could seriously put the financial condition of the firm in danger.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Act in a timely manner to ensure effective management, and where necessary mitigation, of material risk exposures, in particular those that are close to or exceed the approved risk appetite and/or risk limits.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Incorporate risk appetite into the firm's compensation and decision-making processes, including business planning, new products, mergers and acquisitions, and risk assessment and capital management processes.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Independently monitor business line and legal entity risk limits and the firm's aggregate risk profile to ensure they remain consistent with the firm's risk appetite.</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Establish and approve appropriate risk limits for business lines and legal entities that are prudent and consistent with the firm's risk appetite statement.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Ensure the integrity of risk measurement techniques and MIS that are used to monitor the firm's risk profile relative to its risk appetite.</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establish a process for reporting on risk and alignment (or otherwise) of risk appetite and risk profile with the firm's risk culture.</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Actively monitor the firm's risk profile relative to its risk appetite, strategy, business and capital plans, risk capacity, as well as compensation programs.</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
collectively keep risks within the overall risk appetite for the business line.

The FSB has made clear that the business line heads do have responsibility for ensuring the business activities are aligned to risk appetite. Table 4 sets out the accountabilities the FSB expects regarding risk appetite for the CEO, CFO, CRO and business line heads.

### 5. Risk appetite as a precommitment mechanism

In terms of accountability, risk appetite can be seen as a way of establishing a precommitment approach throughout a bank. There was interest in the late 1990s, led by the Federal Reserve, in the possible use of precommitment approaches to align incentives with risk taking. This was seen as a way to overcome the challenges of setting market risk capital requirements for trading books. Kupiec and O’Brien suggested that banks could pre-commit to the regulators a maximum loss that they would make and would face penalties if losses rose above this level. Goodhart et al. raised the idea of extending the concept from market risk to the banking book and counterparty risk.

For a bank, the risk appetite is a way of achieving precommitment by the bank’s senior management that risk levels will be contained within a set amount. The overarching risk appetite statement, if included in the annual report, is a precommitment by the board to

---

18 Kupiec and O’Brien (1997)  
19 Goodhart et al. (2013)
shareholders. By cascading the risk appetite down into business lines and sub-business lines, the management lower down is also committing to top management and the board that risk will not exceed preset levels.

This has implications for the design of the risk appetite statement. The precommitment approach was not pursued in banking regulation because it was backward looking. Penalties could not be effectively imposed on the bank once substantial losses had been made. Fines are not really practical at that point and even imposition of higher capital requirements, looked at by Kupiek and O’Brien, might not be practical. Goodhart et al.\(^\text{20}\) said, “in order for ex-ante incentives against … excessive risk taking … to work effectively, it is essential that banks perceive the (ex-post) imposition of penalties to be credible.” The backward nature of the test created a further problem, identified by Goodhart et al., which was that it could not prevent a firm from gambling for resurrection if the franchise had been eroded.

For risk appetite to be an effective precommitment approach within a bank, it is essential that individual accountability includes estimates of risk (i.e., forward looking) and not just the imposition of penalties ex-post if losses arise.

6. Risk appetite metrics and accountability
Risk appetite statements have been set for a number of years, but unfortunately, they have proved difficult to embed in the business lines.\(^\text{21}\) The statements were initially subjective, using qualitative statements, which made true accountability difficult. Subsequently, many banks added a range of quantitative elements, but without incorporating an overall connecting metric - some banks had 15, 20, and even more different metrics, which made it difficult to check consistency let alone accountability.

In the last few years, there has been a shift towards a smaller number of metrics. Additionally, it is also important to have accountability against a true forward-looking metric that picks up the risk in extreme environments, given the cyclicality of the banking industry. This then provides a double incentive with regard to individual accountability. An individual would be held to account regarding the forward risk metric as well as backward-looking against a loss made, through deferred bonuses, clawback, etc.

Banks are starting to move towards the use of forward loss in an extreme environment as a key element in the central risk appetite statement for financial risk. Table 5 analyzes the information provided by globally systemically important banks (GSIBs) in the EY/IIF\(^\text{22}\) survey on risk governance. Almost all the GSIBs sampled have a minimum acceptable tier 1 capital ratio as part of the primary risk appetite metric. Almost all also have a forward extreme loss metric - loss in extreme events, enterprise VaR, or stress test results. 15 of the 18 banks have stress test results in the risk appetite statement. This is consistent with also having a minimum tier 1 ratio because the maximum loss set would have to be consistent with that minimum capital ratio. The advantage of the forward loss figure is that it can be allocated to regions, risk types and business lines. Judgments have to be made about diversification benefits, but that is tractable.

To date, although forward loss is part of many banks’ primary risk appetite at group level, the next step has not been taken to allocate the loss to individual business lines. Once this allocation has taken place, region heads and business line heads can be held accountable for ensuring the strategy and business decisions do not take the risk of forward loss above the figure set for their business line.

Clarity around this will support real accountability. The region or business line head can then be made responsible for ensuring that limits, controls and strategies keep risk within that set amount. The means of testing this would have to be agreed upon, for instance, by using the ICAAP stress testing, covering a stress equivalent in severity to the last crisis. The scenario used would be open to debate internally, but because the ICAAP stress testing is subject to supervisory oversight, there is an element of external accountability. Boards, too, are required to scrutinize the ICAAP stress testing. In some jurisdictions, there are also regulatory stress tests (US, EU, UK) and these too could form part of the test.

However, it is noticeable from looking at the GSIB metrics from the survey that banks still have a large number of metrics in their primary risk appetite statements (with a number covering similar territory) which will confuse accountability. The average number of metrics is close to 10 per bank. In addition, some metrics could encourage risk-taking, for example ROE and growth measures.

\(^{20}\) Goodhart et al. (2013)

\(^{21}\) Shifting focus: Risk culture at the forefront of banking, EY/IIF (2014)

\(^{22}\) EY/IIF (2014)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital ratios</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Tier 1 ratio</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Stress test results</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Earnings at risk</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Loss in extreme</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterprise-wide</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VaR</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operational losses</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Funding and liquidity</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illiquid investment</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>levels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROE</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings volatility</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic capital</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>RAROC</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RWA</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of risk</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal ratings</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected loss</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operating leverage</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VaR</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tail VaR</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limits</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provisions</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concentration limits</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth measures</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5: Primary risk appetite metrics used by GSIBs**
Source: Data collected for the EY/IIF Risk Management Survey, 2014
Risk accountability and risk appetite: enhancing risk culture

The Journal of Financial Perspectives

used by a minority of the banks.

In the past, many banks have used risk-weighted assets (RWAs) as a core risk appetite metric and 11 banks in the sample continue to use RWA. However, RWA has disadvantages as a risk measure, particularly because assets are multiplied by the pillar 1 risk weights, which assume asset diversification. This measure, therefore, does not reflect a firm’s risk concentrations. In addition, the pillar 1 risk weights do not take into account a range of risks, such as interest rate risk in the banking book and pension risk. These should be captured in the regulatory pillar 2 capital requirements, but are not included in the RWAs. If RWA were the main metric, this gap would affect accountability. Risk concentrations would, in contrast, be picked up by the stress testing results, of course, depending on the shape of the stress test.

The various conduct breaches that have come to light since the crisis (LIBOR, FX, money laundering controls, tax, sanctions and retail product mis-selling) underline the importance of nonfinancial risks being covered by the risk appetite framework and the cascade of responsibilities. This is clearly one of the FSB’s objectives. One way to achieve this is through maximum operational risk losses being set (again for extreme environments). This is the path that many banks are taking. The operational loss figure can then be allocated to business lines. However, the operational risk pool is an amalgam of a large number of different components – conduct, fraud, IT, legal and so on. It would probably be better to treat each separately. This would make it easier to focus on the intrinsic risks and the controls to keep losses in check. There would then be more focus on behavior levers. However, estimating forward loss for each is very difficult and thus another route to expression of risk appetite is needed.

Considering the events that could cause severe difficulties for the bank, and then setting either specific actions that are expected ex ante or risk tolerances (for example, downtime for primary systems), and holding management to account for those, might be more fruitful than just setting an operational loss figure. For example, in the area of market manipulation, senior management could make clear that the heads of the trading lines have two requirements: (1) ensuring that an analysis has been done regarding whether the markets and transactions in their area are vulnerable to manipulation, and (2) in light of that analysis, an assessment of whether the controls are adequate. Again, this would result in responsibilities being cascaded down the chain of command within a business line. This is far more likely to be successful in achieving real accountability than, for example, just saying the business line heads are responsible for the effectiveness of, say, 7,000 end-to-end controls or that the business line head should keep operational losses within GBP 500m (approximately USD 708m).

Subsidiaries of global banks are increasingly being required to tighten risk governance structures. The creation of an appropriate risk appetite framework for that entity and appropriate accountability structures plays an important role.

7. Risk transparency and risk accountability

A very important part of the implementation of risk accountability through the risk appetite framework is risk transparency. Without a view of risk, which fully reflects the downside, the risk appetite framework (and the precommitment it is achieving) will be undermined. The framework clearly requires a forward-looking analysis of aggregate risk in extreme periods, including identifying the risks to assess. One problem in the run-up to the crisis was that severe risks were not fully identified, in part because plausibility filters had been applied at the business unit level – risk was taken and indeed acknowledged, but the crystallization was deemed to be implausible and, therefore, the risk was not debated at a bank-wide level nor stress tested. An example is the structured investment vehicles (SIVs) in London. These were off-balance sheet vehicles that invested in structured products and were funded through the issuance of three-month commercial paper. In the traditional US design, the vehicles had prearranged undrawn bank funding lines to be utilized if sufficient commercial paper could not be issued. To increase profits, those designing the vehicles in London took the decision not to include the lines. The risk was fully understood, but as a drying-up of the commercial paper market was deemed implausible, it had not been debated at the top of the banks nor was it included in stress testing. This led to massive funding needing to be injected by bank sponsors; the alternative was risking the reputational damage caused through the failure of SIVs.23 This problem of a

23 Jackson (2014b)
24 Morcroft (2007)
Another factor is the tendency to concentrate on headline risks rather than those buried in the technicalities of particular instruments. There were many examples of this in the run-up to the GFC. One example is that the recession stress testing carried out precrisis did not include the default rate triggers on the pools in the securitization funding vehicles. This points to the need to develop better judgment tools to identify business fault lines at a business division level, which some banks are now utilizing. These can then inform the holistic bank-wide stress tests.

Monitoring also has to be developed to ensure that risk does not start to move away from risk appetite. These monitoring metrics also need to be carefully designed. Currently, monitoring nonfinancial risk types is relatively underdeveloped. The focus is on controls and control self-assessments, and reporting on the outcomes, as well as reporting on a wide range of different metrics, such as customer complaints. More focus is needed on forward risk, and here a paradigm shift is needed to move to assessing risks and not just relying on controls, under the assumption that they will prevent the risk being taken on. Consistent analytic frameworks to assess intrinsic risk in the nonfinancial risk types would also help to support accountability.

To take one example, product suitability risk, banks have introduced more rigorous product approval processes and escalation rules, some of which are backed up by hard analytic criteria. This process should sift out the most problematic products. But banks also need metrics at the top of the organization so that the pockets of higher risk can be identified as well as the trends in the risk profile across the bank – for example, are products generally becoming more complex? The first step is the identification of intrinsic factors for the risk. In this example of product suitability, the risk is driven by a combination of the complexity and opacity of the products being sold, volatility of returns, and the sophistication (or lack of sophistication) of the investors. Some banks are starting to develop scoring approaches so that the trends in product risk can be tracked. Risk appetite could be set by the maximum score for suitability risk that the bank is willing to see in a business line or product. The scores could also be used in the escalation processes to more senior committees. It is essential though that this is a common analytic framework, which requires judgment rather than being a “tick box” exercise. A scoring approach would enable changes in the product characteristics, delivery channels, incentives or client base during the life of a product to be picked up.

8. Conclusion

Risk accountability has an essential role to play in the strengthening of risk culture and thus risk governance. Currently, too much of the responsibility for the amount of financial and nonfinancial risk (such as conduct) de facto sits with the control functions rather than those taking the risks. This goes right up to board-level thinking and reactions. When a severe problem occurs, very often just the head of risk or head of compliance is fired. This does not put accountability in the right place. The control functions can police the business units, but are in a much weaker position to influence day-to-day actions and behavior. Banks have tried to deal with this by emphasizing remuneration structures and balanced scorecards to reflect wider cultural behavior. But this does not recognize the other ways that individuals receive reward, in particular, from their colleagues in the business line in terms of informal status. It also does not recognize the range of business line-driven incentives (positive and negative) outside the remuneration framework. It is the business line management that is best placed to influence behavior and is closest to the financial and nonfinancial risks being taken.

The FSB has made it clear that the business areas must be responsible for all the risks attached to their activities. Some national regulators have gone further and have tightened accountability of designated individuals. However, both moves raise the question of the standard regarding acceptable levels and types of risk against which an individual would be held accountable. Here, risk appetite can be used as a mechanism to deliver a clear precommitment by management in the business lines, regions and legal entities that risks will not go above pre-set levels, but this will require an evolution of the risk appetite framework in many banks. A move to forward loss in extreme environments as a key accountability metric would supplement the existing responsibility for not breaching a range of preset limits. This could include risk appetite and measured forward looking using, for instance, stress testing. Indeed, the business

plausibility filter being applied is akin to “disaster myopia”. 25

25 Gutten-tag and Herring (1986)
line would have the responsibility for ensuring that controls and limits would deliver risks not in excess of the risk appetite. Non-financial risks pose different challenges, but also need to be clearly within the accountability framework. Here, clarity over acceptable tolerances and expected actions would help.
References


Guttentag, Jack M. and Herring, Richard J. (1986). Disaster Myopia in International Banking, Essays In International Finance No 164, September, Princeton University


Lopez del Olmo, Jose Luis (2014). In Jackson, Patricia (ed.), Risk Culture and effective risk governance, London: Risk Books


Are banks still special?

Thomas F. Huertas¹
EY Global Risk Network Chairman

¹ The author is a Partner in EY’s Financial Services Risk practice and chairs the organization’s Global Regulatory Network. The views expressed here are personal.
Abstract
Banks have long played a special role in the financial system. Individuals and institutions use banks to access the payment system, and central banks rely on banks to transmit monetary policy to the real economy. Hence, financial and economic stability has rested on the stability of the banking system, in particular, on the safety and soundness of systemically important banks. This is the primary reason why banks have access to central banking lending facilities, as well as why banks are regulated and supervised. It has also served as the overriding rationale for the reform of regulation and resolution that the G20 initiated and implemented in the wake of the financial crisis of 2007–08.

But, banks are not inherently special. Banks are only as special as central banks make them. Via quantitative easing (QE) as well as eligibility easing (EE), central banks have broadened the transmission mechanism beyond banks. As a result, banks have become less special. This in turn has significant implications for central banks’ responsibilities for liquidity provision, and for the regulation and supervision of financial institutions.

Banks’ special role could erode further if central banks introduce central bank digital currencies. Such an innovation would not only replace cash, but could also displace deposits. Central banks could not only impose significantly negative rates of interest; they could potentially determine the volume, distribution and pricing of credit, so that the transmission mechanism becomes direct. That in turn would have significant and not necessarily positive implications for banks, for financial markets and for the economy at large.
1. The “traditional” financial system

Traditionally, the financial system has operated on a hub-and-spoke basis (see Figure 1). In this framework, banks are special.\(^2\) Nonbanks (individuals and institutions) have their accounts at banks; banks have accounts at the central bank. Individuals and institutions therefore use banks to access the payment system. Banks lend to nonbanks and provide the economy at large with a liquidity backstop.

In this framework, financial stability depends largely on ensuring the stability of the banking system and that, in turn, hinges heavily on maintaining the continuity of banks’ critical economic functions.

Central banks align their functions to this framework. They restrict access to their payment systems to banks. They transmit monetary policy to the economy at large via banks. They provide liquidity to banks, and they play a significant role in their regulation and supervision. In sum, central banks act as bankers’ banks.

2. A new framework for finance is emerging

However, the crisis has, together with technology and central banks’ concerns about their own risk management, eroded the basis for this traditional framework. A new framework is emerging, one in which central banks are banks, not only to banks, but to financial institutions generally and, to an increasing extent, to the economy at large.

2.1 The transmission mechanism has become multi-channel

To be effective, central banks need a transmission mechanism or an ability to translate policy decisions into marketplace reality. Two aspects are important: first, what the mechanism does to the economy as a whole (i.e., how it affects output, employment and inflation) and second, what the mechanism does to the market(s) in which the central bank chooses to intervene.

To impact the economy as a whole, the central bank needs to send a strong and steady signal. Changes in the signal should result from policy decisions rather than market static or noise. In designing its signal, the central bank has to decide with whom it will deal, how it will transact and what assets it will buy or refinance.

However, with the asset(s) in which it chooses to intervene, the central bank will be no ordinary market participant. Given the size of its portfolio as well as its ability to determine the eligibility of assets as collateral, the central bank is likely to have a dominant position in any asset in which it chooses to invest or to accept as security. Prices of such assets are likely to reflect not only the risk of the borrower, but the degree to which the central bank’s position “overhangs” the market for particular issues and the possibility that the central bank will change the level of its holdings, the pace of its purchases or the haircuts that it applies. In other words, for such assets the central bank will be a price maker and its actions will confer (or constrain) liquidity on such assets.

Traditionally, central banks have principally used banks to transmit monetary policy to the economy at large. The policy rate set by central banks has either been the rate at which the central bank lends to banks\(^3\) or the rate at which banks can borrow central bank money in the market.\(^4\)

---

\(^2\) Corrigan (1982). Separately, various scholars (Gande and Saunders (2012)) have argued that banks are special, due to the monitoring that they provide in connection with loans. However, other financiers (notably private equity firms) also monitor firms in which they invest or to whom they extend credit.

\(^3\) For example, the Bank of England employs bank rate to set sterling interest rates.

\(^4\) In the US the Federal Reserve (Fed) targets the federal funds rate.
Central banks have largely executed monetary policy by conducting open market operations with banks or by varying the amount that they lend directly to banks. This alters the level of reserves at banks, with knock-on effects on banks’ capacity to lend to individuals and institutions. That in turn drives changes in output, employment and prices.

Yet, banks (as well as other financial intermediaries) had practically no place in the macroeconomic models that central banks used to determine monetary policy. Indeed, monetary economics generally regarded the banking system as neutral: it was merely a transmission mechanism without any effects on the real economy.\(^5\)

However, during and after the financial crisis of 2008, this transmission mechanism broke down. Banks failed, credit contracted and the real economy went into a tailspin. Although government intervention rescued the banks, it did not fully restore the transmission mechanism. Nor has regulatory reform. Although such reform has greatly strengthened banks’ condition, credit growth has been weak and the recovery sluggish, particularly in the Eurozone.

Central banks have responded by resorting to extraordinary measures. They have suppressed interest rates to zero and, in some cases, pushed them below zero. They have also initiated a program of QE and implemented a policy that could be labeled as EE (see below).

QE has helped to keep interest rates at very low levels for a very long period of time. QE has allowed the central bank to set the risk-free rate as well as the term structure of such rates (if it buys securities across the entire yield curve). This has in turn helped the economy get on and stay on the recovery path.

But, QE has also diluted the distinction between monetary and fiscal policy. This is perhaps most clearly the case in the Eurozone, where the European Central Bank (ECB)’s purchases of government bonds have contributed to a narrowing of spreads between the more highly indebted, deficit-prone, lower-rated “peripheral” member states and the less indebted, higher-rated member states.

The narrowing in spreads has reduced the government’s interest expense, creating in turn the potential to sustain government spending on goods and services or to reduce the overall budget deficit. In effect, the ECB has given member states the opportunity to create a virtuous circle. Lower-rated governments can “invest” the savings in interest expense to improve their budgetary position so that spreads need not rebound after QE ends.

Under QE, central banks also expanded the range of assets acquired directly via open market operations. In the US, the Federal Reserve (the Fed) bought massive amounts of mortgage-backed securities.\(^6\) In the Eurozone, the ECB has started to purchase corporate bonds as part of its supplemental asset purchase program.\(^7\)

These central bank purchases have supported the price of such assets, enhanced their market liquidity and reduced the spread over the government rate of equivalent maturity that borrowers have to pay. As a result, the creditworthiness of institutions holding such assets improved.

Taken together, QE and EE proved remarkably effective in stabilizing the world economy. Indeed, in my view, it is one of the principal reasons that economists describe 2008 to 2010 as the Great Recession and not as the start of the Greater Depression.

2.2 Payment systems are becoming robust

Well-functioning payment systems help generate the confidence and trust on which any currency is ultimately based. They are critical to maintaining financial stability and generating economic growth. Indeed, if payment systems cease operation, so will financial markets and the economy at large. For this reason, ensuring the integrity and continuity of payment systems is a major public interest.\(^8\)

Traditionally, banks have acted as the front end of the payment system. Individuals and institutions make payments by ordering their bank to debit their deposit account and transfer funds to another person’s or entity’s account. If the beneficiary has its account at another bank, the payer’s bank utilizes a payment

---

\(^5\) Adrian & Shin (2008, 301); Huertas (2011a, 103); King (2012)

\(^6\) FRB (2017)

\(^7\) ECB (2017)

\(^8\) Committee on Payments and Securities Settlement (2001)
system to make a transfer to the payee’s bank. The payee’s bank then credits the beneficiary’s account.

For much of the 20th century, the failure of a major bank could cause the payment system to fail, with knock-on effects on other participants and on the economy at large. Banks were exposed to one another via multilateral netting arrangements, and central banks were exposed to banks via the daylight overdrafts that central banks routinely extended to banks.

Following the failure of Herstatt and Continental Illinois in 1984, central banks led a long and ultimately successful effort to make payment systems robust so that the payment system would be able to continue in operation, even if one or more of its major participants were to fail. This involved shifting central bank payment systems to real-time gross settlement and putting private, multilateral netting systems, such as the clearing house interbank payments system (CHIPS), on a sound basis.

In almost all jurisdictions, the central bank operates its own payment system to enable banks to transfer funds to one another. Settlement occurs via book transfers of “central bank money”: the deposit (reserve) account of the sending bank is debited and the deposit (reserve) account of the receiving bank is credited. Central banks generally guarantee the payments made over their systems so that receiving banks have no exposure to the sending bank if the sending bank were to fail. This provides certainty, immediacy and finality to both sending and receiving banks, and their respective clients.

The central bank itself does not take any risk on the sending bank. The systems that central banks own and operate are now generally on a real-time gross settlement (RTGS) basis, so that banks have to debit their reserve account at the central bank as soon as they initiate a payment instruction. Central banks have drastically curtailed the provision of daylight overdrafts to banks in connection with payment systems.

If a bank wants to initiate a payment, it has to have the money in its reserve account. Private payment systems have also become robust. They have practically eliminated any extension of credit to a sending bank and have instituted frequent intraday settlements of net exposures via special zero-balance accounts at the central bank. Together with the initiation of RTGS in central bank payment systems, the improvements in private payment systems have greatly reduced systemic risk.

Making payment systems robust has, in turn, served as the basis for making other financial market infrastructures robust. Directly or indirectly, central banks ensure that the “P” works as it should in delivery versus payment (DVP) for securities settlement systems and in payment versus payment (PVP) in foreign exchange settlement. In each case, the “settlement asset” is a claim on the central bank. This not only ensures certainty, immediacy and finality, but it also creates neutrality – each member uses the safest asset (a deposit at the central bank) to settle its obligation to other members of the settlement system.

2.3 Central banks are easing eligibility requirements

Central banks act as the ultimate provider of liquidity. This takes two forms: ordinary facilities and lender-of-last-resort (LOLR) facilities (also known as emergency liquidity assistance (ELA)).

Central banks extend ordinary facilities to eligible counterparties on the basis of eligible assets.

Conceptually, ELA encompasses (1) loans to eligible counterparties on the basis of ineligible assets, (2) loans to

9 However, significant operational risk remains particularly in connection with cybercrime and cyber terrorism (http://www.bankofengland.co.uk/education/Documents/ctcs/handbooks/pdf/cc kịchb31.pdf). The US Federal Reserve detected more than 50 cyber breaches between 2011 and 2015, many of which were suspected to have involved hackers or spies http://www.reuters.com/article/us-sri-lanka-gov-cyber-attack-idUSKBN10N310. In 2016, hackers stole US$81 million from the Bangladesh central bank account at the New York Fed via false orders on the SWIFT network http://www.reuters.com/article/us-bangladesh-heist-fbi-insight-idUSKCN00X28. For example, final settlement for CHIPS occurs via a similar zero-balance account at the US Federal Reserve Bank of New York and final settlement for the EURO 1 payment system operated by EBA Clearing occurs via a similar account at the ECB.

12 Comitoto (2011)

14 CLS offers FX settlement services. See also Kahn, Quinn, & Roberds (2014)
ineligible counterparties on the basis of eligible assets and (3) loans to ineligible counterparties on the basis of ineligible assets (see Figure 2).

Traditionally, banks have been the only counterparty eligible to access ordinary central bank facilities, and the only collateral eligible to pledge was high-quality securities and loans. During the crisis, however, central banks supplemented QE with EE. They expanded the collateral eligible to support normal central bank liquidity facilities\(^\text{15}\) and they extended the range of counterparties eligible to access normal central bank liquidity facilities.\(^\text{16}\) This reduced the need to resort to LOLR/ELA. It also reduced the special role of banks.

2.4 Supervision is broadening beyond banks

Banks’ special role has long served as the rationale for the regulation and supervision of banks as well as the rationale for the central bank’s involvement in such functions. In particular, central banks need to be sure that the banks it extends credit to are solvent (particularly, if the bank is making a request for ELA). In fact, the prudential standards that regulators impose on banks are akin to the covenants that banks themselves put into loan agreements with corporate borrowers.\(^\text{17}\) But, the supervision that central banks and other authorities exercise over banks is far stricter than the control that banks can exercise over borrowers via the monitoring and enforcement of covenants.

In the wake of the crisis, authorities strengthened regulation and sharpened supervision. To reduce the probability that banks would fail, Basel III increased capital requirements and introduced global liquidity standards. To make banks “safe to fail,” jurisdictions reformed their resolution regimes to enable the authorities to ensure that a failing bank’s critical economic functions could continue, even in the absence of taxpayer support. To enforce these tougher regulations, jurisdictions granted supervisors broader powers, especially to central banks.\(^\text{18}\)

The crisis and its aftermath confirmed that systemic risk could not be controlled simply by regulating and supervising banks. More was required and more has been accomplished, including strengthening the stability of the derivatives market, bringing shadow banking under control and introducing the concept of macro-prudential supervision under the aegis of systemic risk boards.\(^\text{19}\)

3. The choice facing central banks and society at large

In sum, the current financial system differs from the traditional one. The transmission mechanism is now multi-channel. Payments systems no longer depend on banks: they can continue to operate, even if a bank fails. Nor is access to payment systems restricted to banks. Under “open banking,” banks have to grant access to their systems to third-party providers.\(^\text{20}\) Central banks no longer extend credit or provide liquidity exclusively to banks. Finally, prudential supervision now extends beyond banks. Banks are no longer as special as they once were.

Should the authorities keep this current approach? If so, what measures should authorities take to improve it? If not, what are the alternatives? We consider two measures: one that is conceivable for the near future, namely, reverting to banks as the single transmission mechanism; and one that could be feasible in

\(^{15}\) Breeden & Whisker (2010)
\(^{16}\) BoE (2014)
\(^{17}\) Dewatripont & Tirole (1994)
\(^{18}\) For example, the UK dissolved the Financial Services Authority and returned responsibility for prudential supervision of banks to the Bank of England. In the Eurozone, the Member States established a Single Supervisory Mechanism under the aegis of the Single Supervisory Board at the ECB.
\(^{19}\) Financial Stability Board (2017); Llewellyn, Nieto, Huertas, & Enoch (2017)
\(^{20}\) In the EU, banks are required to grant such access from January 2018 under the terms of the Second Payments Services Directive.
In the not too distant future, namely, eliminating the intermediaries entirely and transmitting policy directly to financial markets and the economy at large.

3.1 Strengthening the current multi-channel framework
The case for retaining the multi-channel approach is strong. It is consistent with the hypothesis that the transmission mechanism works through total credit, not just bank credit or bank money. The multi-channel approach has also apparently worked in practice, at least as a means to arrest recession and to foster recovery.

The case for the multi-channel approach would be stronger still if the authorities aligned liquidity provision to the transmission mechanism. Now that the recovery is finally taking hold, central banks are struggling with the question of when and how to end QE.

A similar debate needs to take place with respect to EE. In particular, the debate should consider whether QE and EE should be regarded as temporary expedients or permanent macro-prudential tools. The bias must be in favor of the latter, especially if policymakers can exit from these programs without disrupting the recovery.

In my view, the authorities should consider transforming EE into a macro-prudential tool. In such an approach, as an upturn progressed, the central bank would narrow the range of assets eligible as collateral for ordinary central bank facilities and broaden the range as a downturn took hold. In effect, this would create a countercyclical “accordion” (see Figure 3). Note that the central bank could supplement these measures by varying the haircut on assets that do remain eligible as well as varying the term of its ordinary facilities.21

With respect to the counterparties eligible to access ordinary central bank facilities, it would seem sensible to align this to the transmission mechanism. As this has become multi-channel, so should central banks' liquidity provision. Eligible counterparties would, therefore, include not only broker-dealers, but also, potentially, all financial institutions.

A separate question concerns the counterparties to whom the central bank can extend ELA or act as a lender of last resort. At a minimum, this should include counterparties that have access to ordinary central bank facilities. Should the central bank have broader powers? This certainly proved useful in the crisis: the Fed made ample use of its extraordinary powers under Section 13(3) of the Federal Reserve Act, but the Dodd-Frank Act repealed this authority.

Some have argued that this should be reinstated, subject to certain safeguards.22

![Figure 3](image)

<table>
<thead>
<tr>
<th></th>
<th>Ambiguity</th>
<th>Certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicability</td>
<td>LOLR/ELA</td>
<td>Ordinary facilities</td>
</tr>
<tr>
<td>Basis</td>
<td>Privilege</td>
<td>Right</td>
</tr>
<tr>
<td>Precondition</td>
<td>Solvency check</td>
<td>Eligible collateral</td>
</tr>
<tr>
<td>Conditionality (in addition to normal regulation and supervision)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Commitment fee</td>
<td>No</td>
<td>Yes?</td>
</tr>
</tbody>
</table>

Table 1

---

21 Huertas (2011, 106-110)
22 Calomiris et al. (2017)
Finally, there is the question of the terms and conditions on which central banks should provide liquidity facilities. Here, central banks should draw a distinction between LOLR/ELA and ordinary liquidity facilities. In the former, ambiguity is essential: no one should have the right to receive LOLR/ELA (the review of EE will determine which counterparties have the right to request LOLR/ELA). In contrast, certainty is constructive with respect to ordinary central bank facilities: eligible counterparties should be able to expect that the central bank will provide liquidity upon the pledge of eligible collateral. This distinction is particularly important in connection with resolution (see table 1).

In sum, EE deserves as much attention as QE. It too played a significant role in containing the crisis. The time has now come to determine its future role, both as a macro-prudential tool and as a determinant of what really constitutes the “last” resort when it comes to LOLR/ELA.

4. Central bank decisions on liquidity can determine course of resolution

A central bank’s decisions can affect both the timing and outcome of resolution. If a bank asks for LOLR/ELA, this may be a sign that it has reached the point of non-viability (PONV). Correspondingly, the central bank must rapidly decide whether to grant or decline the request for LOLR/ELA. In making this decision, the central bank should first consider whether the bank requesting LOLR/ELA is solvent. If it is not, LOLR/ELA could amount to what under normal bankruptcy procedures might be judged a fraudulent conveyance. However, the central bank should also take into account that the decision itself (whether or not to grant LOLR/ELA) may have an influence on asset prices and hence on the solvency calculation.

The central bank’s decision also affects the bank’s continuity. Granting the bank’s request for LOLR/ELA permits the bank to continue to function, but it also creates the possibility that the authorities exercise forbearance. Indeed, without liquidity support from the central bank, the supervisor alone cannot (except where the bank finances itself exclusively with insured deposits) exercise forbearance.

Declining the request will almost certainly lead to the immediate failure of the bank and force the supervisor/resolution authority to put the bank into resolution, regardless of the views that these authorities may have regarding the bank’s condition.

Financial stability will be enhanced if markets know what will happen next. To this end, the central bank and resolution authority should make clear that they will pursue the following presumptive path: they will put the bank into resolution. They will immediately bail-in instruments qualifying as total loss absorbing capacity (TLAC) so that the bank is not only solvent, but also meets minimum requirements for CET1 capital. They will ensure that the bank in resolution retains access to financial market infrastructures. Finally, they will make ordinary central banking facilities available to the recapitalized bank on the basis of the bank’s unencumbered assets.

In particular, the central bank should stand ready to take over both the financing and the collateral from repo providers. It makes no sense for the central bank to create the impression that it will refuse to grant the recapitalized bank access to ordinary central bank facilities. Such a position would undermine practically any resolution plan that the resolution authority might devise.
4.1 Reverting to banks as the single transmission mechanism

In contrast, the case for attempting to reinstate banks as the single channel for the transmission mechanism is weak, particularly in the US. First, banks account for a diminishing share of total credit. The role of nonbanks, including so-called shadow banks, is growing. These entities do not necessarily depend on banks for their own financing, particularly during the upswing of the business cycle.

Second, banks and central banks have different perspectives. Banks (or at least the major banks) operate across many jurisdictions. A central bank must focus primarily, if not exclusively, on its own jurisdiction. Legislation sets the central bank's objectives (price stability, financial stability, and, in some cases, growth and full employment) in domestic terms.

As a consequence, focusing the transmission mechanism solely on banks is likely to be ineffective. Although the central bank can determine the risk-free rate (and therefore the appropriate floor price for credit), it cannot determine the volume of credit. In an upturn, borrowers can turn to nonbank sources of credit. If credit is fuelling the boom, restricting bank credit alone will not necessarily reduce the pace of expansion in overall credit to a sustainable rate.

Nor will focus on banks alone ensure that credit will flow at the trough of the business cycle in amounts sufficient to initiate and sustain the recovery. Banks lack the capacity to do so, and banks lack the incentive. In theory, macro-prudential supervisors would give banks the capacity to expand credit at the trough by rescinding the countercyclical capital surcharge that they had imposed prior to the peak. In practice, however, banks find that capital held to meet the surcharge is needed to meet the hurdle rate imposed under the stress test (or tests) that micro-prudential supervisors conduct.

Central banks, therefore, face a conflict between their macro-economic policy objectives and their supervisory responsibilities. Extending credit at the trough can facilitate the recovery. As borrowers spend the proceeds of the loan, this will stimulate output and employment. Enough lending can produce enough stimulus to jump-start the recovery.

But there is no guarantee that it will do so. At the trough of the cycle, borrowers find it prudent not to count on any improvement in economic conditions until the "green shoots" of recovery are well on their way to becoming sturdy plants. Until such time, borrowers are likely to be looking to strengthen their balance sheets and the higher-rated borrowers are more likely to be successful in doing so. Many of the borrowers who do not strengthen their balance sheets cannot. They lack the current income to do so and/or the prospective future income necessary to attract new equity. Accordingly, the demand for bank credit at the trough of the cycle is likely to come disproportionately from lower-rated borrowers - precisely the segment bank supervisors will be most concerned about. Indeed, at the trough of the cycle, supervisors are much more likely to be urging banks to deal with their nonperforming loans than to extend new credit to those who are more likely to have difficulty in paying it back.

In sum, for the near future, banks are likely to remain semi-special. Although they will remain the most important element in the transmission mechanism, they will by no means be the only one. Although they will be the primary interface with the payment system, the advent of open banking will further reduce the benefits banks derive from this position.

However, in the not too distant future, banks may not be special at all. A central bank could shift to direct transmission: they would interact directly with the individuals and institutions that constitute the "real" economy rather than indirectly via banks or financial institutions. This may not be feasible today, but could well be within the next five to 10 years.

How might central banks do this? One possible route is central bank digital currency (CBDC). This is essentially today's currency in digital form. In fact, a CBDC will be far superior to alternative digital currencies (ADCs), such as bitcoin. What ADCs lack, central banks have. ADCs do not function well as a store of value: their price is too volatile, their defenses against hacking are too weak and their backing is nonexistent. Central bank money is the quintessential store of value. What ADCs have (distributed ledger technology), central banks can and will acquire. This will enable central banks to issue CBDC.
A number of central banks are already considering how this might be done, including the Bank of England,23 the Sveriges Riksbank,24 the Norges Bank,25 the Bank of Canada,26 the Monetary Authority of Singapore,27 the European Central Bank,28 the Federal Reserve29 and the People’s Bank of China.30 Central banks are not only exploring how digitization might enhance the current role of central bank money in clearing and settlement infrastructures. They are also considering whether central banks could and should issue digital currency more broadly to the general public either as a token or in the form of an account at the central bank.31

<table>
<thead>
<tr>
<th></th>
<th>Transmission mechanism for monetary policy</th>
<th>Access to central bank payment system</th>
<th>Access to central bank liquidity facilities</th>
<th>Supervisory focus of central bank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Banks</td>
<td>Banks</td>
<td>Banks</td>
<td>Banks</td>
<td>Banks</td>
</tr>
<tr>
<td>Current Multi-channel</td>
<td>Banks via RTGS</td>
<td>All FI?</td>
<td>All systemic institutions Macro-prudential supervision</td>
<td></td>
</tr>
<tr>
<td>Near future Strengthen multi-channel</td>
<td>Greater indirect access for all via “open” banking</td>
<td>Align to transmission mechanism</td>
<td>Align to transmission mechanism</td>
<td></td>
</tr>
<tr>
<td>Not-so-distant future Direct</td>
<td>All direct via RTGS</td>
<td>All against eligible collateral</td>
<td>Solvency/capital check only</td>
<td></td>
</tr>
</tbody>
</table>

Prospectively, everyone would have access to a CBDC. Indeed, if governments begin to use CBDC to distribute benefits, collect taxes, pay interest on government bonds and pay suppliers, practically everyone will have to have a CBDC account at the central bank. Once an individual or institution has such an account, they can use it for other transactions, such as receiving their salary or paying their bills. They can also allow cash to accumulate in the account as an investment. For macroeconomists, a CBDC holds a certain charm.32 It would expand the policy tool kit: it would greatly increase the ability to employ negative interest rates33 and it would facilitate the distribution of “helicopter money”34 or the introduction of basic income.

For banks, however, a CBDC should hold considerably less charm. A CBDC would not only replace cash, it could also displace deposits. CBDC outstandings will be direct senior obligations of the central bank. They will effectively be backed by the full faith and credit of the government. They will not be subject to bail-in.

They will, therefore, have lower risk than bank deposits, particularly uninsured deposits.35 This could lead to a significant shift of funds from banks to CBDC accounts, particularly if the central bank pays interest on funds held in CBDC accounts – something the central bank is likely to want to be able to do.36 With the advent of CBDC, the central bank balance sheet could grow very large indeed. Over time, the central bank could well become the largest single source of credit to the economy as a whole as well as to specific institutions. This prospect intensifies the debate over the type of assets that the central bank should acquire, and over the impact that the central bank would have on the markets for such assets and financial markets overall.

First, the central bank will need to decide the framework for its liquidity facilities. If central banks continue to grant all account holders access to ordinary lending facilities, this would imply that anyone with a CBDC account could borrow from the central bank against a pledge of eligible collateral. The central bank would then have to decide the basis on which such lending could take place. The simplest – as well as the one that facilitates the operation of the direct transmission mechanism – would be for the central

23 Bordo & Levit (2017)
24 Sveriges Riksbank (2017); Kingsley (2016)
25 Nicolasi (2017)
26 Wikins (2017); Fung & Halaburda (2016)
27 Monetary Authority of Singapore (2017)
28 Mersch (2017)
29 Powell (2017)
30 Fan (2017)
31 Bech & Garratt (2017); Hileman & Rauchs (2017:88)
32 Bordo & Levit (2017)
33 Rogoff (2016); Haldane (2015); Broadbent (2016); Wissenschaftlicher Beirat beim Bundesministerium für Wirtschaft und Energie (Germany) (2017)
34 Rather than dropping dollar bills from helicopters, the central bank could simply credit CBDC accounts with additional funds. If CBDC accounts were linked to tax identifier numbers, there is the additional possibility for such distributions of “helicopter money” to be targeted toward lower income individuals (who have a higher propensity to spend).
35 We assume that bank deposits will continue to be exchangeable with currency on a one-to-one basis in unlimited amounts as long as the bank issuing the deposit remains in operation. In this respect, the analysis here differs from those who envision a variable exchange rate between the CBDC and bank deposits that is either set by the market bidding for a limited amount of CBDC, Barrdear & Kumhof (2016), or set by the authorities in a managed fashion. Agarwal & Kimball (2015)
36 Indeed, the ability to vary that rate of interest would constitute another policy tool. In effect, it would be the deposit rate on bank reserves writ large.
bank to give account holders the right to borrow. Under such an approach, the central bank would commit (possibly upon payment of a commitment fee) to refinance any eligible collateral that the account holder had prepositioned with the central bank. Note that the actual amount that the account holder could draw (the advance rate) and the interest rate the account holder would pay would depend on the haircuts and rates in force at the time.37

Second, the central bank will need to decide whether it should assume a more direct role in the extension of credit. To date, central banks have generally not played a prominent role—aside from their activities vis-à-vis banks—in either the origination of credit or in the exercise of creditors’ rights against obligors who violate covenants or actually default. They have restricted purchases of corporate bonds and asset-backed securities to the secondary market. Additionally, they have taken care to ensure that their holdings of such securities do not constitute a majority of any one issue or class of debt so that private creditors retain the primary responsibility for dealing with troubled borrowers.

Could central banks take on such a task? Probably, yes. Central banks already collect, collate and calibrate credit information on the obligors issuing the instruments that the central bank purchases outright or accepts as collateral. This enables central banks to assess the risks that they incur as well as to monitor overall credit conditions.38

Should central banks take on such a task? That is a different question entirely. An argument in favor is that they may need to do so. As outlined above, a CBDC may be superior to bank deposits, especially uninsured deposits. This could lead to a situation where the central bank has excess funds to invest, while banks struggle to find the funding to finance the loans that individuals and institutions are requesting banks to grant. In such a situation, central banks’ direct extension of credit could help ensure an adequate flow of credit to the “real economy.”

However, this course of action is filled with well-known dangers: it politicizes the extension of credit as well as the exercise of creditors’ rights—hardly a situation in which central banks are likely to be able to maintain their independence. At a minimum, legislatures and governments will want to review the criteria that the central bank uses to allocate credit. But, it is far more likely that legislatures and governments will seek to set the criteria along political as well as economic lines. The criteria would include both the terms and conditions on which the central bank extends credit as well as the rigor and vigor with which the central bank would seek to exercise any remedies available to it as creditor. This could easily lead to credit allocation toward favored sectors as well as forbearance for troubled creditors within such sectors. Neither of these outcomes would be good for efficiency or growth.39

As yet, no central bank is ready to introduce a freely available central bank digital currency. In part, this is because they are acutely aware of the technology, security, privacy and legal challenges that such an instrument would have to meet before a central bank could stake its reputation, its citizens’ wealth and its nation’s fortunes upon it.40 A CBDC would have to be universally and continuously available to eligible participants as well as able to handle a vast volume of transactions at speed and with accuracy. In addition, the CBDC would have to be robust against the cyber attacks that it would undoubtedly attract as a single point of failure for the economy as a whole. Finally, decisions regarding privacy would be required, not the least the entities with whom and the procedures by which the central bank might share the detailed information that a CBDC could generate about the transactions, finances and even possibly the location of participants.

In part, however, central banks hesitate to introduce a CBDC due to the disruption to the banking system that such an innovation could cause. As Vítor Constâncio, Vice President of the European Central Bank, recently remarked:

“(T)he use of the blockchain by central banks to create digital currency open to all citizens without limits would be really disruptive. This would be a radical political choice that

37 King (2016)
38 For example, the ECB is laying the foundation for an analytical credit database (AnaCredit) see European Central Bank (2016). This builds on prior work in various member states to create and maintain a credit register. Globally, the Legal Entity Identifier (LEI) project, LEI ROC (2015), assigns a unique number to each legal entity. This facilitates aggregation of exposures to that entity and provides the potential to extend the AnaCredit approach to other jurisdictions on a standardized basis.
39 In addition, central bank digital currency potentially aggravates privacy concerns. It would facilitate government tracking of an individual’s receipts and expenditures as well as their physical presence at the point at which such expenditures are made.
40 Powell (2017); Hileman & Rauchs (2017:103)
Are banks still special?

For the near future, that is undoubtedly true. But, that may not always be the case. Over the next five to 10 years, technology will continue to advance and this may enable central banks to satisfy their security, privacy and legal concerns. In other words, nothing would stand in the way of a central bank issuing a CBDC. Will they always refrain from doing so?

There are two circumstances in which they might not. If ADCs were to gain significant traction, this could threaten a central bank’s ability to exercise its macroeconomic mandate. If the central bank perceived the threat to be serious, it might well respond by issuing a CBDC.

A second possibility is perhaps more plausible. Suppose there were another crisis, and suppose that banks were to blame (or seen to be so). That could give rise to political pressure to “end banking as we know it,” and a CBDC could well be the way to do it.

As noted above, the introduction of a CBDC enables the central bank to perform directly the critical economic functions currently performed by banks. In particular, via the CBDC and its associated accounts at the central bank, institutions and individuals can hold deposits, make and receive payments, and potentially obtain credit.

Accordingly, banks and other financial institutions would become less systemic and that in turn would make the resolution of failing banks much more straightforward.

Resolution reform has already led to a reordering of the creditor hierarchy so that deposits have preference in liquidation (and insured deposits a first preference)42 (see Figure 5). Provided the authorities intervene in a timely manner,43 banks could potentially be resolved by transferring – together with an equivalent amount of the failed bank’s unencumbered good assets – the failed bank’s insured deposits (and possibly all deposits) to the central bank. The central bank would then credit each customer’s deposit to its CBDC account44 and customers could continue to conduct critical economic functions via their CBDC accounts. After the deposit transfer, a rump bank would remain. This would be restructured or liquidated. Thus, if a CBDC existed, it could facilitate resolution and enhance financial stability.

This is perhaps one more reason to characterize central banks’ current verdict on CBDC as “not now” rather than “never.” There may come a time when central banks opt or are told to “end banking as we know it”45 by introducing a CBDC. In that case, banks would not be special at all.

5. Conclusion

In sum, this paper argues that banks are only as special as central banks make them. Traditionally, that was very special indeed, for central banks restricted access to their payment systems to banks, while central banks themselves used banks as the transmission mechanism for monetary policy. To facilitate the transmission mechanism, central banks extended credit facilities to banks. This in turn propelled the central bank into supervising banks.

41 Constâncio (2017)

42 Note that according insured deposits, a first preference very significantly reduces the risk of deposits (and correspondingly the risk to the deposit guarantee scheme), particularly if there were substantial layers of subordinated liabilities below the insured deposit layer. This is essentially the case under the EU Banking Recovery and Resolution Directive (BRRD), particularly for significant institutions that will be required to maintain TLAC equal to 18% of its risk-weighted assets.

43 The key to timely intervention is timely valuation. This would enable the authorities to avoid forbearance and to institute resolution, while the bank still had positive net worth.

44 To be able to effect such a transfer smoothly, the “single view of the customer” contained in the bank’s resolution plan could be adjusted to include the number of the CBDC account to which the deposit would be transferred in the event the bank failed.
However, over time, banks have become less special. Central banks have moved payment systems to a RTGS basis.

The transmission mechanism for monetary policy now has many channels, not just banks. Central banks now extend credit to a broad range of institutions, not just banks. Supervision is broader as well.

Central banks are now considering whether to introduce central bank digital currencies. If they do, practically everyone may be able to have an account at the central bank. This would allow central banks to transmit monetary policy directly to the economy at large, to extend credit facilities to all, and to move away from much if not all of bank regulation and supervision. But, this could politicize credit as well as harm efficiency and growth. In the process, banks would not only become less systemic, they could well become collateral damage, for a central bank digital currency would not only replace cash. It would displace deposits.
References


German banks in financial centers: how risky is their business?

Cornelia Kerl
Deutsche Bundesbank

---

1 Ludwigstr. 13, 80539 München, Germany, cornelia.kerl@bundesbank.de
Abstract
Before as well as after the financial crisis of 2008, German banks' financial center affiliates have been on aggregate four times as large as the affiliates located elsewhere, and their balance sheet total has been half the size of the German parent banks' aggregate total assets. In addition, they are strongly connected with financial players in other financial centers, making them susceptible to distress in financial markets. German banks' affiliates in financial centers operate predominately as branches, as opposed to subsidiaries. This promotes the transmission of shocks to the parent bank due to the balance sheet consolidation. Financial center affiliates constantly have to roll over large amounts of short-term debt. As a consequence, they required larger injections of liquidity from their parent banks during the recent financial crisis. Balance sheet risk for parent banks is most likely to arise from financial center branches, as they are, in general, weakly capitalized. A change in accounting rules in December 2010 revealed their strong, formerly off-balance sheet involvement in derivatives trading.
This paper was partly written during research visits to the Deutsche Bundesbank while working as an external researcher. I would like to thank the Bundesbank and its staff for their continuous hospitality and support. The paper has benefited from valuable comments by Claudia Buch, Rainer Frey, Heinz Herrmann, Sebastian Hügelschäffer, Leif Lengelsen, Alexander Lipponer, Cordula Munzert and Peter Tillmann. All remaining errors and inaccuracies are my own. The paper represents my personal opinion and does not necessarily reflect the views of the Deutsche Bundesbank or its staff. I gratefully acknowledge financial support from the University of Giessen, which I received while working on this paper.

1. Introduction
While expanding their business internationally during the past decade, many banks established important affiliates in major financial centers, such as Luxembourg or the Cayman Islands. The financial crisis revealed the potential of financial shocks being transmitted via international financial platforms, where assets are traded globally and the interdependence of financial institutions is very high. It is, therefore, particularly important to understand how banks have made use of international financial centers.

This paper seeks to clarify the role of foreign affiliates of German banks located in financial centers, and to provide the first steps in analyzing whether they represented a source of risk to the stability of German multinational banks around the time of the financial crisis, as well as in its aftermath. The paper is intended to contribute to the policy debate regarding the regulation of financial centers, by providing often missing quantitative insights based on empirical research. For this purpose, the development of the financial center affiliates’ assets and liabilities during the financial crisis is examined carefully, relative to other affiliates of German banks located outside these centers. Detailed data on German banks’ foreign branches and subsidiaries reported to the Deutsche Bundesbank allow a deeper analysis of the affiliates’ individual balance sheets, which, due to another focus and a different type of granularity, is not possible using reports made to the Bank for International Settlements (BIS) or data provided by individual host countries.

Many empirical studies on international banking include a lump-sum control for characteristics of banks located in financial centers, but do not discuss these differences in more detail. Partly owing to the lack of bank-level data, the literature on the role of foreign banks’ affiliates in financial centers is rather limited. With a macroeconomic perspective, Errico and Musalem investigate legal and tax regimes of financial centers, and state that greater leeway for balance sheet management in these centers leads to higher solvency risk for banks. They highlight the role of financial center affiliates of Asian and Latin American banks in several regional crises during the 1980s and 1990s. These affiliates had built up unhedged exposures that were concentrated on very few asset types. In addition, they had provided extensive funding to their parent banks, which ran into trouble as their financial center affiliates experienced large losses.

Williams et al. analyze the costs and benefits for countries of becoming offshore financial centers, using data on realized government revenues in existing financial centers. They point to potential risk stemming from bank affiliates in financial centers, as they become larger than their parent banks. Dixon compares consolidated and locational claims of BIS-reporting banks on financial centers in order to determine the intermediation function of these platforms for several BIS-reporting countries. She stresses the potential risk stemming from unobserved off-balance sheet activity of banks in financial centers, on which no data exist. Rose and Spiegel regard this very aspect as one of the major risks arising from financial centers. They assess the influence of offshore financial centers on neighboring countries and come to the conclusion that loose regulation in financial centers encourages “bad behavior” on the part of source country banks, e.g., the building up of large off-balance sheet activity. Focusing on Chinese direct investment into financial centers, Sharman, on the other hand, emphasizes the advantage of these locations as points of access to international capital markets. Furthermore, he states that they are platforms where investors can establish links to profitable investments in, for example, developing countries.

Financial centers have evolved in places that either used to have the most capital to export (e.g., New York and London) or aimed at diversifying away from agriculture and tourism (e.g.,

---

2 Errico and Musalem (1999)
3 Williams et al. (2005)
4 Dixon (2001)
5 Rose and Spiegel (2007)
6 Sharman (2012)
German banks in financial centers: how risky is their business?

The Journal of Financial Perspectives

With new technology becoming available, some financial centers have become more and more specialized in providing international managerial and support services such as settlement and clearing (e.g., Luxembourg). In order to encourage this development, favorable fiscal and legal systems have been established in many financial centers, which make financial sector business more profitable. As early as 1999, Errico and Musalem called the number of financial centers that had evolved over time, and the subsequent volume of transactions carried out in these centers, a “pervasive practice.”8 The locational claims of BIS banks in the Cayman Islands, for example, amounted to 600 times its nominal gross domestic product (GDP) at the end of 2012. In comparison, BIS banks’ locational claims on Germany were 0.5 times nominal German GDP.

The International Monetary Fund (IMF) describes financial centers as “jurisdictions whose financial sector accounts for a significant – and disproportionate – share of its domestic economy.”9 These jurisdictions are also often called “offshore” financial centers, highlighting the fact that the lion’s share of financial interactions of these locations is carried out with nonresidents. In addition, these locations are often said to be “jurisdictions where offshore banks are exempt from a wide range of regulations that are normally imposed on onshore institutions.”10 However, no clear-cut definition of financial centers exists. In 2000, the Financial Stability Forum (FSF) launched an assessment program in order to address regulatory deficits in several financial centers, and drew up a list of countries hosting major financial centers for this purpose, on which this study relies mainly.11

This paper sheds some light on the role of financial center affiliates by investigating the timeframe from 2006 to 2012. Well after the financial crisis, the aggregate size of German banks’ financial center affiliates still exceeded the size of affiliates outside financial centers by a factor of three. In addition, they were about half the aggregate size of their German headquarters. Both figures highlight the importance of financial center affiliates for German banks. Activities in financial centers are mainly carried out via branches, which usually do not have to fulfill local regulatory requirements to the fullest and, therefore, allow greater flexibility regarding the management of balance sheets. This is reflected by the low levels of capitalization. In their loan portfolio, financial center affiliates, and financial center branches in particular (a share of almost 90%), focus on lending to banks and firms in financial centers. However, in doing so, they lend more across borders than to affiliates located outside financial centers. Furthermore, they are highly involved in trading securities, which put particular pressure on their stability during the financial crisis, and required increased support from parent banks. It is also striking that financial center affiliates have to constantly roll over large amounts of short-term debt. During the financial crisis, funding obtained from the parent bank was found to compensate for the difficulties of financial center affiliates, in particular, to tap

### Table 1: German banks’ affiliates in jurisdictions with major financial centers

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>Aggregate size of branches and subsidiaries of German banks (in billion euros)</th>
<th>Number of parent banks operating in this jurisdiction via branches or subsidiaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cayman Islands</td>
<td>178.109</td>
<td>10</td>
</tr>
<tr>
<td>2. Hong Kong SAR</td>
<td>31.395</td>
<td>7</td>
</tr>
<tr>
<td>3. Ireland</td>
<td>85.537</td>
<td>11</td>
</tr>
<tr>
<td>4. Luxembourg</td>
<td>387.167</td>
<td>26</td>
</tr>
<tr>
<td>5. Malaysia</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6. Mauritius</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7. Netherlands</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8. Philippines</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9. Singapore</td>
<td>68.725</td>
<td>10</td>
</tr>
<tr>
<td>10. Switzerland</td>
<td>24.467</td>
<td>13</td>
</tr>
<tr>
<td>11. Channel Islands</td>
<td>19.380</td>
<td>3</td>
</tr>
<tr>
<td>12. UK</td>
<td>940.446</td>
<td>25</td>
</tr>
<tr>
<td>13. US</td>
<td>300.966</td>
<td>12</td>
</tr>
</tbody>
</table>

Total in financial centers: 2,041.114
Total without US and UK: 799.732
Total via branches in financial centers: 1,544.126
Total via subsidiaries in financial centers: 497.018

Source: Deutsche Bundesbank and own calculations

NB: This table provides a list of jurisdictions defined as financial centers (see section 2.1), in which German banks had established affiliates (branches or subsidiaries) as of 2006m12. The (–) symbol signifies data not shown here on grounds of confidentiality. The UK Channel Islands comprise Guernsey, the Isle of Man and Jersey.

---

7 Lewis (1999)
8 Errico and Musalem (1999)
9 Darbar et al. (2003)
10 Errico and Musalem (1999)
11 FSF (2000)
short-term wholesale funding markets. Since December 2010, new accounting rules have required banks to report trading portfolio derivatives as part of the balance sheet position “other assets.” This dramatically increases the share of these assets in the total balance sheets of financial center branches and likely reveals part of the off-balance sheet risk that these entities have built up.

The remainder of the paper is structured as follows. Section 2 defines the group of financial center jurisdictions and investigates the location and size of German banks’ financial center affiliates. Section 3 analyzes in detail the asset structure and the funding sources of branches and subsidiaries in financial centers before, during and after the financial crisis, and compares them with affiliates outside financial centers. In section 4, paths for further research and data limitations with regard to the analysis of risk in financial centers are discussed. Section 5 concludes the findings.

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>Aggregate size of branches and subsidiaries of German banks (in billion euros)</th>
<th>Number of parent banks operating in this jurisdiction via branches or subsidiaries</th>
<th>Jurisdiction</th>
<th>Aggregate size of branches and subsidiaries of German banks (in billion euros)</th>
<th>Number of parent banks operating in this jurisdiction via branches or subsidiaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Australia</td>
<td>-</td>
<td>27. New Zealand - -</td>
<td>10. Czech Republic</td>
<td>10.804 4</td>
<td>36. Slovak Republic - -</td>
</tr>
<tr>
<td>3. Austria</td>
<td>140.233</td>
<td>16 28. Norway - -</td>
<td>12. Denmark</td>
<td>2.323 5</td>
<td>42. Taiwan - -</td>
</tr>
<tr>
<td>4. Belgium</td>
<td>8.615 6</td>
<td>29. Pakistan - -</td>
<td>13. Estonia - -</td>
<td>- -</td>
<td>43. Thailand - -</td>
</tr>
<tr>
<td>5. Bosnia and Herzegovina</td>
<td>-</td>
<td>30. Poland 14.479 7</td>
<td>14. Finland - -</td>
<td>- -</td>
<td>44. Turkey - -</td>
</tr>
<tr>
<td>8. Canada</td>
<td>6.046 3</td>
<td>33. Romania - -</td>
<td>17. Greece 2.323 5</td>
<td>5 42. Taiwan - -</td>
<td>47. Vietnam - -</td>
</tr>
<tr>
<td>9. Chile</td>
<td>-</td>
<td>34. Russian Federation 10.119 5</td>
<td>18. Hungary 15.092 7</td>
<td>7 43. Thailand - -</td>
<td>- -</td>
</tr>
<tr>
<td>10. China (mainland)</td>
<td>5.201 7</td>
<td>35. Saudi Arabia - -</td>
<td>19. India - -</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>11. Czech Republic</td>
<td>10.804 4</td>
<td>36. Slovak Republic - -</td>
<td>20. Indonesia - -</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>12. Denmark</td>
<td>-</td>
<td>37. Slovenia - -</td>
<td>21. Iran - -</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>13. Estonia</td>
<td>-</td>
<td>38. South Africa - -</td>
<td>22. Italy 50.290 16</td>
<td>14 41. Sweden 2.443 5</td>
<td>- -</td>
</tr>
<tr>
<td>14. Finland</td>
<td>-</td>
<td>39. Spain 28.515 11</td>
<td>23. Japan 44.621 6</td>
<td>6 Total outside financial centers 459.480 42</td>
<td>- -</td>
</tr>
<tr>
<td>15. Serbia and Montenegro</td>
<td>-</td>
<td>40. Sri Lanka - -</td>
<td>24. Latvia - -</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>Montenegro</td>
<td>-</td>
<td>Total via branches 195.271 34</td>
<td>25. Lithuania - -</td>
<td>- -</td>
<td>Total via subsidiaries 264.209 20</td>
</tr>
<tr>
<td>16. France</td>
<td>31.140 14</td>
<td>-</td>
<td>-</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>17. Greece</td>
<td>2.323 5</td>
<td>Total via subsidiaries 264.209 20</td>
<td>-</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>18. Hungary</td>
<td>15.092 7</td>
<td>-</td>
<td>-</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>19. India</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>20. Indonesia</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>21. Iran</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>22. Italy</td>
<td>50.290 16</td>
<td>-</td>
<td>-</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>23. Japan</td>
<td>44.621 6</td>
<td>-</td>
<td>-</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>24. Latvia</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>25. Lithuania</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>- -</td>
<td>- -</td>
</tr>
</tbody>
</table>
German banks in financial centers: how risky is their business?

2. How large is the German banks’ investment in financial centers?

2.1 Defining financial centers

This paper relies on the FSF’s list of financial centers. All “financial centers with significant offshore activities” are included, as well as “major financial centers,” which also received the questionnaires for offshore supervisors. Furthermore, in this study, the UK and the US are classified as financial centers. This decision is based on the following considerations. Zoromé, following the initiative of the IMF, developed an identification scheme for offshore financial centers on the basis of macroeconomic characteristics. His results strengthen the IMF’s definition, but also call for the UK to be regarded as a financial center. Because of the presence of foreign banks and the dominance of international banking, Cassis sees both the UK and the US as offshore financial centers. For both countries, BIS-consolidated foreign claims are smaller than locational foreign claims, which hints at a strong presence of foreign banks in both locations transacting with other countries. Almost all of the existing German banks’ affiliates in the UK are located in London, one of the world’s largest financial centers. Similarly, with negligible exceptions, the existing German affiliates in the US are located in New York or Delaware. The US has established the “International Banking Facilities (IBFs),” a booking concept that allows banks to conduct international banking under more favorable tax rules. According to the Federal Reserve Bank of New York (FRBNY), the majority of IBFs are registered in New York, as New York has introduced laws to facilitate the establishment of IBFs and has exempted net income from state and local taxes derived from such facilities. Delaware has become the most popular state in which to host holding companies, under which many foreign banks operate, since it has a very favorable tax and legal environment.

12 FSF (2000)
13 These are Andorra, Anguilla, Antigua, Aruba, Bahamas, Bahrain, Barbados, Belize, Bermuda, British Virgin Islands, Cayman Islands, Cook Islands, Costa Rica, Cyprus, Gibraltar, UK Channel Islands (Guernsey, Isle of Man, Jersey), Lebanon, Liechtenstein, Macau, Malta, Marshall Islands, Mauritius, Monaco, Nauru, Netherlands Antilles, Niue, Panama, St Kitts, Saint Lucia, St Vincent, Samoa, Seychelles, Turks and Caicos Islands, and Vanuatu.
14 These include Hong Kong SAR, Ireland, Luxembourg, Malaysia, Singapore and Switzerland.
15 Zoromé (2007)
16 Cassis (2006)
17 Mileti-Ferretti et al. (2010)
18 FRBNY (2007a)

Figure 1: Asset portfolio of German banks’ foreign affiliates.

Source: Deutsche Bundesbank and own calculations.

NB: The four panels show the composition of aggregate total assets of German banks’ foreign branches and subsidiaries located in financial centers as well as outside those centers. “Claims” refer to accounts receivable. “Firms” stands for the nonbank private sector and “foreign” stands for all jurisdictions except Germany.

2.2 Location and size of German banks’ financial center affiliates

The data used in this study comes from the reporting of banks to the Deutsche Bundesbank. All banks registered in Germany report, on a monthly basis, balance sheet characteristics of the German part of the bank, as well as of all its foreign affiliates (branches and subsidiaries). Branches usually do not fulfill regulatory capital requirements in their host countries to the fullest, but are consolidated into the balance sheet of the parent bank, whereas subsidiaries represent own legal entities. While subsidiaries are covered individually in the reports, activities of a bank’s foreign branches are aggregated by foreign jurisdiction. In addition to the standard balance sheet positions, all of the mentioned entities of the bank file a foreign positions report.

This allows for the identification of foreign activities vis-à-vis individual foreign countries and sectors within those countries.

19 Reports of subsidiaries are submitted if the German bank is the majority shareholder.
20 Fiorentino et al. (2010)
Table 1 provides a list of jurisdictions with major financial centers, in which German banks had established foreign affiliates by the end of 2006. Table 2 serves to compare the presence of banks in financial centers with their engagement in other foreign jurisdictions. Both the large number of banks with foreign affiliates in financial centers and the size of these affiliates underline the importance of the centers for German banks, which did not significantly change after the crisis.

By the end of 2006, 39 German banks had established affiliates in 13 financial center jurisdictions, almost as many as across all the other 47 countries (42 German banks). The majority of banks had representations in the financial centers of Luxembourg (26) and the UK (25), while Italy and Austria were the most popular among the nonfinancial center countries (in both countries, 16 banks had established affiliates by December 2006). The aggregate balance sheet size of financial center affiliates is roughly four times the aggregate size of nonfinancial center affiliates (roughly €2t versus €0.5t). Even if the US and the UK are excluded from the list of the 13 financial center countries, the remaining financial center affiliates in 11 countries are, on aggregate, larger than the nonfinancial center affiliates in the other 47 countries. The distribution of total assets between financial center affiliates and nonfinancial center affiliates does not change much between 2006m12 and 2012m12. The same applies to the number of banks active in the two categories of countries.

Individual financial centers host very large affiliates. For example, in the period under review, the overall size of affiliates in the Cayman Islands on its own is already larger than the size of all affiliates in Austria and France taken together. In view of the large balance sheet size of German banks’ financial center affiliates, it is of particular importance that parent banks control the risk associated with these affiliates effectively. For this purpose, it is important that supervisors of home and host country cooperate in order to achieve consolidated supervision of the respective bank holding companies.21

In financial centers, aggregate balance sheet size of affiliates between branches and subsidiaries is 76% to 24%, whereas the difference in aggregate size is much smaller in nonfinancial center countries. There, branches account for roughly 42% of the aggregate size of all affiliates. German banks thus operate in financial centers mainly through branches, while subsidiaries are more important outside financial centers. The dominance of branches in financial centers means that potential risk is more easily transmitted to the parent bank, as the common balance sheet facilitates transfers between the two entities of the bank holding company. Furthermore, the aggregate size of all parent banks at that time was around €4t. This means that all German parent banks together were only twice as large as their financial center affiliates (€2t, see table 1). This ratio highlights the relevance of financial center affiliates for German multinational banks and the need to control the risk taken by these affiliates.

Figure 2: Affiliates’ lending to financial centers: lending to banks and firms in financial centers relative to total lending to foreign banks and foreign firms carried out by the different groups of affiliates.
Source: Deutsche Bundesbank/own calculations.

NB: “Firms” stands for the nonbank private sector and “foreign” stands for all counterparties located outside Germany.

21 FSF (2000)
German banks in financial centers: how risky is their business?

3. The role of financial center affiliates for German banks
3.1 Strong focus on financial markets
During the recent financial crisis, the risk associated with some operations in financial markets was largely underestimated. Banks that were highly active in these markets, therefore, had to cope with substantial losses after the collapse of the US subprime market triggered a worldwide financial crisis and negatively affected the functioning of financial markets. In general, however, investments made in financial markets, particularly the trading of securities, are more volatile than, for example, bank lending to the real sector. This section analyzes the extent to which German banks’ affiliates in financial centers focus on intermediation in financial markets versus traditional bank lending.

3.1.1 Securities holdings and portfolio trading characterize financial center affiliates
The four panels in figure 1 show a breakdown of German banks’ foreign affiliates’ assets at three points in time: 2006m12 characterizes the situation before the outbreak of the financial crisis, and 2009m12 captures the structure of balance sheets after the first phase of the crisis and before the beginning of the sovereign debt crisis. 2012m12 stands for the situation found after the financial crisis. All statistics distinguish between branches and subsidiaries, and show aggregates of the two types of affiliates for those located in financial centers and those located elsewhere.

When comparing branches and subsidiaries in financial centers (top row, figure 1) with their equivalents outside financial centers (bottom row, figure 1), it is striking that both types of affiliates in financial centers hold a larger percentage of their total assets as securities. At the end of 2006, the share of securities holdings in total assets shrank during the financial crisis. Figure 3: Local lending: the extent to which different groups of affiliates lend locally as a share of their total lending to foreign banks and foreign firms.

Source: Deutsche Bundesbank/own calculations.

NB: “Firms” stands for the nonbank private sector and “foreign” stands for all countries except Germany.

Figure 3: Local lending: the extent to which different groups of affiliates lend locally as a share of their total lending to foreign banks and foreign firms.

Source: Deutsche Bundesbank/own calculations.

NB: “Firms” stands for the nonbank private sector and “foreign” stands for all countries except Germany.

Figure 4: Maturity structure of claims on foreign firms and foreign banks: short-term claims (accounts receivable) are displayed as a share of total claims on foreign banks and firms by affiliates in and outside financial centers.

Source: Deutsche Bundesbank and own calculations.

NB: “Firms” stands for the nonbank private sector and “foreign” refers to counterparties located outside Germany.

Figure 4: Maturity structure of claims on foreign firms and foreign banks: short-term claims (accounts receivable) are displayed as a share of total claims on foreign banks and firms by affiliates in and outside financial centers.

Source: Deutsche Bundesbank and own calculations.

NB: “Firms” stands for the nonbank private sector and “foreign” refers to counterparties located outside Germany.
The term “firms” is used to describe lending to the foreign nonbank private sector. The term “foreign” applies to all countries but Germany, hence it also applies to the affiliate’s country of residence.

NB: “Parent bank funding” refers to liabilities to the German headquarters, “deposit funding” comprises all liabilities to nonbanks, “wholesale funding” consists of funding via foreign banks (i.e., banks outside Germany) and own bonds and notes issued, and “short-term” refers to original maturities of less than one year.

On the one hand, in the event of a shock hitting financial markets first and foremost, a high concentration of assets in these markets, as can be found particularly in the case of financial center branches, most certainly limits the affiliates’ ability to compensate possible losses. This, in turn, increases the risk for parent banks of having to step in and provide emergency assistance. On the other hand, this focus on financial markets brings about deeper and more liquid markets with a highly specialized labor force. Parent banks may profit from this, as information on the development of global credit and funding markets is transmitted more quickly to headquarters. Affiliates outside financial centers are in a similar situation. Although they may be more sheltered from shocks hitting financial markets, they are highly exposed to economic slowdowns hitting the nonbank private sector. At the same time, their specialization in lending to the real sector of their country of residence may contribute significantly to the bank’s local success.
A large exposure to financial centers, however, could lead to higher balance sheet risk than a comparably large exposure to other economies. This arises from the fact that financial markets are highly interconnected. As mentioned above, the lion’s share of lending by financial center affiliates is directed to financial centers (figure 2). Furthermore, as shown in figure 3, the share of lending to local rather than foreign banks and firms is much smaller for financial center affiliates than it is for nonfinancial center affiliates. While financial center subsidiaries have lent 60% locally as of 2012, subsidiaries outside financial centers grant 80%-90% of their total bank and firm loans to local counterparties. Branches in financial centers lend only around 40% locally, while their equivalents outside financial centers focus on local lending to the tune of 80%. A shock hitting one financial center might, therefore, quickly affect other financial centers. This risk of contagion is highlighted by Garratt et al.23 In contrast, if an affiliate located outside a financial center suffers from a local economic downturn, this will be less likely to negatively impact other affiliates, because those located outside financial centers are mostly exposed to their country of residence. This very concentrated exposure certainly is disadvantageous from a diversification point of view if local problems arise. Still, parent banks then have the chance to isolate and solve the temporary difficulties. By contrast, in the case of a financial center crisis, necessary support measures for affiliates may soon exceed the parent banks’ capacity as a result of the high interconnectedness of financial centers.

3.2 Maturity structure of assets and liabilities

3.2.1 Short-term assets dominate financial center affiliates’ lending portfolio

Figure 4 shows the share of short-term claims in total claims on banks and firms for all four groups of affiliates between 2006 and 2012. Both branches and subsidiaries in financial centers (top panels) hold relatively more short-term claims on foreign firms than their equivalents outside financial centers (bottom panels). For branches, this difference is particularly striking. While branches outside financial centers grant only around 35% of their loans to foreign firms on a short-term basis, financial center branches lend roughly 60%-70% on a short-term basis. This higher level of short-term lending reflects once more the high interaction between financial centers.

Branches and subsidiaries in financial centers probably lend more to financial firms, such as hedge funds or investment vehicles, which engage more in short-term assets and securities trading activities and less in longer-term investments, for example, plant and equipment. These activities, in general, increase the default risk of loans, particularly in a financial crisis.

However, due to the shorter-term structure, balance sheets of branches and subsidiaries in financial centers seem to be more liquid. This likely increases their potential to buffer shocks to their portfolio, thus reducing the risk of transmitting disruptions to the parent banks. After the outbreak of the financial crisis, branches in financial centers, whose balance sheets were the most liquid before the crisis, seem to have used this buffer most often. They were at the forefront of letting short-term loans to foreign firms expire, which led to a drop in the share of short-term lending to foreign firms by 12 percentage points.

When comparing the situations before and after the financial crisis (2006m12 and 2012m12), the maturity structure of loans granted to foreign firms lengthened slightly across all four groups of affiliates. Compared with foreign banks, financial center affiliates (top panels, figure 4) reduced or kept stable their

23 Garratt et al. (2011)
short-term claims relative to longer-term claims. In contrast, affiliates outside financial centers (bottom panels, figure 4) increased slightly the share of loans to banks that are granted on a short-term basis. A possible explanation for this development could be that, in the light of increasing opacity of risk incorporated in many banks’ balance sheets, branches and subsidiaries outside financial centers had been downsizing their longer-term investments in other banks, instead concentrating on their core business — longer-term lending to firms. In addition, they could have tried to increase the liquidity of their asset portfolios after the financial crisis had rendered the economic environment less predictable. Assuming that investments in other banks have become more risky since the financial crisis, owing to remaining toxic assets in many banks’ balance sheets, then it may be that affiliates outside financial centers have recently downsized the risk of their lending portfolio by shortening the maturity of loans to foreign banks.

3.2.2 Large rollover risk in financial centers

Financial center affiliates not only issue more short-term loans relatively, but also finance their business to a larger extent on short-term wholesale markets. The liability structure of all four groups of affiliates is depicted in figure 5 as of 2006m12, 2009m12 and 2012m12. Branches in financial centers use short-term wholesale funding most often. This includes interbank funding with an original maturity of less than one year as well as own bonds and notes issued. Before the financial crisis, this group raised 43% of its total funding via short-term wholesale markets, compared with 25% obtained by branches outside financial centers. When located in financial centers, subsidiaries used short-term wholesale funding to a similar extent as branches, but they reduced or had to reduce their funding via this market at an earlier point in the financial crisis. By the end of 2012, both types of financial center affiliates had raised 25% of their total funding on short-term wholesale markets.

However, the development of the branches’ liability portfolios is again overshadowed by the change in the accounting rules implemented at the end of 2010. In the liability composition as of 2012m12, liabilities arising from trading portfolio derivatives are reported as “other liabilities.” This increases the share of this funding source in total liabilities and reduces the share of short-term wholesale funding. Yet, as most of these newly reported liabilities are probably of a short-term nature, the dominance of short-term liabilities of financial center branches is still striking. In total, over 50% of the liabilities of financial center branches were likely short-term at 2012m12.

Subsidiaries (and branches to a certain extent) outside financial centers can rely much more than their financial center equivalents on deposit funding, i.e., on funding obtained from nonbanks. This funding source adds stability to the liability portfolio of nonfinancial center affiliates, as it is less volatile and generally longer-term than wholesale funding. During the financial crisis, the share of deposits in total liabilities remained fairly stable for all types of affiliates except branches in financial centers. The higher degree of interaction with financial firms might explain the relatively large decline in deposits as a share of total liabilities. It has to be kept in mind that at the time, counterparties, such as special purpose vehicles, were realizing dramatic losses due to the collapse of the subprime market. Therefore, they may have withdrawn their deposits from banks. And again, the change in accounting rules might have also impacted the variation in deposits relative to total liabilities.

The comparably large reliance on short-term funding by affiliates located in financial centers versus outside financial centers creates greater exposure to rollover risk. When providers of short-term funds, such as other banks, lose confidence in market participants, which is what occurred during the financial crisis, then financial center affiliates will be hit sooner and more strongly than other affiliates because of the maturity structure of their liabilities. In a systemic crisis, it has proved to be advantageous if a bank has access to longer-term deposit funding. Many banks have, therefore, aimed at increasing the share of funding from nonbanks in total liabilities. Parent banks holding affiliates that rely largely on short-term funding have to be prepared to compensate possible shortfalls during a funding crisis.

3.3 Capitalization

3.3.1 Branches in general are very weakly capitalized

As can be seen in figure 5, equity capital represents only a small fraction of total liabilities of all groups of affiliates.
Subsidiaries are generally better capitalized than branches, both in and outside financial centers. Figure 6 traces the development over time of the affiliates’ equity capital to total liabilities in more detail. The large difference in the level of capitalization between branches and subsidiaries is striking. While, in the period under review, branches were having only 1%-2.5% of equity capital relative to total liabilities, subsidiaries were holding roughly 4%-7% of equity capital. This reflects the obligation of subsidiaries to fulfill local regulatory rules, including minimum capital requirements. However, their stock of equity capital is likely to include buffers. In comparison, German parent banks were holding on aggregate roughly 4% of equity capital relative to total liabilities. Being own legal bank entities themselves, subsidiaries have therefore been rather well-capitalized. Branches, by contrast, would have been rather weakly capitalized if they did not have unconditional access to the parent banks’ resources.

When capitalization is low, affiliates are unable to compensate large losses by themselves. Branches are, therefore, most likely to need support from the parent bank, particularly when they are located in financial centers. During the financial crisis, owing to their exposure to the disruptions on financial markets, banks located in financial centers registered very large defaults on loans and massive depreciations of securities. Branches in financial centers probably experienced the largest losses of all groups of affiliates, and were at the same time the least sufficiently capitalized group. To avoid an erosion of their capital base, many foreign affiliates of multinational bank holding companies most likely tapped the internal capital market organized by their parent banks.

3.3.2 Financial center subsidiaries are increasingly well-capitalized

According to figure 6, financial center branches and subsidiaries dispose of, in the period under review, less equity capital relative to total liabilities than their equivalents outside financial centers. While the difference between the two types of branches amounts to only half a percentage point on average over time, financial center subsidiaries had an equity capital ratio of roughly one to one-and-a-half percentage points below the ratio of subsidiaries outside financial centers. However, from mid-2009 to the end of 2012, financial center subsidiaries have increased their aggregate capital ratio to almost the level of nonfinancial center subsidiaries. Both then held between 6% and 7% of equity capital relative to total liabilities. This adjustment may be driven by the increasing need for capital buffers in connection with impending losses in the sovereign debt crisis. Branches, by contrast, slightly reduced their capital buffers during the same time, although the changes occurred on a much smaller scale.

Banking regulation has attached even greater significance to capital ratios since the financial crisis. US regulators, for instance, have adopted a stricter leverage rule for bank holding companies above a certain size. This entails insured depository institutions, such as subsidiaries of these bank holding companies being considered “well-capitalized” only if they have a capital ratio (also called “leverage ratio”) of around 8%. The ratio might be slightly lower if IFRS is used as the reporting standard rather than US GAAP, on which this figure is based.

24 Board of Governors of the Federal Reserve System (BGFRS) (2014)
25 Most branches of foreign banks in the US are not insured depository institutions. FRBNY
3.4 Reliance on funding provided by the parent bank
The lower level of capitalization of branches compared with subsidiaries, and to a lesser degree their use of deposit funding suggest that they rely more on parent bank funding. The liability portfolio compositions in figure 5 demonstrate this.
Irrespective of their country of residence, German banks’ foreign branches generally use more parent bank funding on a regular basis (accounts payable to the parent bank) than subsidiaries.26 Before the crisis (in 2006m12), branches in financial centers relied on parent bank funding to an extent of 21%, while subsidiaries in financial centers only received about 14% of total funding from headquarters (see top row, figure 5). The difference between branches and subsidiaries is even more dramatic for entities located outside financial centers (bottom row, figure 5). These branches used parent bank funding for 37% of their business in 2006m12, while subsidiaries outside financial centers only received 9% of their funding from parent banks (bottom row, figure 5). A larger reliance on parent bank funding generally increases the responsibility for parent banks to provide liquidity assistance on a regular basis.

3.4.1 Parent bank funding replaced short-term wholesale funding in the crisis
In the financial crisis, all four groups of affiliates increased the share of parent bank funding in total liabilities. This increase mirrors the decline in short-term wholesale funding during the first stage of the financial crisis. This enforces the impression that during the crisis parent bank funding was used to compensate for the decline in interbank and market funding, not only by financial center affiliates. The largest increase in parent bank funding in total liabilities occurred for subsidiaries outside financial centers (+14 percentage points), which had relied the least on parent bank funding before the crisis. However, funding via parent banks decreased for these subsidiaries between 2009m12 and 2012m12, and instead they returned to a higher degree of short-term wholesale funding.

Subsidiaries in financial centers continued to expand their

borrowing from their parent banks between 2009m12 and 2012m12 as well, but could not return to higher levels of market funding, unlike subsidiaries outside financial centers. Those subsidiaries may have been more deeply involved in the sovereign debt market, which was impacted by the subsequent stage of the crisis when banks incurred extensive write-downs on sovereign bonds from European peripheral states. This development suggests that a high level of borrowing from the parent bank on a regular basis does not increase the risk for parent banks stemming from the affiliates’ balance sheets per se. During the crisis, affiliates with the most difficulties in keeping up short-term wholesale funding increased their reliance on parent bank funding as a share of total liabilities the most. This development was not solely linked to affiliates residing in financial centers. These difficulties arose during the financial crisis because of increasing uncertainty about the riskiness of the affiliates’ asset portfolio as well as the worsening economic outlook.

3.4.2 Net support by parent banks in the crisis was greatest for financial center affiliates
As mentioned above, funds regularly flow between parent banks and affiliates, not just in times of crisis. Liquidity may be shifted to affiliates for funding purposes, but may also be withdrawn from affiliates by the parent bank in order to prioritize other activities, such as domestic lending. The net liabilities of affiliates to the parent bank, therefore, provide another insight into the role of the different groups of affiliates within the banking organization.
Figure 7 shows this net position, i.e., the aggregate net borrowing of affiliates from the parent bank, scaled by aggregate total foreign assets of the affiliates (hence, by the volume of the business that these affiliates carry out abroad).

Before 2007, foreign subsidiaries of German banks, both in and outside financial centers, were net providers of funds to their parent banks. This means that they were used as net funding sources by parent banks, possibly owing to their greater ability to raise funds locally compared with branches. Subsidiaries, being own legal entities, and often formerly stand-alone banks, usually have a greater network in place to attract deposits. This is reflected in their negative net borrowing position to parent banks. However, the situation changed in the first stage of the financial crisis. After mid-2007, subsidiaries were net borrowers

26 While subsidiaries report their assets and liabilities to the parent bank explicitly, it has to be approximated for branches. Frey and Kerl (2015) find that branches’ claims and liabilities on the German banking sector (excluding the central bank) provide a fairly good approximation for the position of the parent bank, as branches rarely interact with domestic banks other than their own headquarters. This approach is implemented here.
from their parent banks, as were branches. Compared with net borrowing of nonfinancial center affiliates, net reliance of financial center affiliates on their parent banks increased more strongly after the collapse of the subprime market in mid-2007 and again with the bankruptcy of Lehman Brothers in September 2008. This development was probably due to losses that financial center affiliates had to realize, as defaults and impairments had to be recognized in relation to receivables and securities in the aftermath of the crisis events.

Net borrowing from the parent bank relative to total foreign assets seems to have experienced a downward trend after 2008. At that time, German banks may have started to repatriate funds in order to stabilize domestic lending after their own access to market funding became more limited. Branches in financial centers needed increasing net support again starting in mid-2010. This also may signal deeper involvement of financial center branches in the trading of sovereign debt. Some of these investments had possibly turned risky, particularly with some European peripheral states struggling against insolvency, as they were unable to shoulder the support of their over-indebted banking systems.

Hence, whereas before the financial crisis, branches and subsidiaries outside financial centers had relied more on parent bank funding than their respective equivalents in financial centers, in the course of the crisis, the increase in net borrowing from headquarters was larger for financial center affiliates. This possibly reflects the higher risk that was incorporated in their balance sheets, which materialized with the disruptions on financial markets.

4. Paths for further research and data limitations

This paper provides a first insight into the relevance of financial centers to German multinational banks before, during and shortly after the financial crisis, and potential risk that is associated with affiliates located in these centers. Banks shift business, such as the intermediation of international bank lending to financial centers, if this appears to be more profitable than conducting the same business elsewhere within the bank holding company.

New regulatory approaches, which are being implemented within the framework of the Basel III accords, aim at strengthening the equity capital base of banks and put new limits to the growth of multinational bank holding companies at low levels of equity. It is, however, unclear how banks will react to these new rules. Regulators are struggling with the implementation of uniform regulation across country borders to create a “level playing field.” This opens up leeway for jurisdictions that refrain from implementing tighter restrictions on financial transactions. Some offshore financial centers will continue to provide or will enlarge their platforms where regulation is less strict, so that they will attract bank business from more tightly regulated countries.

It is, therefore, possible that the home countries of multinational banks will experience a further loss of control over parts of their banks’ balance sheets despite tighter regulation. As discussed in this paper, financial center affiliates of multinational banks can increase much more in size compared with their home institutions. This development will probably continue as restrictions on home institutions tighten. Further research should, therefore, analyze the extent to which the size of financial center affiliates actually contributes to the spillover of financial stress to parent banks.

For this purpose, it is advantageous to model in more detail the relationship between parent banks and their financial center affiliates, as well as between financial center affiliates and nonfinancial center affiliates. For German banks, data reports available since June 2010 provide further insight into claims and liabilities of German banks’ foreign affiliates compared with each other. As longer time series become available, an analysis of this data can refine an assessment on how stress can be transmitted within banking groups and to other banks. Models that assess the risk of contagion within and between networks can provide useful tools. Garratt et al., for example, conduct such an analysis on the country level.

As is the case with other studies on the role of financial centers, a very detailed balance sheet data reported to bank supervisors do not reveal the extent to which banks make off-balance sheet investments. A change in accounting rules implemented in Germany in December 2010 offers more insights into this off-balance sheet activity by requiring the reporting of trading 28 Garratt et al. (2011) 29 Lane and Milesi-Ferretti (2010)
portfolio derivatives. For branches in financial centers, this has changed the composition of assets dramatically, as this position accounted for roughly 40% of the balance sheet after the change of rules (see section 3.1).

However, the lack of data on activities in financial centers likely leads to an underestimation of the risk that parent banks ultimately have to bear. In particular, complex unreported ownership structures of financial firms, including hedge funds and special purpose vehicles, make it even harder to assess the size and the riskiness of banks’ investments in financial centers. Therefore, more data on relationships between banks and nonbanks should be collected and analyzed by supervisors. Arteta et al., for instance, assemble data on ownership and sizes of special purpose vehicles that had invested in the US subprime market prior to its collapse. They find that the buildup of too much risk by these vehicles was possible because of manager agency problems but also because of the lack of government control of ownership structures. This path of research should be extended in order to assess more accurately the risk incorporated in financial center investments.

5. Conclusion

Financial centers are often accused of encouraging tax fraud and money laundering, but many of them have adopted stricter legislative frameworks in their financial systems in order to address these concerns. Nevertheless, these countries still offer an advantageous tax and legal environment for financial transactions. Their importance for global banking has continued to increase in terms of the number of existing financial centers and the volume of financial transactions intermediated. This study, therefore, analyzes the role played by financial center affiliates for German multinational banks and provides first steps towards assessing the risk for parent banks incorporated in these affiliates’ balance sheets.

Financial center affiliates of German banks were, even well after the financial crisis, four times as large as their affiliates located elsewhere. In addition, the German parent banks of these affiliates were, on aggregate, only twice as large as the financial center affiliates themselves. Because of their large size and their particularly strong interactions, financial center affiliates are likely to transmit distress to their parent banks whenever disruptions to financial markets occur. Furthermore, German banks operate in financial centers mainly through branches that are consolidated into the parent banks’ balance sheets. This promotes the transmission of shocks. When several financial center affiliates need support from the parent bank at the same time, the latter may quickly become overtaxed by the support needed, particularly when taking into account the large size of financial center affiliates.

Judging from the analysis before and after the financial crisis, branches both in and outside financial centers would be very weakly capitalized if they were not consolidated into the parent banks’ balance sheets. On the contrary, after the financial crisis, financial center subsidiaries increased capitalization almost to the level of nonfinancial center subsidiaries, and were subsequently better capitalized than the aggregate of their parent banks. This development reduced the risk incorporated in their balance sheets.

When the recent crisis unfolded, parent bank funding was used by all types of affiliates to compensate for a loss in short-term wholesale funding. Only subsidiaries outside financial centers, up to the end of 2012, managed to return to their pre-crisis level of market funding. In general, affiliates in financial centers, particularly branches, constantly need to roll over large amounts of short-term debt. This increases the risk that parent banks may have to provide financial assistance during downturns on funding markets. Therefore, it was to be expected that net borrowing of financial center affiliates from their parent banks would increase more sharply after the collapse of the subprime market in mid-2007.

With the experiences of the financial crisis, regulators in advanced economies are tightening regulation for large banks and are setting limits to their size when weakly capitalized. The introduction of leverage ratios, for example, compares equity capital with total assets rather than only with risk-weighted assets. In case of a bank defaulting, this is expected to reduce the problem of institutions being “too big to fail.”
The new rules may, however, encourage shifting business off the balance sheet. Furthermore, the largely unregulated hedge fund industry or other market participants offering bank-type services can pick up investments that are no longer profitable for banks. Financial centers have hosted mostly banking activity, but the fund industry has made some ground.\textsuperscript{35} As a result of lower tax burdens and their platform character within the global banking network, financial centers represent an attractive place not only for banks but also for hedge funds and other financial firms. For example, in the early 2000s, the hedge fund industry grew in the Cayman Islands by 30\% per year.\textsuperscript{36}

The relevance of financial centers as international financial platforms is, therefore, likely to increase in the near future. In order to avoid excessive risk taking in these countries, supervisors should aim at increasing the transparency of local activities and work closely together with regulators of the international banks’ home countries. Tax regimes and supervisory standards should be further harmonized.\textsuperscript{37} Effective supervision of bank holding companies on a consolidated level may help to avoid the buildup of risk exposures, which could threaten the solvency of international banks and ultimately endanger global financial stability.

\textsuperscript{35} Mileti-Ferretti et al. (2010)
\textsuperscript{36} McGuire (2005)
\textsuperscript{37} FSF (2000)
References


Stress testing banks: whence and whither?

Pavel Kapinos, PhD
Federal Reserve Bank of Dallas

Christopher Martin
Federal Deposit Insurance Corporation (FDIC)

Oscar Mitnik
Inter-American Development Bank

1 Corresponding author: Federal Reserve Bank of Dallas, 2200 N Pearl Street, Dallas, TX 75201; tel: +1 214 922 6494; email: pavel.kapinos@dal.frb.org. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Federal Reserve System; Federal Deposit Insurance Corporation; or the Inter-American Development Bank, its Board of Directors, or the countries they represent. We thank Arthur Micheli and Shibani Putatunda for their excellent research assistance, and our colleagues, especially Emily Johnston Ross, Myron Kwast and John O’Keeffe, for feedback on a preliminary draft of the paper.
Abstract
This paper provides a brief overview of the recent practice of stress testing banking institutions, focusing on capital adequacy. We argue that stress testing has been successfully used to mitigate bank opacity; quantify systemic risk under extreme but plausible stress; keep the participants mindful of severely adverse shocks, thereby mitigating “disaster myopia” and concomitant financial instability; and improve the data collection and analytical capabilities of financial institutions. Our paper then reviews several critiques of stress testing made by policymakers and academics. We also propose several modifications of the current stress testing practice, such as the fusion of liquidity and capital adequacy stress testing, expansion of granular data availability, and explicit modeling of sectors inextricably connected to banking as well as the feedback mechanisms from these sectors. Addressing these issues is likely to keep stress testing highly relevant for promoting financial stability in the future.
1. Introduction
The 2008-09 global financial crisis and the ensuing Great Recession brought about arguably the largest financial reforms since the Great Depression of the 1930s. In the US, much of this new regulation was implemented following the passage of the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act or the DFA) in 2010. One of the DFA’s numerous provisions aimed at improving the stability of the US financial system calls for larger banks and bank holding companies (BHCs) to undergo so-called “stress tests” that require both bank managers and regulators to understand what would happen to banking institutions if they were subjected to “exceptional but plausible macroeconomic shocks.” Similar reforms have been undertaken in other countries as well, with stress testing emerging as a commonly used instrument in the toolkit of financial supervision and regulation. The purpose of this paper is to survey current stress testing practices, discuss their relative advantages and disadvantages, address some of the criticism leveled against stress testing exercises and propose improvements to future stress tests. We restrict our analysis to the banking sector and to stress testing mandated by regulatory rules, as opposed to purely internal stress tests conducted by financial institutions.5

We argue that regulatory stress testing is a useful tool for addressing classical concerns with the robustness of financial architecture for the following reasons: firstly, it could complement Basel capital regulation in capturing systemic risk. Secondly, it may reduce the opacity of the banking industry and reduce asymmetric information among market participants. Thirdly, it forces both regulators and financial institutions to periodically assess the possible effects of highly adverse scenarios, which tend to be ignored in good times. Fourthly, it is useful as a regulatory tool to incentivize banks to collect better data, expand their quantitative analytical capabilities, and engage in more robust and holistic risk management practices.

The first benefit of stress testing is its potential to complement Basel capital regulation in capturing systemic risk. The Group of Ten defines systemic risk as “the risk that an event will trigger a loss of economic value or confidence in, and attendant increases in uncertainty about, a substantial portion of the financial system that is serious enough to quite probably have significant adverse effects on the real economy.” Early efforts of stress testing, as well as Basel capital requirements, have not been fully effective in addressing spillovers across financial institutions or feedbacks to the macroeconomy. In this sense, both programs have remained largely microprudential, though recent work has moved best practice in the direction of these systemic or macroprudential ideas. One of the main arguments of this paper is that more work is needed in this area for stress testing to realize its potential to serve as both a forward-looking and macroprudential complement to Basel capital regulations.

Second, stress testing may reduce the opacity of the banking industry and reduce asymmetric information among market participants. To the extent that the results are made public, stress tests can make financial institutions more transparent, effectively acting as another form of financial disclosure. This reduction of asymmetric information can be particularly valuable during crises. In addition, Bookstaber and colleagues argue that disclosure “enhances market discipline as a tool for financial stability, strengthens the incentives for financial institutions to meet supervisory standards, and reinforces confidence in the functioning of the financial system.”

Two strands of the literature are rapidly emerging to address this issue. Theoretical papers evaluate the desirability of disclosing stress testing results. Goldstein and Sapra argue that disclosing such results improves financial stability by imposing market discipline, which leads to improvements in resource allocation.7 They caution, however, that bank-specific effects may impose costs, such as

---

2. Sorge (2004) provides useful definitions and an overview of the early stress testing practices that preceded the Great Recession. Drehmann (2009) and Kapinos and Mitnik (2016) describe methodological updates to his early contribution. Not all stress testing in the US is mandated by the DFA; for example, the National Credit Union Administration (NCUA) requires credit unions with more than US$10b in assets to undergo stress testing, even though the DFA does not mention credit unions in its stress testing provisions. See NCUA Rules and Regulations §702–Capital Planning and Stress Testing for details.

3. The DFA also mandates stress testing of nonbank “systemically important financial institutions” (SIFIs), as determined by the Financial Stability Oversight Council (FSOC). In this paper, we do not address stress testing of these institutions.
inducing excessively volatile reactions to results releases, thereby reducing the incentive for analysts to monitor banks. Goldstein and Leitner study a trade-off between disclosure that reduces market breakdowns and also possibly reduces banks’ risk-sharing incentives. They argue in favor of partial disclosure in stressful conditions and no disclosure in normal times. On the empirical front, the cumulative abnormal return methodology has been used to evaluate the stock market reaction to the release of stress test results. Morgan et al. find that the market largely knew which banks would pass or fail the Federal Reserve’s 2009 Supervisory Capital Assessment Program (SCAP), but the release of the size banks would pass or fail the Federal Reserve’s 2009 Supervisory Capital Assessment Program (SCAP), but the release of the size of capital shortfalls provided useful information. Candelon and Sy review the evidence for subsequent stress tests and argue that while the market reaction to their releases has decreased from the SCAP levels, it continues to be significant. Petrella and Resti discuss evidence that the 2011 European stress testing exercise provided useful information for market participants. Therefore, stress tests do appear to reduce the opacity of the banking industry.

Third, stress testing can keep the participants mindful of severely adverse shocks. As Minsky and Guttentag and Herring argue, financial systems exhibit fundamentally destabilizing behavior. Minsky refers to the idea as the “financial instability hypothesis”, and Guttentag and Herring call it “disaster myopia”, but the intuition is similar: periods of economic and financial calm cause the financial system to become less resilient to extreme shocks because market participants perceive a reduced probability of their realization and engage in accordingly riskier behavior. By subjecting individual banks and the financial system to severely adverse hypothetical scenarios in times of stability, stress testing may uncover emerging fragility in bank portfolios. Hence, stress testing can serve as a formal structure to remind market participants that such hypothetical adversity may yet happen in practice, deterring destabilizing behaviors.

Fourth, stress testing is useful as a regulatory tool to incentivize banks to collect better data, expand their quantitative analytical capabilities, and engage in more robust and holistic risk management practices. Schuermann argues that financial institutions have had to evaluate more holistically their approach to risk management, which, in most cases, has driven improved data collection, model development and validation capabilities. As we discuss below, this aspect of regulatory stress testing has been especially challenging for smaller financial institutions.

Notwithstanding its usefulness, stress testing faces a number of important technical challenges and conceptual critiques. One such challenge relates to the modeling of systemic impacts, and spillovers of financial stress and bank failures. Addressing this challenge will involve the inclusion of liquidity risk and firms’ interaction in regulatory stress tests, an area of ongoing but incomplete progress. Adequately stressing liquidity risk, in turn, may require the incorporation of market, as opposed to book, equity into stress testing models, given the empirical relationship between market equity and liquidity. Another technical challenge facing stress testing is that reasonable assumptions regarding difficult-to-model elements of bank balance sheets can lead to meaningfully different results. Finally, we address several critiques of stress testing, including that it may implicitly commit regulators to bail out banks that perform unexpectedly poorly; that it represents an undue burden on banks, particularly on smaller institutions; and that, as currently conducted, it represents an unduly opaque regulatory process.

Our view is that there is considerable room for improvement to the current practices of regulatory stress testing. Addressing the technical challenges above will likely make the process more useful for regulators and banks alike. Our discussion of the critiques of the fundamental appropriateness of stress testing, however, suggests that most of the criticism is misplaced and that stress testing is a fundamentally useful regulatory tool, albeit one that can and should be streamlined and improved.

8 Goldstein and Leitner (2015)
9 Morgan et al. (2014) also provide a brief review of the literature on bank opacity whose balance, with caveats, appears to support the notion that the banking industry is more opaque than other sectors of the US economy.
10 Candelon and Sy (2015). Similarly, Glasserman and Tangirala (2016) find that the Comprehensive Capital Analysis and Review (CCAR) results are becoming increasingly predictable.
11 Petrella and Resti (2013)
13 Schuermann (2015)
14 Bookstaber et al. (2014)
15 IMF (2012a) and Annaert et al. (2013). Indeed, Kuritzkes and Scott (2009) and Flannery (2013) argue that stress testing should focus on something closer to market equity. Particularly during times of stress, both fundamental solvency and market equity tend to diverge from book equity valuation. Flannery (2005) shows that managers have incentives to inflate book value in these cases. This concern is emphasized by the stylized fact from the above references and Kapinos and Mitnik (2016) that most banks that failed during and after the 2008-09 crisis had book equity well above regulatory thresholds just before their failure.
The rest of this paper is organized as follows. Section 2 provides a broad survey of the current practice of stress testing, as conducted in the US, the European Union (EU) and the UK, and discusses both differences in methodological approaches and outcomes from the practical implementation of stress tests. Section 3 discusses some of the known challenges to the current practices and offers ways of addressing them. Building on this technical assessment, section 4 reviews critiques that have been recently advanced against stress testing. Finally, section 5 summarizes issues that, if addressed, would change existing stress testing practices; suggests ways for implementing them; and offers conclusions and recommendations on the future of stress testing.

2. The practice of stress testing: history and implementation

This section discusses current stress testing practices and provides a brief history of supervisory regimes. Section 2.1 briefly reviews the history of stress testing, while section 2.2 describes the current landscape in capital adequacy stress test modeling, focusing on the technical details.

2.1 A brief history of stress testing programs

2.1.1 Precrisis evolution: from 1980s to 2007

Stress testing began in the 1980s as a relatively minor risk management tool applied largely to individual risks in isolation. Many private institutions and regulators began stressing interest rate risk in the 1980s. Later, the Market Risk Amendment of 1995 to the Basel Capital Accords required many larger financial institutions to stress market and liquidity risks in their trading books. Based on discussions with market participants and a formal survey of stress testing practices, the committee on the global financial system documented that stress testing was, by the late 1990s, commonly used to assess market, liquidity and even occasionally credit risk in the trading book. In addition to stress testing their trading portfolios, a report issued in 1999 indicated that some firms were developing techniques to stress investment and credit portfolios as well.

With the advent of stress testing by the International Monetary Fund (IMF) under the aegis of the Financial Sector Assessment Program (FSAP), it began to address broader stability concerns. Introduced in 1999, the FSAP includes stress tests conducted by IMF staff or by a combination of IMF and in-country supervisory staff on a periodic, country-by-country basis. Unlike earlier stress tests, FSAP tests focus on aggregate financial stability and jointly stress a wider array of risks, notably including credit, interest and exchange rate risks. While the early FSAP stress tests tended to use models and scenarios that were simplistic by current standards, more recent tests are comparable in sophistication with current supervisory stress tests. In addition to any direct financial stability benefits, the FSAP tests served to generate research interest and led some regulators to begin conducting routine stress testing of their financial systems, even before the financial crisis.

The last element of precrisis stress testing worth mentioning serves as something of a cautionary tale, from which we draw several conclusions in sections 3 and 4. Beginning in 2002, the Office of Federal Housing Enterprise Oversight (OFHEO) stress tested the interest rate and credit risk exposures of Fannie Mae and Freddie Mac, the two US Government-sponsored enterprises (GSEs) central to financing the US housing market. Much like recent stress tests, the results were used to determine capital adequacy and effective capital requirements. However, Frame et al. describe the stress test exercise as “a spectacular failure,” owing to the fact that the tests suggested the GSEs’ capital was adequate until just a few months before they became insolvent and were placed into conservatorship by the US Government. Frame et al.’s analysis suggests that three major lessons should be drawn from the OFHEO experience. First, models should be conceptually evaluated and updated, as well as re-estimated, regularly in order to maintain accuracy and rigor. Perhaps because of cumbersome and detailed publication requirements for the model, OFHEO’s stress test model and its coefficients were static from their implementation in 2002 onward. Frame et al. find that this failure to update the model resulted in a substantial understatement of stressed losses. Second, regulators must take care to select appropriately stressful scenarios.

17 BCBS (1996)
18 Committee on the global financial system (2000, 2001). See also Fender et al. (2001) for a consolidated summary of the findings of the survey.
20 Blaschke et al. (2001)
21 Regulatory authorities in Mexico and Luxembourg routinely conducted stress tests before the advent of FSAP tests. Several others began after their participation in the FSAP, including Finland, Hungary and Poland. IMF (2003)
22 Frame et al. (2015)
Stress testing banks: whence and whither?

Stress only sets in after an extended period of time in the OFHEO scenarios, at which point existing loans would have acquired considerable seasoning, reducing the probability of their default. Finally, assumptions on balance sheet behavior, especially asset growth, over the stress scenario are critical. The OFHEO stress tests assumed that the GSEs would have no new business during the scenario. This served to limit the base of loans to suffer losses, and it also interacted with the profile of loss rates by mortgage vintage to further decrease estimated losses. These three lessons should be applied to more recent stress testing, as we discuss in more depth below.

Table 1: Summary of stress testing exercises

<table>
<thead>
<tr>
<th>Program</th>
<th>Description</th>
<th>Year</th>
<th>Results</th>
<th>Scenario variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCAP</td>
<td>Stress test of 19 US BHCs with greater than US$100b in assets. Covered credit risk in banking book and market risk in trading book.</td>
<td>2009</td>
<td>Ten BHCs capital deficient, with a total shortfall of US$75b. Only one bank used Treasury recapitalization program.</td>
<td>GDP, unemployment, HPI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>Four banks breached capital limit</td>
<td>Same plus CRE prices, VIX, international</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2013</td>
<td>Two BHCs failed</td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>Five BHCs failed</td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015</td>
<td>Two BHCs failed</td>
<td>Same</td>
</tr>
<tr>
<td>DFAST¹</td>
<td>Quantitative stress test and qualitative analysis of capital planning process at all federally regulated US financial companies with assets over US$10b.</td>
<td>2013</td>
<td>One BHC dropped below 5% tier 1 common ratio</td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>One BHC dropped below 5% tier 1 common ratio</td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015</td>
<td>No BHCs breached 5% threshold</td>
<td>Same</td>
</tr>
<tr>
<td>NCUA</td>
<td>Similar to CCAR, but applied to federally insured credit unions.</td>
<td>2014</td>
<td>Results not published</td>
<td>Similar</td>
</tr>
<tr>
<td>EU-wide</td>
<td>Test of credit risk in banking book and market risk in trading book. 2009 covered 22 cross-border banks with 60% of EU banking assets. 2010-11 covered 90 or 91 largest banks to capture at least 50% of banking assets by country. 2014 covered 130 banks.</td>
<td>2009</td>
<td>No bank’s tier 1 ratio fell below 6%</td>
<td>At least EU-wide GDP, unemployment, and RE prices</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>Seven banks failed</td>
<td>Country-level GDP, CPI, unemployment, yields, RE prices, international</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>Eight banks failed</td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>25 banks failed</td>
<td>Same</td>
</tr>
<tr>
<td>UK</td>
<td>Similar to 2014 EU-wide stress test but tailored to UK-specific risks and covered eight banks.</td>
<td>2014</td>
<td>Three banks failed</td>
<td>Same</td>
</tr>
</tbody>
</table>

Table 1: Summary of stress testing exercises

Sources: FRB (2009a, 2011, 2012, 2013a,b,c, 2014a,b,c, 2015a,b, 2016a,b, 2017a,b), NCUA (2014, 2015, 2017), CEBS (2009, 2010), EBA (2011a,b, 2014, 2016a,b), ECB (2011, 2014), European Systemic Risk Board (2014) and Bank of England (2014a,b, 2015a,b, 2016a,b, 2017). 1 As noted in section 2.1, the full institutional coverage and disclosure of DFAST results is fragmented and new. Consequently, the test results presented here reflect only the Federal Reserve’s disclosure of the results from its DFAST projections for the same set of institutions subject to CCAR. Wider results for all banks are not centrally available. 2 Additional stress testing of smaller EU banks is conducted at the national level and not discussed here. 3 Banks with capital shortfalls were considered not to have “failed” because of capital actions already undertaken, namely between the stress test “as of” date and the conclusion of the stress test.

2.1.2 Developments since the crisis: from 2008 to present

The financial crisis dramatically accelerated both the use of stress testing by regulators and research on the topic. Because of the first major stress test conducted during the crisis, the Federal Reserve’s SCAP was widely perceived as successful as it has served as the basis for the techniques and scale of stress testing exercises performed since then.²³

Conducted in early 2009, the SCAP was intended both to estimate the scale of recapitalization needed by the financial sector and to reduce asymmetric information, which was elevated

²³ For an early survey of alternative methods for stress testing credit risk, Foglia (2009).
during the peak of the crisis.\textsuperscript{24} The SCAP covered the 19 US BHCs with at least US$100b of consolidated assets. These 19 institutions held about two thirds of the assets in the US banking system. The supervisory agencies and BHCs estimated credit losses on 12 portfolios within the banking book and on investment securities (both held-to-maturity and available -for-sale) for each bank. They also estimated pre-provision net revenue for institutions with trading book assets in excess of US$100b, market and counterparty credit losses on their trading books. Bank liquidity was not stressed. Projections for these variables were made at the annual frequency for 2009 and 2010, though appropriate measurement of the allowance for loan and lease losses required loss estimates for 2011 as well. The projections of financial institution health were made for two hypothetical scenarios provided by the regulators, “baseline” and “more adverse” provided by the regulators. The scenarios comprised paths for real GDP growth, the national unemployment rate and a house price index. Firms were encouraged to develop paths for additional macroeconomic variables, consistent with the three regulator-provided series, as needed. Banks whose capital was forecast to fall below the threshold were required to raise additional capital.\textsuperscript{25} Firms were encouraged to obtain additional private capital, such as by common stock issuance. However, any firm failing to recapitalize privately by November 2009 would have to receive a capital injection from the Treasury and would be disallowed from making dividend distributions. Of the 19 firms tested, 10 were found to have a capital deficiency totaling US$75b.

Though the aims of the SCAP were relatively narrow, the program’s architecture provides a solid basis for stress tests for several reasons. First, the techniques to provide credible assessments of stress losses and capitalization are equally necessary for estimating recapitalization needs or for mitigating disaster myopia. More generally, the modeling, supervisory oversight and disclosure practices of the SCAP were designed to ensure clarity, accountability and the reduction of bank opacity. Finally, the SCAP laid the groundwork for wider and systematic data collection and analysis by banks and their regulators. This data collection is useful for many purposes, including the development of highly granular stress test models. We argue below that these highly granular models seem more promising than highly aggregated models. We also maintain that, suitably redacted, loan-level data gathered for stress testing should be made available to researchers to spur further development of bottom-up models.

Subsequent stress tests in the US have fallen under several programs; table 1 enumerates the US and European stress tests and provides summary information.\textsuperscript{26} BHCs with consolidated assets in excess of US$50b, as well as nonbank financial institutions designated as systemically important by the Financial Stability Oversight Council (FSOC), are required to undergo the CCAR annually.\textsuperscript{27} CCAR includes a quantitative stress test of BHCs’ credit risk in the banking book and market risk in the trading book, conducted by the Federal Reserve. It also includes a qualitative review of BHCs’ capital planning and stress testing procedures.\textsuperscript{28} BHCs can fail CCAR on either quantitative or qualitative grounds, with failure typically implying restrictions on BHCs’ planned dividend payouts or share repurchases.\textsuperscript{29} The basic stress testing paradigm for CCAR is quite similar to the SCAP architecture, though modeling, data collection and disclosure have improved as the process has evolved. The CCAR exercise is a useful financial stability tool beyond the benefits derived from the SCAP. Specifically, because regulators can restrict firms’ capital plans, CCAR is clearly a useful forward-looking approach to assessing capital adequacy and assigning capital requirements, complementing Basel regulation. Moreover, the emphasis on the qualitative aspects of capital planning and risk management ensures that CCAR is useful for incentivizing improved data collection and risk management, our fourth dimension of value for stress testing.

Two other US stress test programs closely resemble CCAR and SCAP. First, Dodd-Frank Act Stress Testing (DFAST) requires all federally regulated financial institutions with assets in excess

---

\textsuperscript{24} FRB (2009a, 2009b, 2009c)

\textsuperscript{25} As noted by Hirtle and Lehnert (2015), the requirement that the BHCs raise a particular dollar value of capital, as opposed to meeting a minimum capital ratio, ensured that the banks could not simply shed assets to address the capital shortfall. Regulators arguably wished to discourage such de-leveraging during the crisis.

\textsuperscript{26} Of course, many other supervisory agencies conduct stress tests, besides the US, EU and UK. Among others, CBOI (2011, 2012) discusses the Irish stress tests; IMF (2012b), Oliver Wyman (2012) and Roland Berger (2012) discuss Spanish stress tests.

\textsuperscript{27} FRB (2014d, 2015a)

\textsuperscript{28} These stress tests are also intended to capture bank operational risks, though these risks are measured using rather simple and reduced-form techniques. Because operational risk is largely unrelated to the state of the macroeconomy, it cannot be easily integrated into a macro-driven stress test in any deep or structural way.

\textsuperscript{29} As we discuss in section 4 below in more detail, starting in 2017, the Federal Reserve has eliminated the qualitative requirements for the smaller banking institutions previously subjected to it.
of US$10b to conduct company-run stress tests.\textsuperscript{30} Covered institutions submit their stress tests to their primary regulators for review and are required to disclose the results.\textsuperscript{31} While DFAST is quite similar to CCAR and thus serves many of the same core purposes, the tests are applied to a much larger set of firms and are not directly used to set capital requirements, though they are used to inform supervisory oversight and enforcement actions. Figure 1 describes the evolution of the share of assets belonging to banks undergoing stress testing in total banking assets. This share has increased from about 60\% under SCAP to about 85\% under the CCAR and DFAST as of Q2 2017.

Finally, though not required by any specific legislation, the NCUA began stress testing credit unions with assets exceeding US$10b in 2015.\textsuperscript{32} The NCUA provides a description of the framework for carrying out these tests, but their results are not publicly reported.

Also summarized in Table 1 is the EU-wide stress testing program.\textsuperscript{33} The first EU-wide stress test, conducted by the Committee of European Banking Supervisors (CEBS) in 2009, covered only 22 cross-border banks and disclosed very little information. As such, these early stress tests were of relatively limited financial stability value. Subsequent stress tests, conducted by the CEBS’s successor agencies – the European Banking Authority (EBA) followed by the European Central Bank (ECB) – have applied to a wider number of banks and have been much more thoroughly disclosed. In broad terms, the EU-wide stress tests are similar to their US counterparts, emphasizing credit and market risk. Recent scenarios have featured some elements of sovereign risk, though they fall short of fully assessing the impacts of sovereign default.\textsuperscript{34} The tests are coordinated by the EU banking regulators but are largely overseen by country-level regulators. Unlike many other countries, the EU-wide stress tests do not occur every year: no tests were conducted in 2015 or 2017, and the next test will be in 2018. In addition to any direct financial stability benefits, EU-wide stress testing fosters testing by national authorities, such as the Bank of England.\textsuperscript{35} These national-level tests are often structured to roughly coincide with the EU-wide stress tests, in terms of timing, model structures and variables, but often include additional scenarios, banks or risks assessed.

Stress testing in the UK has evolved at a more rapid pace than in many other countries. The 2014 stress test by the Bank of England was conducted jointly with the EU-wide stress test and used the same scenarios with some modifications to tailor them to risks in the UK. Since then, annual UK stress testing has been largely independent of the EU-wide stress tests, partly because the EU-wide stress tests have not consistently occurred every year.

The 2015 stress tests were largely comparable to the 2014 tests, though the scenario was designed solely by UK supervisory authorities and emphasized broader and global stresses.\textsuperscript{36} Beginning in 2016, the Bank of England implemented a novel approach to generating scenarios. The main stress scenario is referred to as the “annual cyclical scenario” (ACS) and is intended to be countercyclical: in booms, when downside risks are judged to be high, the scenarios are very severe; in downturns, the scenarios are less severe. More generally, the scenarios are

\textsuperscript{30} FRB (2014b, 2015b)
\textsuperscript{31} The DFAST program is newer than the CCAR, having started only in 2013. The first year of DFAST covered only CCAR firms. Coverage of institutions and disclosure requirements have been phased in to match those described above and required by law. In addition, the conduct and disclosure of DFAST is more fragmented than the CCAR, reflecting the fact that it is overseen by multiple supervisory agencies.
\textsuperscript{32} NCUA (2014, 2015, 2017)
\textsuperscript{33} CEBS (2009, 2010); EBA (2011a, 2014, 2016a,b); ECB (2014)
\textsuperscript{34} Schuermann (2014) and Goldstein (2014) argue that the approach to stressing sovereign exposures in the EU-wide stress tests has not been sufficiently rigorous.
\textsuperscript{35} Bank of England (2014a, 2014b)
\textsuperscript{36} Bank of England (2015a, 2015b)
meant not to explore new areas of emerging risk, but rather to serve as a stable benchmark across years. Complementing these relatively rigid annual scenarios, the Bank of England also conducts a test every other year using a “biennial exploratory scenario” (BES). The ACS-based tests began in 2016 and involve a pass/fail outcome for each bank. The BES-based tests began in 2017, are not “graded” on a pass or fail basis, and their results are not released.\(^{37}\)

Importantly, none of these stress testing regimes include stress tests of bank liquidity, which is an area of particular concern that we discuss below. Several central banks do conduct separate liquidity stress tests, though. For example, the Federal Reserve conducts the Comprehensive Liquidity Assessment and Review (CLAR),\(^ {38}\) though the data, methodology and results are confidential. In addition, many central banks have recently developed stress test models (typically fairly aggregated models) that include liquidity and solvency interaction. However, these models are generally separate from the primary, large-scale stress tests whose results are publicized and studied by private sector analysts. The models are so far mainly used for internal policy analysis. In general, liquidity stress testing is too often conducted confidentially and separately from other forms of stress testing, which is regrettable given the substantial interaction of liquidity and solvency risks. In a recent policy speech, Tarullo has recognized the calls for the liquidity and capital stress testing frameworks to be merged together and acknowledged the lack of progress on that policy front. More work in that direction is urgently needed.\(^ {39}\)

Even though the Financial CHOICE Act would imply large changes in the regulatory landscape for large financial institutions in the US, the belief that this bill will be enacted into law in its current form is generally lacking among lawmakers, pundits and policymakers.\(^ {41}\) In a June 2017 report, the US Treasury called for similar changes to the stress testing practice, in particular recommending the following changes: (1) raise the participation threshold from US$10b to US$50b in total assets, while potentially exempting institutions whose business models do not warrant stress testing; (2) eliminate the midyear DFAST cycle and reduce the number of supervisory scenarios from three to two (excluding adverse and retaining baseline and severely adverse); (3) conduct the CCAR exercise every two years, matching the duration of the stress scenarios; (4) provide off-ramp exemptions from stress testing and other regulatory requirements for banks maintaining sufficiently high capital levels; (5) subject stress testing and capital planning frameworks to regular public notice and comment processes to increase transparency and accountability; and (6) eliminate the CCAR qualitative exercise as the sole objection to a capital plan owing to its subjectivity and lack of transparency. The Treasury report also contains a large number of recommendations that are not directly related to stress testing.\(^ {42}\)

The administration change in the US in January 2017 portends potentially significant changes for the current financial regulatory landscape. Both the new administration and Congress have shown interest in reducing the scope of financial regulations, including the extent of the mandatory stress testing exercise. In June 2017, the House of Representatives passed the Financial CHOICE Act that either weakens or reverses several DFA regulations. The bill, among other things, repeals the Orderly Liquidation Authority and replaces it with a new chapter in the Bankruptcy code; repeals the authority of the FSOC to designate firms as systemically important financial institutions (SIFIs); and proposes to rely on the leverage ratio as the measure for assessing capital adequacy, and offers those institutions with a leverage ratio of 10% exemptions from many regulatory requirements.\(^ {40}\)


\(^{38}\) Tarullo (2014a)

\(^{39}\) Tarullo (2016)

\(^{40}\) Those financial institutions deemed as qualifying, by maintaining a leverage ratio above 10% will be exempt, according to Section 602 of the Financial CHOICE Act from (among others):

\(^{41}\) E.g., Rappeport (2017)

While the direction of the proposed changes is clearly aimed at reducing the regulatory burden, as of the writing of this article, the likelihood of their implementation remains uncertain. In section 4, we discuss some of these policy proposals whose likely impact has become reasonably well understood among policymakers and practitioners. We lay the foundations for that discussion by providing a more technical overview of the stress testing practice in sections 2.2 and 3.

2.2 Current modeling approaches
With the historical context in mind, we now turn to the technical aspects of stress test implementation. Today, there are two general modeling paradigms for conducting capital adequacy stress tests: direct and reverse stress testing. We discuss these methods, dedicating the most attention to direct stress testing, since it has been the main method used for regulatory compliance purposes.

2.3 Direct stress testing
Direct stress testing, which includes the stress testing regimes developed since the crisis, is essentially a three-step process. First, empirical relationships between a banking variable and exogenous stressors, such as macroeconomic variables or financial indexes, are estimated. Second, a stress scenario for the exogenous stressors is produced by a researcher or risk manager, or delivered by a regulator or third-party consultant. The scenario may replicate a historical episode or may be hypothetical, intended to assess a particular risk. Third, the stress scenario is applied to the empirical models to understand the effect on the banking variable of interest. The process may be repeated for several banking variables, and additional and possibly assumption-based calculations may be performed to obtain results for final measures, such as capitalization levels.

Banks and regulators both follow this general procedure in their stress testing activities, though with a high degree of heterogeneity in the particular banking variables modeled, the modeling approaches and even the scenarios considered.

Two empirical methods fall under the direct stress testing umbrella: top-down and bottom-up. In the top-down approach, the empirical relationship between a banking variable and an exogenous stressor is assumed at the portfolio level of low granularity. A bank's charge-off rates on an entire portfolio of loans may depend on the unemployment rate, for example.44 In the bottom-up approach, the empirical relationship is estimated at the highest possible level of granularity of a banking variable. For example, this model may capture the effect of the local unemployment rate on the probability of borrower default for individual loans. The main advantages of the top-down approach are that it requires only readily available bank-level data and that it allows for the horizontal comparison of stress scenarios for many institutions. The main disadvantage of the top-down approach – and the main advantage of the bottom-up approach – is that idiosyncratic aspects of a bank may be lost at the aggregate level but may be more easily modeled when using granular data. Importantly, under both of the direct stress testing approaches, the number of stress scenarios is typically small, which may limit the sources of risk considered, both for individual banks and for the financial system as a whole.

Much of the academic literature on stress testing has focused on the direct top-down approach, while some of the stress test regimes discussed above, such as CCAR, are bottom-up, creating a substantive gap between academic research and supervisory implementation of stress testing. This disparity likely exists because macroeconomic and portfolio-level data is publicly available, whereas loan-level data, necessary for bottom-up modeling, is available only from commercial sources and within supervisory agencies. Several authors provide examples of top-down models that have attempted to replicate regulatory analyses.45 While Hirtle et al. assume that all banks are equally sensitive to exogenous stressors, Kapinos and Mitnik and Covas et al. describe ways of modeling heterogeneity of this sensitivity.46 The latter is important because it yields stronger banking distress for a given macroeconomic stress scenario. As was mentioned earlier, Frame et al. describe the bottom-up model used by the OFHEO to stress test the GSEs in the run-up to the most recent crisis.47

43 In this paper, we follow the US definitions for these concepts, whereas in Europe “top-down” may refer to stress testing conducted by regulators and “bottom-up” by banks. Burrows et al. (2012) suggest that direct stress testing may include other options along the tradeoff between theoretical coherence and empirical detail, such as dynamic stochastic general equilibrium (DSGE) models. While potentially useful, we are not aware of stress tests being incorporated into these models; hence, we do not discuss them in this paper.
44 Sangha and Lin (2013) critique such direct stress testing models from the risk management perspective.
45 Hirtle et al. (2016) and Kapinos and Mitnik (2016).
46 Hirtle et al. (2016); Kapinos and Mitnik (2016); Covas et al. (2014).
47 Frame et al. (2015).
Relatedly, integrated liquidity stress testing is an idea that has become relatively developed in the academic (and, typically, top-down) literature, but whose practical implementation is incomplete.

2.4 Reverse stress testing
An alternative paradigm to the direct approaches is reverse stress testing. The main objective is to determine the range of exogenous driver scenarios that may “break the bank.” Alternative methods for conducting reverse stress tests have been provided. While it may be difficult to envision a significant regulatory application of the reverse stress testing paradigm, it is likely to be a useful tool for identifying the types of stress scenarios that may be used for regulatory direct stress testing purposes.

3. Continuing technical challenges
Although the theoretical and empirical research has advanced modeling various income statement and balance sheet items and projecting them forward under hypothetical stress scenarios, several technical and qualitative challenges remain. Further efforts to address them must be made for stress testing to realize its full potential as a financial stability tool. Below, we describe some of the more significant issues in order to motivate future work.

First, stress testing currently stops short of analyzing systemic stress. Although it analyzes the impact of a macro scenario on each firm’s capitalization, mainly through credit losses and income, stress testing does not allow the firm’s results to affect other firms or the real economy. This approach leaves several important questions open. Suppose a firm that does poorly in the stress scenario actually failed. What would happen to the other firms? Would they facet liquidity runs? Would the macroeconomic scenario worsen? If so, would this further affect financial stability? Even barring a major bank failure, what effects does lower capitalization have on an institution’s interaction with other players? How might banks react in an effort to mitigate their risk? As banks become less capitalized without breaching regulatory thresholds, do they become more likely merger targets, which may lead to higher concentration and increase systemic risk? These questions regarding the feedback between tested institutions’ results, and between their results and the macro scenario must be answered in order to understand the full system-wide effects of a particular stress scenario.

Focusing on the transmission of stress among tested firms, recent experience suggests that transmission largely occurs through liquidity channels. When firms face steep losses and declining capital, they frequently lose funding sources, precipitating fire sales of assets. These fire sales, in turn, depress asset (and collateral) prices for all institutions and drive liquidity hoarding. Even in the absence of fire sales, loss of confidence in either firm or market liquidity may lead to rapidly rising collateral haircuts and loss of funding capacity or liquidity. In this manner, solvency and liquidity stress is quickly transmitted to other firms. Indeed, recent empirical evidence points to strong links between bank liquidity positions and the ability of banks to meet capitalization requirements. For example, banks experience liquidity problems when markets expect their insolvency in a crisis. Similarly, Schmieder et al. provide simulation-based evidence for a nonlinear relationship between liquidity and solvency risks triggered by an adverse funding shock. Finally, on average, large US and European banks decrease their regulatory capital ratios when faced with higher illiquidity, whereas small banks strengthen their solvency standards in this situation.

In addition to this empirical evidence, policymakers clearly believed during the 2008-09 crisis that the risks posed by illiquidity at least rivaled the risks stemming from bank insolvency. This belief is supported by the scale of the official liquidity intervention during the crisis. The Federal Reserve reports that the peak value for the total of its emergency liquidity facilities in the US during the crisis was US$1.5t. It is worth noting that this amount does not include several prominent liquidity programs such as the FDIC’s Temporary Liquidity Guarantee Program, the Treasury’s Temporary Guarantee Program for money market mutual funds, or deposit insurance, which can obviously...
Stress testing banks: whence and whither?

be viewed as a form of liquidity provision. In comparison, official solvency interventions were comparatively small: the combined size of the Treasury’s Troubled Asset Relief Program’s components directed at recapitalizing the banking industry, the Capital Purchase Program and the Targeted Investment Program, was US$245bn. Moreover, the book capitalization levels of banks that either failed or underwent assisted takeovers during the financial crisis were typically well above regulatory minimums in the last quarter before their failure.55 This raises the concern that current stress tests, which focus almost exclusively on book equity, may yield relatively benign outcomes by not explicitly considering liquidity risk. In part, this may reflect the fact that book and market equity values tend to diverge during times of stress, and that declining market capitalization drives liquidity stress. Thus, incorporating liquidity and systemic stress may require a more thorough treatment of market equity.

The systemic links between solvency and liquidity risk have prompted some policymakers and analysts to call for a combined approach to the regulation of liquidity and capital adequacy at banks.56 However, liquidity risk modeling remains underdeveloped, particularly compared to solvency tests driven by macro stress scenarios.57 In many stress testing exercises, a puzzling disconnect remains between the related concepts of liquidity and capital adequacy stress testing. The clearest examples of this disconnect are the two separate exercises run by the Federal Reserve for the US BHCS: CCAR and its liquidity counterpart, CLAR. While much of the information regarding the former is public, the latter remains shrouded in mystery.

To be fair, some regulators have advanced models that attempt to address these issues, though the models have yet to be integrated into the primary stress testing regimes.58 Several models have emphasized bank heterogeneity and systemic implications of stress; for example, in the context of the Bank of England’s Risk Assessment Model for Systemic Institutions (RAMSI).59 RAMSI is a top-down stress testing model that features bank heterogeneity and systemic risk driven by shocks to funding liquidity. It follows earlier work on the Austrian Central Bank’s model.60 The Bank of Canada’s MacroFinancial Risk Assessment Framework (MFRAF) has also been discussed.61 As a final example of this literature, Van den End provides a related example of a liquidity stress test where bank heterogeneity and systemic risk exacerbate the level of stress experienced by individual banks for the Dutch banking system.62 Finally, Basel III bank regulations include liquidity regulation, but given the above considerations, we maintain that such requirements should be integrated with solvency regulation in a forward-looking and systemic manner.63 Thus, in spite of the empirical links and the developing theoretical basis discussed here, many regulatory exercises still treat liquidity and capital adequacy stress tests as separate. We contend that this is perhaps the most critical area for future work.

Given the complexity of the models and sensitivity of firm-level data required to integrate capital and liquidity stress testing, future assessments of systemic risk may need to be conducted only by regulators. At present, many stress testing regimes include (or wholly comprise) stress tests conducted by banks and reviewed or benchmarked by regulators. However, capturing systemic feedback at a granular level and in a consistent manner across firms requires such excessive data volumes and extraordinarily complex modeling architecture as to make the exercise very challenging for individual institutions. Instead, regulators can build and run models using data obtained from the banks and make assessment regarding the levels of systemic risk. Indeed, this is the general direction taken in the above-cited papers on such testing.

A second technical issue concerns the fact that some of the key quantities that are essential for obtaining measures of bank capitalization are assumed, rather than modeled explicitly. Moreover, these assumptions can significantly affect the results, but their impact is not fully understood. Specifically, paths for the size of the balance sheet (mainly asset growth) are important

55 Kapinos and Mitnik (2016)
56 For instance, Tarullo (2013, 2016)
57 BCBS (2013a)
58 BCBS (2013b)
59 Aikman et al. (2009); Alessandri et al. (2009)
60 Discussed in Elsinger et al. (2006) and DeHaan (2006)
61 Gauthier et al. (2010); Anand et al. (2014)
63 Basel III includes the liquidity coverage ratio (LCR) and the net stable funding ratio (NSFR). The LCR rule requires that banks hold enough high-quality liquid assets to fund 30 days of liquidity needs under a hypothetical liquidity stress scenario. Thus, while the LCR can be thought of as a liquidity stress test, it is not integrated with solvency stress tests. Structured more like traditional Basel capital requirements, the NSFR requires banks to maintain stable funding (defined as a function of the contractual maturity and estimated stability of liabilities) at least equal to the bank’s funding needs. These needs are determined by applying weights (analogous to risk weights) to assets and off-balance-sheet exposures. See BCBS (2013c, 2014) for more on the LCR and NSFR.
for determining capital ratios but are difficult to credibly model as functions of macroeconomic scenarios. Such quantities often depend on bank-level management decisions, making consistent statistical modeling across firms, or even within a given firm, especially challenging. As a result, for regulatory stress tests, these quantities are often set by simple assumptions, such as zero asset growth.\textsuperscript{64} US stress tests typically allow assets to grow at a rate chosen by the banks, while EU-wide tests assume no asset growth. The Central Bank of Ireland's Financial Measures Programme\textsuperscript{65} and OFHEO's stress testing of the GSEs even assumed that loans were not replaced as they matured.

The extent to which these varying assumptions affect the results or their comparability is not fully understood. Sensitivity analysis on asset growth assumptions in top-down models shows that different assumptions have sizeable effects on capital projections.\textsuperscript{66} Moreover, as noted in section 2.1, Frame et al. finds that OFHEO's decision to not assume any new business during the stress scenario caused a substantial understatement of losses, because the probability of default (PD) is especially high during the first few years of a mortgage's life cycle.\textsuperscript{67} This suggests that bottom-up stress tests (such as OFHEO's), where individual assets are modeled, may be especially sensitive to asset growth assumptions. However, reflecting the limited published literature on bottom-up stress testing, an issue we return to shortly, Frame et al. is the only study that assesses this sensitivity in such models. Complicating the issue further, stress testing results may be sensitive to the use of total versus risk-weighted assets, which need not grow at the same rate.\textsuperscript{68}

Third, some seemingly innocuous modeling choices may affect capital projections under stress in unanticipated ways. For example, the standard that has emerged in the top-down modeling literature is to use income quantities as shares of assets and chargeoff quantities as shares of loans. Both sets of assumptions make sense: It would be econometrically flawed to use constantly growing dollar quantities in a regression framework; normalizing income by assets makes the quantities from banks of different size comparable; and chargeoff rates have been the standard metric for the quality of loans. However, they make the modeled ratios insensitive to asset and loan growth assumptions in the stress projection period. One cannot sensibly include the denominator of the ratio as an explanatory variable, so the loss rate will be insensitive to asset growth assumptions. This implicitly assumes that the portfolio loan characteristics remain constant in the face of any asset growth rate, which appears implausible. For example, for a bank to grow a portfolio more quickly, it may need to lower underwriting standards, which would be expected to raise loss rates.

Fourth, the focus of the academic stress testing literature has been on the top-down models that, while useful as a first approximation, have severe limitations. Because they use data at low levels of granularity, the top-down models cannot capture bank portfolio drift or the appearance of new types of assets. While the data necessary for use in top-down models is publicly available, the granular data for the bottom-up approach is not.\textsuperscript{69} As a result, a substantial disconnect exists between published studies on stress testing (which generally rely on public data) and the actual practice of state-of-the-art bottom-up stress testing (which generally relies on richer, non-public data). Regulators may need to develop a representative, de-identified (and possibly synthetic), publicly available dataset at the highest level of granularity to spur research activity in this area and yield more informative stress testing models.\textsuperscript{70}

Finally, some institutional issues may need to be addressed to make the stress testing exercise comprehensive and policy-relevant. They primarily concern modeling other sectors related to banking, global and national macroeconomic conditions, and the feedback mechanisms from the banking sector to these entities. An important sector related to the regulated banking industry is shadow banking. Early theoretical work suggests that increasing the regulatory scrutiny on the formal banking sector may increase the level of risk, including systemic risk, in

\begin{itemize}
  \item \textsuperscript{64} Similarly stylized assumptions are often applied to dividend payouts, and stock issuance and repurchases.
  \item \textsuperscript{65} Central Bank of Ireland (2011)
  \item \textsuperscript{66} Hurtle et al. (2016); Kapinos and Mitnik (2016)
  \item \textsuperscript{67} Frame et al. (2015)
  \item \textsuperscript{68} Acharya et al. (2014)
  \item \textsuperscript{69} Currently, micro- and loan-level data is available from commercial sources (e.g., CoreLogic for mortgages) and data consortia among banks (e.g., SAS OpRisk VaR and ORX). However, this data is not particularly useful for stress test modeling because it does not allow for analyzing a bank's portfolio and, therefore, its risk profile.
  \item \textsuperscript{70} Development of such a dataset is likely to be technically difficult, as it will entail addressing issues such as the heterogeneity of credit ratings systems for loans as well as the more obvious confidentiality concerns. Providing more detailed portfolio-level data may be an intermediate step that is more easily implementable, if less informative. The dataset may be based on the annual submissions for the stress testing exercise and not entail additional regulatory burden on financial institutions.
\end{itemize}
the shadow banking industry. Further work along these lines, including the feedback effects from the unregulated to the regulated financial sector, is urgently needed.

4. Critiques of stress testing
Abstraction from the technical concerns discussed above, forceful criticism argues against the fundamental appropriateness of stress testing. The argument is that stress testing, or at least the form of regulatory stress testing performed since the 2008-09 crisis, should not be conducted.

We consider in this section the most salient of these critiques. While some reflect valid concerns and may warrant additional analysis going forward, we conclude that stress testing remains an appropriate and useful regulatory tool.

The first critique concerns the possibility that stress testing represents an implicit commitment by the regulatory authorities to backstop firms that have previously passed stress tests. Market participants may interpret a “pass” as an official endorsement of the health of an institution. If the firm turns out to be unhealthy, participants may expect the regulator to take responsibility. This concern is most clearly present in stress testing during financial crises, when the test may be explicitly aimed at assessing the size of capital shortfalls in preparation for a government recapitalization scheme. Indeed, Schuermann suggests that one reason for the markets’ positive response to SCAP may have been that markets interpreted the stress tests as an implicit assurance that regulators would in some way absorb losses in excess of the stress test estimates. Flannery makes a similar point. This critique has also been extended to stress testing outside financial crises. For example, Kupiec argues that stress testing (and postcrisis financial regulation, more broadly) amounts to regulators operating financial institutions.

He goes on to argue that since regulators are effectively operating the institutions, they may face difficulty in allowing equity holders and creditors to take losses. This critique is accentuated by the potential for financial institutions to “game” the system, designing new products that have low historical sensitivity to macroeconomic variables, yet will suffer losses in a crisis. By designing such products to artificially lower their stress tested losses, financial institutions may have gained another method to quietly extend their access to implicit government guarantees.

We believe stress testing need not increase the likelihood of future bank bailouts, for several reasons. First, additional regulatory provisions, such as the DFA requirement for BHCs with more than US$50b in assets to have an updated “living will” – the company’s strategy for a rapid and orderly resolution under the Bankruptcy Code – substantially mitigate these concerns. In addition to serving as contingency plans, these living wills are useful tools for regulators to encourage firms to conduct operations in a manner that facilitates resolution through bankruptcy and reduces the systemic implications of their failure. Second, stress testing is most widely used in the regulation of firms thought to be systemically important. As such, many market participants already expect that the firms have some form of implicit government support. To the extent that these expectations are true, enhanced supervision, perhaps in the form of stress testing, is warranted to protect the government and society from costs arising from resulting market failures. Third, the worst-case scenario for a bank participating in a stress test exercise is that it learns nothing new about its risks; the best-case scenario is that the regulator helps to identify possible blind spots in the bank’s risk management strategy. As we elaborate below, stress tests should be designed with that goal in mind.

A closely related concern is that stress testing exposes regulators to reputational risk. If markets perceive that regulators give a particular firm or financial system a passing grade, only for it to fail soon thereafter, the regulators’ reputations may be compromised. Indeed, this seems to have occurred with the EU-wide stress tests of banks in both

71 Kolm (2015)
72 This can happen even in the absence of extending regulation to the shadow banking sector, as suggested by Fischer (2015), for example.
73 Importantly, at least some large bank executives seem to disagree. Dimon (2015), in his annual letter to JPMorgan Chase shareholders, described CCAR as “good for our industry in that it clearly demonstrates the ability of each and every bank to be properly capitalized, even after an extremely difficult environment” and the stress test’s key assumptions as “appropriately conservative.”
74 Practitioners frequently refer to this concern as creating a “safe harbor” for banks that enables them to thwart the intended purpose of stress testing regulation.
75 Schuermann (2014); Flannery (2013)
76 Kupiec (2014)
77 This requirement is mandated by §165(d) of the DFA and also applies to nonbank financial firms designated by the FSOC as systemically important. In addition, the FDIC requires insured depository institutions (rather than BHCs) with assets over $50 billion to submit comparable resolution plans.
78 Hirtle and Lehnert (2015)
Ireland and Spain, and it was detrimental to the reputation of European banking authorities.\textsuperscript{79} We see this as an argument for effective and accountable stress testing rather than an indictment of the practice. Undoubtedly, the choice of scenario severity involves a difficult trade-off akin to the type I versus type II errors in statistics. On one hand, regulators need to be mindful of the large costs (mostly borne by shareholders) associated with falsely determining that a financial institution may be insolvent under stress. Subjecting banks to implausibly severe stress is likely to reduce their operational efficiency and make them too cautious to fail.\textsuperscript{80} On the other hand, regulators need to balance those costs with the potential costs to taxpayers (or insurance funds), as well as their own reputation, from failure of an institution that has passed a stress test. Treading this fine line, however, is inherent to the nature of banking regulation in general and not to stress testing in particular.

A third critique is that the requirement to conduct stress testing represents an unreasonable burden on financial institutions, especially smaller and less complex institutions.\textsuperscript{81} Fundamentally, this critique relies on a (perhaps unstated) value judgment that the benefits of stress testing in terms of financial stability are not worth the costs to banks or regulators. We argue that this value judgment should be formally stated. In that spirit, we contend that the net benefit of stress testing is almost surely positive for the largest institutions and declines for successively smaller and less sophisticated banks. At some threshold, the costs outweigh the benefits. The largest institutions pose the greatest systemic risk, require advanced risk management practices consistent with their high degree of complexity, and have the resources to undertake stress tests. In contrast, midsize and smaller institutions are less complex. Because they have limited capabilities to conduct stress tests, many midsize and smaller banks rely on external consultants to conduct tests, both reducing their value in encouraging improved risk management and exposing them to consultant-driven, correlated model risk. Moreover, midsize regional banks and other smaller banks are likely less sensitive to national macroeconomic and financial conditions described in regulatory scenarios, as their stability depends more on local conditions or the health of individual counterparties. Current supervisory requirements and expectations for the various stress test regimes are consistent with our assessment of the net benefits: they are graduated to reduce the burden on small institutions, and the smallest are not required to conduct tests. From this perspective, this critique is essentially a debate about precisely where the graduations or thresholds should fall, rather than about the practice of stress testing per se.\textsuperscript{82} Debate on this topic is ongoing among the senior leadership of US regulatory agencies responsible for stress testing. For example, Tarullo argues that a higher asset threshold should be used to designate those banks required to participate in CCAR.\textsuperscript{83} The recent report by the US Treasury discussed in section 2.1 has formalized this proposal by calling for the exemption of institutions with assets below $50b from stress testing exercises.\textsuperscript{84} Furthermore, the Federal Reserve Board has finalized a rule for the 2017 CCAR cycle that “removes the qualitative assessment of CCAR for large and noncomplex firms, which are BHCs and US intermediate holding companies of foreign banking organizations with total consolidated assets between US$50b and US$250b, total nonbank assets of less than US$75b, and that are not identified as global systemically important banks (GSIBs).”\textsuperscript{85} In general, while we agree that for smaller institutions the costs of stress testing may outweigh the benefits, we highlight that in periods of banking consolidation, which typically coincide with financial and macroeconomic distress, smaller institutions may well contribute to the systemic risk of their larger acquirers. Hence, further research to assess whether the US’s current “dual system” of stress testing by bank size can be improved upon seems warranted.

More broadly, this critique highlights the value of additional analyses characterizing the costs and benefits of stress testing. Such research is scarce but would provide rigor to the discussion on the appropriate breadth and depth of stress testing.\textsuperscript{86} Available estimates of the cost of the DFA regulation suggest that bank costs related to stress testing are small.\textsuperscript{87} It is necessary to

\begin{itemize}
\item \textsuperscript{79} Schuermann (2014); Central Bank of Ireland (2012)
\item \textsuperscript{80} Smith (2014)
\item \textsuperscript{81} Deseret News (2013); McLannahan (2015); US Treasury (2017)
\item \textsuperscript{82} Recall that figure 1 provides a visual description of the marginal asset coverage by DFAST over CCAR.
\item \textsuperscript{83} Tarullo (2014b)
\item \textsuperscript{84} US Treasury (2017)
\item \textsuperscript{85} FRB (2017c)
\item \textsuperscript{86} Cumming (2015)
\item \textsuperscript{87} S&P/PRS (2012) estimates that compliance with all of the DFA provisions would reduce the pretax earnings of Bank of America, Citigroup, Goldman Sachs, JPMorgan Chase, Morgan Stanley, PNC Financial Services, US Bancorp and Wells Fargo by $22 to $34 billion annually. Stress testing would fall under the higher regulatory compliance expenses whose upper
\end{itemize}
recognize that stress testing does require the resources of banks and that those resources must be leveraged to maximum effect as a risk management tool, rather than for a mere compliance exercise. However, many of the benefits of stress testing, such as improved data collection and risk management, are hard to quantify. Moreover, these benefits are relevant even to smaller financial institutions. Insofar as stress testing reduces asymmetric information and mitigates systemic risk, smaller banks receive an additional layer of insulation from adverse industry-wide shocks. Future work to quantify these benefits is highly advisable.

Further, Kupiec argues that stress testing represents dangerous overreach into bank management, forcing institutions to submit to an opaque process whose outcome-generating model is not known to the banking industry.88 Making the stress testing models transparent, however, is likely to generate the same problems experienced by OFHEO, which conducted stress tests under maximum transparency. Further, such transparency is likely to maximize so-called “model risk,” where both the regulators and the entire industry are driven by the results of just one particular model, preventing the identification of blind spots using techniques such as reverse stress testing. Maintaining a certain degree of process opacity, therefore, appears to be appropriate. Moreover, regulators should remain vigilant that the industry not settle for a few consultant-driven stress testing models known to be given a pass by regulators, thus institutionalizing model risk.

Another frequently made critique argues that historical data will not be informative in predicting the next crisis. This shortcoming of historical data may arise from the fact that new products will be developed with little history; loan underwriting standards may change; or historical data may be contaminated with events that may not repeat, such as a government capital and liquidity support; among other things. In addition, stressful events are infrequent, compounded by the fact that institutions and their interconnections, positions and leverage change over time.89 As a result of these limitations, stress testing may provide a false sense of security without actually addressing risks. This is a legitimate concern, and addressing it calls for granular, loan-level modeling to reduce the model risks posed by new products or business practices. It also calls for a stress testing program that evolves over time and avoids converging to a model monoculture.

A related, fundamental concern involves the implementation of stress testing as a tool to attenuate disaster myopia. We argue that stress testing is useful in this regard because it provides a rigorous structure within which market participants can consider unlikely adverse events. For this to be true, regulators must avoid becoming complacent and myopic, which will require an ongoing commitment to rigorous modeling and strenuous scenario selection. The OFHEO experience suggests that maintaining up-to-date models and providing sufficiently stressful scenarios are critical to the success of stress testing. These activities should not simply “fight the last war,” but should instead proactively study emerging risks. So far, regulators’ performance in this regard is mixed.90 While modeling techniques are advancing, the importance of scenario design is not widely emphasized. For example, the most stressful scenarios in US stress tests have borne substantial resemblance to the 2008–09 crisis. We hope that the scenarios tested will become more diverse and exploratory as the crisis fades into history. Also noted above, the EU-wide stress tests have stopped short of including an actual sovereign default in their stress scenarios. This has harmed the usefulness of the tests and likely points to significant political limitations within the EU banking regulatory system. Going forward, regulators must continue to design stressful scenarios, perhaps relying on reverse stress testing techniques to uncover the emerging sources of risk. Continued effort is also required to ensure that both models and scenarios retain some heterogeneity and do not progress to a state of monoculture.

A final, fundamental concern is that financial shocks, rather than macroeconomic shocks, may represent the greater risk to financial stability. In short, the concern is that stress testing is looking in the wrong place. For the most part, current stress tests are driven primarily by a macroeconomic scenario, which is then translated into financial distress. An alternative narrative for financial stress might be that it originates as a shock to the financial system, which then causes macroeconomic stress. Whether macroeconomic or financial shocks drive financial cycles is the subject of a large literature beyond the scope of this paper.91

---

88 Kupiec (2014)
89 Bookstaber et al. (2014); Jacobides et al. (2014)
90 Hittle and Lehnert (2015) and Schuermann (2014) suggest that if sufficiently detailed data on the banks is released, market participants can effectively conduct their own stress tests using their own scenarios.
91 This question is closely related to a large literature that includes Bernanke et al. (1999).
This preliminary discussion, however, suggests that a more comprehensive modeling approach is strongly warranted.

5. Concluding remarks and recommendations for the future of stress testing

This paper has provided an overview of the theory and practice of stress testing banks. On balance, we argue that stress testing exercises are a useful addition to the toolkit of bank risk managers and their regulators. From a supervisory perspective, stress testing serves as a forward-looking complement to Basel regulations that can more effectively capture systemic risk while encouraging more holistic risk management.

From a more fundamental perspective, stress testing may be viewed as a means of reducing disaster myopia\[92] and preventing financial fragility that results in a so-called Minsky moment of financial collapse and a subsequent severe macroeconomic slump.

Regulatory implementation of stress testing in the wake of the most recent crisis has arguably already delivered substantial benefits to the banking industry and the public. Wider availability of banking data and its more rigorous scrutiny have curbed bank opaqueness, improved governance and led to better risk management practices than those that were in place before the 2008–09 crisis. Equipped with more detailed data and improved risk modeling techniques, many bank executives are now forced to think through forward-looking risk management issues; an analytical framework addressing them was largely absent prior to the crisis.

While substantial progress has been made, some important issues remain unresolved. First, public availability of some form of granular data needs to be expanded to give researchers the opportunity to develop more precise stress testing models. Also, given the strong empirical link between liquidity stress and capital adequacy, we recommend that these two types of stress testing exercises be merged. In addition, modeling the interaction of banks with other banks and financial institutions, including the shadow banking sector and, ultimately, the broader macroeconomy, is likely to yield important insights into additional sources of liquidity and capital adequacy stress.

Further, the reverse stress testing paradigm needs to encompass banking, financial, macroeconomic and international sources of extreme yet plausible shocks to identify the potential blind spots in bank risk management. Finally, research on the benefits and costs of stress testing may help to better tailor this regulatory tool. While this program is undoubtedly quite challenging, it promises to reap substantial rewards.

\[92\] In the sense of Guttentag and Herring (1986)
Stress testing banks: whence and whither?

References


Stress testing banks: whence and whither?


Minsky, Hyman (1986), Stabilizing an Unstable Economy, New Haven: Yale University Press.


How to capture macro-financial spillover effects in stress tests

Heiko Hesse
Directorate General for Economic and Financial Affairs (DG ECFIN), Italy desk, European Commission; Strategy, Policy and Review Department, International Monetary Fund (IMF)

Ferhan Salman
Middle East and Central Asia Department, International Monetary Fund (IMF)

Christian Schmieder
Member of Secretariat, Financial Stability Board, Basel

1 We thank Eugenio Cerruti, Martin Cihak, Amadou Sy and participants at a Banque de France conference (September 2012) as well as those at a meeting in May 2013 of the Liquidity Stress testing Subgroup of the Basel Committee on Banking Supervision’s Research Task Force for helpful comments and suggestions. The views expressed in this article are those of the authors and do not necessarily represent those of the IMF or IMF policy. This article was previously published as an IMF working paper. This is an updated and edited version, published with permission.
Abstract
One of the challenges of financial stability analysis and bank stress testing is how to establish scenarios with meaningful macro-financial linkages, i.e., taking into account spillover effects and other forms of contagion. We propose an approach to simulate the potential impact of spillover effects based on the “traditional” design of macroeconomic stress tests. Specifically, we examine spillover effects observed during the financial crisis and simulate their impact on banks’ liquidity and capital positions. The outcome suggests that spillover effects have a highly nonlinear impact on bank soundness, in terms of both liquidity and solvency.
I. Introduction

Stress testing has received a lot of attention in recent years, which has spurred numerous conceptual developments. However, overarching approaches to establish macro-financial linkages, explicitly to capture the nonlinearity of shocks (originating from spillover effects and other types of contagion), are still evolving. Such linkages have seen particularly significant growth during the past decade and are, therefore, an important dimension to be captured by meaningful empirical analysis. This paper uses a case study to focus on the design of stress tests which helps to capture spillover effects and demonstrate their potential impact.

The first part of the paper deals with the establishment of macro-financial scenarios that are explicitly informed by spillover effects. Scenario design for macroeconomic stress tests is typically based on an “indirect approach” (Figure 1): first, economic and financial variables are estimated, conditional on a macroeconomic scenario, and then the trajectories of the economic and financial variables are translated into bank solvency and liquidity measures on the basis of the so-called “satellite” or “auxiliary” models. Three approaches have commonly been used to predict economic and financial variables under stress: (i) a structural econometric model, (ii) vector autoregressive methods, and (iii) pure statistical approaches. The satellite models commonly take the form of (panel) regression models. The direct approach, on the other hand, is based on projections of the actual solvency and liquidity parameters without an explicit link to the state of economic and financial variables. While this approach could be equally meaningful in terms of the outcome of stress tests, it does not allow for detailed storytelling and can underestimate the importance of nonlinear macro-financial factors for bank-specific stress tests.

Modeling contagion effects and their impacts is typically challenging. By definition, spillover effects and other dynamic contagion effects are implicitly captured in past data, but not necessarily if one uses structural econometric models, which are usually perceived as being best practice. Even if potential spillover events are captured in past data, this data might not be representative for a future scenario if, for example, linkages between economies and banks have become gradually more intense over time. In this study, we focus on spillover effects originating from the recent sovereign debt crisis. Other spillover catalysts could be, for instance, a macroeconomic downturn in a major world economy or the failure of a large financial institution such as the case of Lehman Brothers.

We aim to develop a stress testing approach that captures spillover effects in detail. Our solution is an amended version of the indirect approach: the starting point is to establish a macroeconomic scenario, typically not informed by potential spillover effects, at least not explicitly. In the second step, the potential marginal increase of stress due to spillover effects is estimated by translating the spillover effects into reduced output paths, i.e., an adverse macroeconomic scenario. In terms of the stylized design of macroeconomic stress tests (figure 1), we thereby implicitly incorporate a quasi-feedback loop into the linear design of traditional stress tests – through a sensitivity type approach. The approach could also include a test for interbank contagion, as shown in figure 1. We build on previous IMF work to establish an explicitly iterative process, i.e., to establish a scenario informed by initial spillover effects on the basis of

---

2 For work on stress testing at the IMF, for example, see Jobst et al. (2013)
3 Frank et al. (2008)
4 Jobst et al. (2013)
5 Most liquidity stress tests have typically relied on the “direct” approach.
6 Foglia (2009)
7 Jobst et al. (2013)
8 Hesse et al. (2014) provide more technical details on the stress testing approach.
9 Further information on macroeconomic scenarios used for FSAPs can be found in Jobst et al. (2013)
How to capture macro-financial spillover effects in stress tests

a structural econometric approach, compute the impact on banks' solvency parameters, recompute the resulting spillover effects and feed them back to the structural model, and so on, until an equilibrium is reached. The approach presented herein uses proxies for the “ultimate” impact of spillovers for different advanced and emerging economies conditional on the evolution of sovereign spreads in the Euro area periphery (that serves as the stress catalyst). Dynamic effects can also be captured via “direct” approaches, but renders the outcome a reduced-form type.

Specifically, we infer from market data the magnitude of sovereign spread spillover effects resulting from an increase in peripheral EU sovereign debt spreads, while controlling for changes in the market sentiment (i.e., risk aversion) and macroeconomic factors. Using market data, we seek to capture point-in-time and dynamic time series effects, while recognizing the limitations of using market data, i.e., that they might not necessarily “only” reflect underlying vulnerabilities and risks. The translation of sovereign spread spillovers into a loss of output is based on recent work at the IMF.

Two approaches are used to capture the spillover effects in sovereign debt markets: panel regressions and a GARCH model. The panel regressions, which are used to establish an “average” impact of spillover effects during periods of stress on advanced market (AM) and emerging market (EM) countries, respectively, suggest that increasing sovereign risk in the Euro periphery was a major driving force behind the spillover effects. As expected, risk aversion, measured through changes in the VIX and high-yield spreads, is found to increase during periods of financial stress, exhibiting a nonlinear pattern. Country-specific macroeconomic factors also matter, but to a lesser degree, and their impact does not appear to change significantly under periods of stress.

GARCH models were run to obtain more granular spillover effects, such as the country-specific co-movements between peripheral European GIIPS sovereign debt spreads and the corresponding spreads in the banks' home countries (i.e., the 25 most systemically important financial systems, the “S-25” sample) for specific points in time. The study reveals significant differences in terms of the spillovers across countries, with a higher impact observed for most core Euro area countries (in particular, during peak periods of the crisis) than for Scandinavian countries, Switzerland, the UK and most non-European countries. The findings also show a flight-to-quality element, i.e., a negative co-movement of GIIPS spreads with German Bunds and US Treasuries.

In the second part of the paper, we illustrate how the established spillover effects would feed through to banks based on a case study for 154 large international banks from the “S-25” country sample. The impact of different degrees of spillover on banks' solvency and liquidity positions is compared with baseline type conditions (which correspond to realized stress scenarios in recent years unlike in “normal” times). Stress at the bank level is simulated on the basis of a recently developed IMF stress testing framework for liquidity and benefits from work on solvency, which together allow running integrated solvency and liquidity scenarios.

The outcome suggests that spillover effects have a highly nonlinear impact on bank soundness, both in terms of liquidity and solvency. It is thereby shown (once more) that the design of stress scenarios is a highly crucial element of stress testing, and is sensitive with respect to the outcome of stress tests. The magnitude of the impact on bank solvency and liquidity could serve as a benchmark for other studies, while recognizing that future spillover channels could be highly different, in terms of both direction and magnitude. In this sense, our study could help to identify potential systemic vulnerabilities ex-ante, a role that stress tests have not necessarily played in the past for a number of reasons.

10 At the IMF, such analyses were carried out by combining the work of Schmieder et al. (2011) and Vitek and Bayoumi (2011) as part of early warning analyses and vulnerability exercises. It should be noted that running such an approach requires close cooperation between staff running macroeconomic forecasts and staff simulating the impact of stress at the bank level (typically done by financial stability departments).
11 As done by Jobst and Gray (2013), for example.
12 IMF (2011, 2012)
13 Vitek and Bayoumi (2011)
14 GIIPS refers to Greece, Ireland, Italy, Portugal and Spain.
15 Schmieder et al. (2012)
16 Schmieder et al. (2011), Hardy and Schmieder (2013)
17 The frameworks were developed in the context of recent FSAPs and IMF technical assistance, extending the seminal work of Čihák (2007), and drawing upon work at the Austrian National Bank (OeNB).
18 Taleb et al. (2012) on how to test the sensitivity (i.e., nonlinearity) of the outcome of stress tests.
19 Boric et al. (2012)
The paper is organized as follows. Section 2 investigates financial spillovers at the sovereign and bank level, based on panel regressions and a GARCH model framework. Section 3 provides a brief overview of the stress testing framework used to simulate the impact of spillover effects on bank liquidity and solvency. Section 4 shows the impact of different degrees of spillover based on a case study. Finally, section 5 concludes and offers some avenues for future research. The appendix shows an illustrative country example.

2. Financial spillovers from the Euro periphery to the rest of the world

2.1 Panel approach

Financial market linkages across economies have grown significantly in recent decades; linkages which were felt strongly when the financial crisis started in 2008 with the failure of Lehman Brothers, and later continued to become a sovereign debt crisis, especially in the European periphery. AM financial spillovers have been a dominant determinant of AM and EM financial soundness during the previous years.

Recent studies identified three important factors for spillover effects: (i) a stress spillover catalyst – in this study, AM sovereign debt yields, (ii) risk aversion in global markets, and (iii) country-specific risk factors.

Herein, we seek to establish benchmark parameters to simulate spillover effects at the bank level. Initially, we construct a risk premium variable for our sample of 35 countries. The risk premium is the spread between 10-year domestic treasuries to US Treasuries for non-European AM countries, to German Bunds for AM countries in Europe and to the JP Morgan Emerging Markets Bond Index (EMBI) for the EM countries.

On the basis of random effects panel regressions, the sovereign spreads are regressed on three sets of peripheral spreads: (i) for Greece, Ireland and Portugal (GIP), and (ii) for Italy and Spain (IT-ES). Risk aversion is identified by two variables: high-yield spreads and the VIX. The former is the difference between yields to maturity of Moody’s Aaa-rated and Baa1-rated US corporate

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIIPS spread</td>
<td>0.237a (0.045)</td>
<td>0.244a (0.047)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIP spread</td>
<td></td>
<td>0.288a (0.046)</td>
<td>0.289a (0.047)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy/Spain spread</td>
<td></td>
<td></td>
<td></td>
<td>0.611a (0.09)</td>
<td>0.653a (0.094)</td>
<td></td>
</tr>
<tr>
<td>High-yield spread</td>
<td>0.666a (0.242)</td>
<td>0.621a (0.229)</td>
<td></td>
<td>0.357 (0.30)</td>
<td></td>
<td>-0.070 (0.291)</td>
</tr>
<tr>
<td>VIX</td>
<td></td>
<td>0.348 (0.238)</td>
<td></td>
<td>0.342 (0.229)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td>0.015 (0.017)</td>
<td>0.015 (0.017)</td>
<td>0.031c (0.016)</td>
<td>0.030c (0.016)</td>
<td>0.025 (0.020)</td>
<td>0.025 (0.021)</td>
</tr>
<tr>
<td>M2/GDP</td>
<td>0.080a (0.017)</td>
<td>0.078a (0.017)</td>
<td>0.061a (0.016)</td>
<td>0.060a (0.016)</td>
<td>0.053a (0.020)</td>
<td>0.051b (0.020)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.297b (0.131)</td>
<td>-0.632 (0.744)</td>
<td>0.256c (0.136)</td>
<td>-0.660 (0.718)</td>
<td>0.700a (0.166)</td>
<td>0.997 (0.912)</td>
</tr>
<tr>
<td>R-squared (within)</td>
<td>0.77</td>
<td>0.70</td>
<td>0.79</td>
<td>0.73</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td>Observations</td>
<td>415</td>
<td>415</td>
<td>435</td>
<td>435</td>
<td>454</td>
<td>454</td>
</tr>
<tr>
<td>T</td>
<td>25</td>
<td>25</td>
<td>23</td>
<td>23</td>
<td>26</td>
<td>26</td>
</tr>
</tbody>
</table>

Table 1: Panel regressions, 2006 Q1-2012 Q2 (dependent variable: sovereign spreads of 35 sample jurisdictions) (quarterly data)

Standard errors in parentheses. a = p<0.01, b = p<0.05, c = p<0.1. *Right-hand-side variables are in logs.

20 Caceres and Unsal (2011)
21 The sample of jurisdictions includes Australia, Austria, Belgium, Brazil, Canada, China (mainland), Cyprus, Denmark, Finland, France, Germany, Greece, Hong Kong SAR, Hungary, India, Ireland, Italy, Japan, Korea, Luxembourg, Malta, Mexico, Netherlands, Norway, Poland, Portugal, Russia, Singapore, Slovenia, Spain, Sweden, Switzerland, Turkey, the UK and the US.
22 The panel regressions adjust for exchange rate changes.
How to capture macro-financial spillover effects in stress tests

The latter is the implied volatility for S&P 500 index options. Trade openness, liquidity (proxied by M2 to GDP and the level of reserves to GDP), inflation rates, GDP growth, the current account, the level of public debt and deficits to GDP ratios are used as macroeconomic control variables to capture country-specific cyclical effects. The regressions are estimated for two time periods on the basis of quarterly data: 2006–12 and 2008–12. The choice of the two sample periods is meant to capture the impact of the systemic stress. The results (displayed in tables 1 and 2) present various model specifications considered useful to identify drivers of spillover stress and their actual impact, respectively. Using the sovereign debt spreads of the 35 sample countries as the dependent variable, table 1 shows the outcome for 2006–12 and 2008–12. The choice of the two sample periods is meant to capture the impact of the systemic stress. The results (displayed in tables 1 and 2) present various model specifications considered useful to identify drivers of spillover stress and their actual impact, respectively. Using the sovereign debt spreads of the 35 sample countries as the dependent variable, table 1 shows the outcome for 2006–12 and 2008–12. The results confirm previous studies’ in that all three factors, i.e., a catalyst, risk aversion and country-specific factors, are actually important to explain financial stress (measured in terms of sovereign spreads), at least for the recent financial crisis:

- Increasing sovereign risk in the Euro periphery was found to be a catalyst for spillover effects
- The global perception of risk magnifies stress conditions, as do expected future interest rates
- Country-specific macroeconomic factors also matter, but to a lesser degree
- While the impact of country-specific factors does not appear to change significantly under stress, the impact of the former two factors is higher during 2008–12, i.e., in the period covering only the crisis years (compared with the full sample period)

For the longer sample period (i.e., 2006–12), a one percentage point change in Euro periphery sovereign spreads (i.e., GIIPS and GIP) translates into a 0.2–0.3 percentage point change of sovereign debt spreads in the 35 sample countries (table 1). Global risk aversion (measured by changes in high-yield spreads) has an even higher impact — a 1 percentage point change in high-yield spreads translates into a circa 0.6 percentage point change in sovereign spreads. As global risk aversion and high-yield spreads are highly correlated during episodes of stress, the joint impact on the peripheral spreads is exacerbated — which is illustrated in a comparison of the coefficients in tables 1 and 2. The transmission of risk premium shocks from Italy and Spain to the countries in the sample is more pronounced than for the GIPs.

Table 2: Panel regressions, 2008 Q1–2012 Q2 (dependent variable: sovereign spreads of 35 sample countries) (quarterly data)

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIIPS spread</td>
<td>0.492a (0.105)</td>
<td>0.463a (0.106)</td>
<td>0.511a (0.090)</td>
<td>0.479a (0.090)</td>
<td>1.002a (0.173)</td>
<td>0.998a (0.175)</td>
</tr>
<tr>
<td>GIP spread</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy/Spain spread</td>
<td>1.042a (0.299)</td>
<td>1.033a (0.279)</td>
<td>0.813a (0.301)</td>
<td>0.735b (0.366)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-yield spread</td>
<td>0.823b (0.322)</td>
<td>0.813a (0.301)</td>
<td>0.034c (0.019)</td>
<td>0.033c (0.019)</td>
<td>0.033 (0.027)</td>
<td>0.032 (0.027)</td>
</tr>
<tr>
<td>VIX</td>
<td>0.018 (0.021)</td>
<td>0.017 (0.021)</td>
<td>0.057a (0.018)</td>
<td>0.056a (0.018)</td>
<td>0.045c (0.025)</td>
<td>0.043c (0.025)</td>
</tr>
<tr>
<td>Openness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2/GDP</td>
<td>0.078a (0.020)</td>
<td>0.075a (0.020)</td>
<td>0.056a (0.018)</td>
<td>0.056a (0.018)</td>
<td>0.045c (0.025)</td>
<td>0.045c (0.025)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.133 (0.222)</td>
<td>-2.418b (1.084)</td>
<td>-0.216 (0.222)</td>
<td>-2.459b (1.022)</td>
<td>0.308 (0.246)</td>
<td>-1.117 (1.307)</td>
</tr>
<tr>
<td>R-squared (within)</td>
<td>0.93</td>
<td>0.78</td>
<td>0.91</td>
<td>0.78</td>
<td>0.91</td>
<td>0.85</td>
</tr>
<tr>
<td>Observations</td>
<td>321</td>
<td>321</td>
<td>357</td>
<td>357</td>
<td>341</td>
<td>341</td>
</tr>
<tr>
<td>T</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>

Standard errors in parentheses. a = p<0.01, b = p<0.05, c = p<0.1. 1 Right-hand-side variables are in logs.

23 Measured in terms of the R-squared and the actual coefficients.
Depending on the model specification, the availability of domestic liquidity and trade openness also contribute to some degree to spillovers.\textsuperscript{24}

The outcome for the crisis period only (covering the years from 2008–12, table 2) indicates that the coefficients for all three major drivers, i.e., European periphery shocks, global risk aversion as well as the slope of the US yield curve are higher than for the period including the pre-crisis years (table 1).

A 1\% shock to Euro periphery spreads translates into a 0.5 percentage point increase in the risk premium of the 35 sample countries if the shock originates in the GIPs and a one percentage point increase in spreads if it originates in Italy and Spain.

Hence, it seems that the size of the peripheral European country determines the size of spillovers, as expected. Moreover, global risk aversion shocks also translate almost one-to-one into spreads.

\textsuperscript{24} For robustness check, a separate set of regressions was run to estimate the impact of expectations of higher interest rates, represented by the slope of the US Treasury yield curve on the global risk premium. Results indicate that a steepening of the curve implies higher costs of borrowing for the periphery countries.
2.2 DCC GARCH approach
The panel regression approach provided the average spillover effect on countries’ sovereign spreads. Below, we complement the above by estimating country-specific daily co-movements in order to differentiate more between countries and produce the range of the potential spillover impacts observed over time. We use a multivariate GARCH framework for the estimation, which allows for heteroskedasticity of the data and a time-varying correlation in the conditional variance. Specifically, the Dynamic Conditional Correlation (DCC) specification by Engle (2002) is adopted, which provides a generalization of the Constant Conditional Correlation (CCC) model by Bollerslev.25, 26 The DCC GARCH models are estimated in first differences to account for the non-stationarity of the variables in the crisis period.

These econometric techniques allow us to analyze the daily co-movement of the GIIPS spreads and the sovereign bond spreads of our sample of AMs and EMs. The GIIPS spreads are included in the model as a conditioning variable, as is the VIX. The methodology is, therefore, closely aligned to that of the panel regression.

We choose as the sample period daily data from 2007 to end of August 2012, with a view to covering the full crisis period.

As before, for the European AMs, we measure the risk premium of 10-year instruments as the difference between the average GIIPS spread as well as those of the domestic treasuries to German Bunds. For the non-European countries, the spread to the 10-year US Treasury bonds is calculated, and for EM countries, we use the EMBI Global spread and the HSBC Asian US Dollar spread for Asian jurisdictions.

As expected, our findings suggest that the spread between GIIPS to German Bunds exhibits a higher degree of co-movement with the risk premiums for European countries than non-European countries (figures 2-5). In particular, implied DCC GARCH correlations with the GIIPS spread were as high as 0.7-0.8 for Austria, Belgium, France and the Netherlands during episodes of systemic stress (figure 2, panels A and B). In contrast, the GIIPS co-movement with the UK spread to German Bunds is relatively low and oscillates between 0 and 0.2, while the model implied correlation with the Swiss spreads reaches a maximum of 0.4 (figure 2, panel C). The results also show that the spreads of the Scandinavian countries, namely Denmark, Norway, Sweden and Finland (with higher average levels though),27 on average exhibit a lower co-movement with the GIIPS spread than their continental European peers (figure 2, panel D). The outcome also suggests a constant level of stress, with some easing toward the end of the observation period, a finding that also applies to the non-European sovereigns.

Co-movements of the GIIPS spread with Australian and Canadian spreads (relative to US Treasury bonds) are rather low, with implied correlations up to 0.2 (figure 3). Looking at the Asian jurisdictions, Hong Kong SAR, Japan and Singapore show a somewhat higher correlation, with the GIIPS spread of up to 0.3 and with one jump to 0.4. In terms of EM countries, results suggest that mainland China’s co-movement with the GIIPS spread is relatively subdued compared with the other EMs: Brazil, Mexico, Russia and Turkey (figure 4). Out of this EM sample, Turkey has the highest implied correlation with the GIIPS during episodes of system stress, up to 0.6.

Since the onset of sovereign debt crisis in 2009, the average GIIPS interest rates exhibit a negative correlation with both the German Bund and US Treasury interest rates (figure 5).
Since 2009, the implied correlation has turned negative for both countries, with lows of -0.4 (US) and -0.6 (Germany), indicating a sudden flight to safety, in line with other recent studies.\(^\text{28}\)

3. Liquidity and solvency stress testing

The area of stress testing has seen a number of advances during recent years. Our study uses a recently developed IMF liquidity stress testing framework to run integrated solvency and liquidity stress tests. The liquidity stress testing framework presented in Schmieder et al.\(^\text{29}\) was developed in the context of recent financial sector assessment programs (FSAPs)\(^\text{30}\) and International Monetary Fund (IMF) technical assistance, extending the seminal work of Čihák,\(^\text{31}\) and drawing upon work at the Austrian National Bank (OeNB).\(^\text{32}\)

In this study, the focus is on scenario design, namely building integrated scenarios for solvency and liquidity risks that take into account spillover effects and feedback loops.\(^\text{33}\) The central question becomes how the findings established in section 2 can be used to inform bank-level stress tests.

Nevertheless, while we attempt to condense a wealth of information and assumptions to establish integrated scenarios, this should not, in any sense, give a false sense of precision. Instead, we recommend running a whole range of scenarios that can build upon the ones established in the study, with varying

---

\(^{28}\) IMF (2011)

\(^{29}\) Schmieder et al. (2012).

\(^{30}\) Examples include Chile, Germany, India, Spain, Turkey and the UK.

\(^{31}\) Čihák (2007).

\(^{32}\) It is complemented by a previously developed solvency stress testing tool by Schmieder et al. (2011). While developing the solvency and liquidity stress testing frameworks, four key facts were accounted for, which constitute key challenges of contemporaneous financial stability analysis: the availability of data varies widely, and lack of data is common; both solvency and liquidity risks have various dimensions, which require multidimensional analysis, thereby integrating risks; designing and calibrating scenarios is challenging, even more so for liquidity risk than for solvency risk (mainly as liquidity crises are relatively rare and originate from different sources); and communication of stress test results is an integral part of the exercise. The answers to these multiple dimensions are Excel-based balance sheet type frameworks.

\(^{33}\) The exercise thereby reflects key principles for liquidity stress testing put forward by the Basel Committee in the aftermath of the first wave of shocks following the default of Lehman Brothers (BCBS, 2008).
degrees of severity. Reverse stress tests can also be included.\textsuperscript{34} This is an important way forward to obtain a better understanding of key solvency and liquidity risks faced by banks, and to gain a more comprehensive view of their respective risk tolerance.

3.1 Liquidity stress testing approach
We apply an implied cash-flow approach to simulate the impact of a bank-run type stress scenario. The banks’ liabilities are broken down into demand and term deposits, short-term wholesale funding (including bank and secured funding), derivatives funding, as well as long-term funding, such as senior debt or subordinated debt. On the asset side, we include a range of potentially liquid asset positions, such as cash, government, trading and investment (both available-for-sale and held-to-maturity) securities, loans and advances to banks and reverse repos and cash collateral. Given European periphery banks’ increasing collateral use of pools of loans (such as covered bonds) for liquidity, we also include a crude definition of banks’ loan level as a portion of their total assets.

3.2 Solvency stress testing approach
We use rules of thumb for solvency stress testing as proposed by Hardy and Schmieder\textsuperscript{35} and thereby, a simplified solvency test.\textsuperscript{36} Credit losses, banks’ pre-impairment income and the trajectories of risk-weighted assets (RWAs) for a two-year horizon were simulated based on GDP trajectories, with and without spillover effects. The capital shortfall was measured against a tier 1 capital ratio (tier 1 capital/risk-weighted assets) of 6%, below which a bank is considered undercapitalized.\textsuperscript{37}

4. Integration of the financial spillover analysis with the stress testing approach
Our integrated approach to simulate stress at the bank level is illustratively shown in figure 6:

1. Scenario design: We use the GDP trajectories of a specific macroeconomic scenario, the WEO baseline scenario for 2013-14 as of April 2012 and add the spillover stress component.
2. Spillover analysis: The outcome of the spillover analysis (see above), measured through a widening of sovereign spreads, worsens the macroeconomic scenario, and is used as a sensitivity analysis. The translation of the spillover effects into the revised macroeconomic trajectories is based on recent IMF work.
3. Soundness of banks: The scenario is translated into bank-level stress parameters to simulate both the banks’ solvency and liquidity positions, drawing on work by Hardy and Schmieder and Schmieder et al. respectively.\textsuperscript{38}

We use bank-level data from Bankscope (from end of June 2012) for large systematically important banks (SIBs). In total, our sample includes 154 large banks from the following 26 jurisdictions: Austria, Australia, Belgium, Brazil, Canada, Switzerland, mainland China, Germany, Denmark, Finland, France, the UK, Hong Kong SAR, India, Japan, Korea, Luxembourg, Mexico, the Netherlands, Norway, Poland, Russia, Sweden, Singapore, Turkey and the US.

Our sample comprises almost the full EBA sample for the European banks (except for the banks in the GIIPS countries) and includes the largest banks in the non-European countries. In total, it captures US$84t of bank assets (i.e., about 50% of the assets held by banks worldwide), US$39t non-bank deposits and around US$7t of government securities held by banks.

\textsuperscript{34} The work by Taleb et al. (2012) and Schmieder and Hardy (2013) for example, could be useful to consider in this context.
\textsuperscript{35} Hardy and Schmieder (2013)
\textsuperscript{36} However, it should be noted that the evidence is based on a comprehensive set of data from 16,000 banks during the past 15 years (as available).
\textsuperscript{37} Please note that this specific choice is meant for illustration only - through a similar level as used for the European stress tests conducted in 2010 and 2011, for example.
\textsuperscript{38} Hardy and Schmieder (2013), Schmieder et al. (2012)
4.1 Scenarios
We refer to four different scenarios: the April 2012 World Economic Outlook (WEO) baseline scenario for 2013–14 (scenario 1); and three spillover scenarios (referred to as scenarios 2.x) conditional on scenario 1 – scenarios that banks could potentially face, where increasing degrees of spillovers affect the general growth trend.

Specifically, scenario 1 is adjusted for an increase of GIIPS spreads by 100 (scenario 2a), 200 (2b) and 300 (2c) basis points, respectively. We further distinguish between the spillover impact observed during periods of substantial financial stress (using the panel regression for 2008–12 and the GARCH model for 2010–12) and during periods of less significant stress (using the panel regression for 2006–12 and the GARCH model for 2008–12), i.e., refer to a total of six spillover scenarios (2a/1, 2a/2, 2b/1, 2b/2, 2c/1, 2c/2).

For the banks’ solvency, we simulate their tier 1 capital ratios by end of 2014, on the basis of the evolution of the main solvency dimensions (banks’ income and losses). For liquidity, we determine the impact of a worst-case idiosyncratic shock to the bank’s liquidity profile on top of the impact on liquidity resulting from the macroeconomic/spillover scenarios.

4.2 Impact on bank solvency
As outlined above, we use the outcome of the 2012 Spillover Report by IMF, which simulates the impact of a 300 basis point increase in peripheral countries’ spreads (including a lower-yield increase for core countries) on European countries’ GDP paths on the basis of the IMF G-35 model.39

We applied our scenarios to a stylized Austrian bank.40 In the first step, we simulated an increase in the Austrian sovereign debt spread using the evidence established in section 2. A 100 basis point shock of GIIPS spreads (scenario 2a) resulted in an increase of Austrian spreads by 24 basis points for less significant spillover stress (scenario 2a/1) and 50 basis points (2a/2) for more substantial spillover stress. When measured in relation to the April 2012 WEO baseline scenario for Austria, suggesting real GDP growth rates of 1.8% (2013) and 2.2% (2014), spillover analysis carried out at the IMF (2012) would predict a drop of real GDP growth by about 0.45 percentage points for scenario 2a/1 (less significant spillover stress), whereby the GDP trajectory becomes 1.4% (2013) and 1.8% (2014). For a period with more significant spillover (scenario 2a/2), the impact is about twice (0.9 percentage points), whereby the GDP trajectory is 0.9% (2013) and 1.3% (2014). For a 200 basis point shock (scenario 2b), growth drops by 1.7 percentage points and for 300 basis

39 Vitek and Bayoumi (2011)
40 Data for this scenario analysis can be obtained from the authors.
How to capture macro-financial spillover effects in stress tests

The Journal of Financial Perspectives

points (scenario 2c) by 2.6 percentage points (per year) under substantial spillover conditions.

We then use the satellite models by Hardy and Schmieder to determine banks’ loan impairment levels and pre-impairment income for 2013 and 2014. For a stylized bank with loss impairment rates of 0.5% and a pre-impairment return on capital of 10% in 2012, loan impairment rates are simulated to decrease slightly under the baseline scenario and mild spillover conditions, while they would increase (nonlinearly) under increasing levels of spillover stress. The same pattern holds for pre-impairment income. This input is used to simulate the bank’s capital, RWAs and capital ratio. Again, the same pattern holds, with a decrease of the stylized banks’ capital ratio to 7.5% under the most severe scenario, which is above the hurdle rate in terms of tier 1 capital to pass the stress test (6%).

The outcome of this solvency stress test applied to the 154 banks presented in figure 6 shows that the large international banks would be in a position to digest the baseline scenario plus some level of spillover stress, while additional stress in the Euro area periphery results would have a highly nonlinear impact on potential capital needs. The nonlinearity results from two factors: the nonlinearity in the satellite models for loan impairment rates and pre-impairment income and the effect of the kick-in of capital needs for banks that fall below the hurdle rate.

4.3 Impact on bank liquidity
For the liquidity stress test, we simulate the impact of stress on both banks’ market liquidity (i.e., their ability to fire sale assets) and funding liquidity (i.e., the potential outflow of funding). Again, we assume that the bank is affected by the shock in its home country. The link between the level of stress and bank liquidity is established on the basis of empirical work of Schmieder et al. We link the GDP trajectories, implied by the changes of sovereign spreads, to funding shocks experienced by the most affected banks during the Lehman crisis. In other words, we simulate highly adverse idiosyncratic liquidity shocks conditional upon macroeconomic conditions.

41 Hardy and Schmieder (2013)
42 For simplification, we assume that banks are affected according to their domestic scenarios, i.e., that their business is predominantly based in their home country.
43 In a few cases, the latest available figures were from 2011.
44 The RWAs are simulated based on work by Schmieder et al. (2011), assuming point-in-time credit risk parameters.

45 Unlike for the solvency scenario, we do not simulate stress for a specific point in time; rather, the simulated stress conditions reflect a worst-case situation resulting from the general macroeconomic conditions as well as an idiosyncratic shock to the bank conditional.
46 In other words, it is assumed that all of its assets are based in the home country, which is a crude simplification.
47 Schmieder et al. (2012)
In line with (very limited) empirical evidence, we expect the relationship between the shock and the potential adverse impact on the bank level to be highly nonlinear (table 3). Under a worst-case scenario, banks would experience a shock equal to a Lehman Brothers type scenario, the severe stress scenario (this shock level represents how the stress at the time of the Lehman Brothers’ event affected the banks that were most severely hit, i.e., overlays a market shock with an idiosyncratic liquidity shock). The stress level relative to the one experienced by banks at the time of the Lehman Brothers crisis is established via the cumulative GDP trajectory under stress compared with the long-term average. For the example of the Austrian bank discussed above, the stress level is at 0.65, i.e., the benchmark funding stress parameters (for the severe stress scenario) in table 3 have to be multiplied by 0.65.

### Table 3: Benchmark stress scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Moderate stress scenario</th>
<th>Medium stress scenario</th>
<th>Severe stress scenario</th>
<th>Very severe stress scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liquidity outflows</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer deposits (term)</td>
<td>2.5%</td>
<td>5%</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td>Customer deposits (demand)</td>
<td>5%</td>
<td>10%</td>
<td>20%</td>
<td>40%</td>
</tr>
<tr>
<td>Wholesale funding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short-term (secured)</td>
<td>5%</td>
<td>10%</td>
<td>20%</td>
<td>40%</td>
</tr>
<tr>
<td>Short-term (unsecured)</td>
<td>25%</td>
<td>50%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Contingent liabilities</td>
<td>0% need funding</td>
<td>5% need funding</td>
<td>10% need funding</td>
<td>20% need funding</td>
</tr>
<tr>
<td>Liquid inflows</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haircut for cash</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Haircut for Government securities</td>
<td>1%</td>
<td>2%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haircut for trading assets/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proxies, specific assets</td>
<td>3%</td>
<td>6%</td>
<td>30%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Equities: 3 Bonds: 3</td>
<td>Equities: 4-6 Bonds: 3-8</td>
<td>Equities: 10-15 Bonds (only LCR eligible ones): 5-10</td>
<td>Not liquid</td>
</tr>
<tr>
<td>Haircut for other securities</td>
<td>10%</td>
<td>30%</td>
<td>75%</td>
<td>100%</td>
</tr>
<tr>
<td>Percent of liquid assets encumbered</td>
<td>10% (or actual figure)</td>
<td>20% (or actual figure plus 10 ppt)</td>
<td>30% (or actual figures plus 20 ppt)</td>
<td>40% (or actual figures plus 30 ppt)</td>
</tr>
</tbody>
</table>

---

1/ The Lehman type scenario would correspond to a scenario encountered by banks that were hit severely during the 30-day period after the Lehman collapse, i.e., a stress situation within a stress period rather than an average. The scenario has been put together based on expert judgment, using evidence as available.

2/ The haircut highly depends on the specific features of the government debt held (rating, maturity, market depth) and can be higher or lower. The figures displayed herein are meant for high-quality investment grade bonds, taking into account recent market conditions. The same applies for the remainder of the liquid assets. For the securities in the trading book, it is assumed that they are liquidated earlier, resulting in lower haircuts.

3/ A haircut of 100% means that the asset is illiquid, i.e., the marker has closed.

4/ The figures account for a downgrade of the bank, which triggers margin calls and higher collateral requirements generally. Please note that the unencumbered portion applies to a gradually narrower definition of liquid assets.
How to capture macro-financial spillover effects in stress tests

The funding available for the specific banks under the European Central Bank's Long Term Refinancing Operations is inferred from country-level data and used as a cushion for the relevant European banks.

Figure 8 shows the outcome of this liquidity stress test. Under the baseline scenario, all banks have sufficient liquidity, as expected. Adding spillover stress triggers a nonlinear increase in liquidity needs (which occur in case the liquidity needs exceed the available liquidity generated via fire sales), and more substantial spillover stress makes the stress highly nonlinear. Measured against tier 1 capital rather than total assets, the substantial spillover stress leads to a maximum liquidity shortfall of 20% for the entire bank sample for scenario 2c/2 (300 basis points spread shock, significant spillover stress) and close to 6% for scenario 2b/2 (200 basis points spread shock), compared with 0.3% and 1% if measured against total assets.48

5. Conclusion

This study attempted to contribute to an important challenge faced by current financial stability analysis, namely to capture spillover effects and other types of contagion that ultimately determine macro-financial stress at the bank level.

By integrating recent IMF work on financial spillover analysis and stress testing, we use a novel framework that allows some light to be shed on the potential impact of spillover effects on bank-level solvency and liquidity. Nevertheless, we recognize that significant additional effort and evidence is needed to make the modeling of dynamic macro-financial linkages more robust, not least owing to the many potential channels of spillover and contagion, the fact that the use of crude data available for stress tests is subject to uncertainty, and other factors that contribute to uncertainty (such as mixed evidence for the use of market data).

The outcome of the stress tests suggests that spillover effects observed for the sovereign debt markets in recent years have a highly nonlinear impact on bank soundness, in terms of both liquidity and solvency. This implies (once more) that the design of stress scenarios is a crucial element of stress testing, and is very sensitive with respect to the outcome of stress tests.

48 We did not explicitly model a central bank response as the lender of last resort (LOLR) to mitigate the estimated liquidity shortfall. In reality and as seen during the crisis period, central banks would provide large liquidity support to solvent banks subject to an appropriate haircut.

The approach used in this paper is meant to be a menu for future analyses of the impact of potential spillovers. Sensitivity analysis and reverse stress tests appear to be an important complement in this context.
References


Appendix A

<table>
<thead>
<tr>
<th>Factor</th>
<th>Ambiguity</th>
<th>Certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sovereign risk</td>
<td>GIIPS spread</td>
<td>Average of Euro periphery sovereign spreads to German Bunds</td>
</tr>
<tr>
<td></td>
<td>GPI spread</td>
<td>Average of Greece, Ireland and Portugal sovereign spreads to German Bunds</td>
</tr>
<tr>
<td></td>
<td>Italy/Spain spread (IS spread)</td>
<td>Average of Italy and Spain sovereign spreads to German Bunds</td>
</tr>
<tr>
<td>Risk aversion</td>
<td>High-yield spread</td>
<td>Difference between yields to maturity of AAA-rated and BAA-rated corporate US bonds</td>
</tr>
<tr>
<td></td>
<td>VIX</td>
<td>Implied volatility of S&amp;P 500 index options</td>
</tr>
<tr>
<td>Macroeconomic environment</td>
<td>Openness</td>
<td>Sum of imports and exports to GDP ratio</td>
</tr>
<tr>
<td></td>
<td>M2/GDP</td>
<td>Broad money to GDP ratio</td>
</tr>
</tbody>
</table>

Main explanatory variables

Source: Authors
Why small portfolios are preferable and how to choose them

Francesco Cesarone  
Department of Business Studies, Roma Tre University

Jacopo Moretti  
Department of Methods and Models for Economics, Territory, and Finance, Sapienza University of Rome

Fabio Tardella  
Head of the Master's course Finance and Insurance, Department of Methods and Models for Economics, Territory, and Finance, Sapienza University of Rome
Why small portfolios are preferable and how to choose them

Abstract

One of the fundamental principles in portfolio selection models is minimization of risk through diversification of the investment. However, this principle does not necessarily translate into a request for investing in all the assets of the investment universe. Indeed, following a line of research started by Evans and Archer almost 50 years ago, we provide here further evidence that small portfolios are sufficient to achieve almost optimal in-sample risk reduction with respect to variance and to some other popular risk measures, and very good out-of-sample performances. While leading to similar results, our approach is significantly different from the classical one pioneered by Evans and Archer. Indeed, we describe models for choosing the portfolio of a prescribed size with the smallest possible risk, as opposed to the random portfolio choice investigated in most of the previous works. We find that the smallest risk portfolios generally require no more than 15 assets. Furthermore, it is almost always possible to find portfolios that are just 1% more risky than the smallest risk portfolios and contain no more than 10 assets. The preference for small optimal portfolios is also justified by recent theoretical results on the estimation errors for the parameters required by portfolio selection models. Our empirical analysis is based on some new and on some publicly available benchmark data sets often used in the literature.
1. Introduction
Since the start of Modern Portfolio Theory in the 1950s, with the seminal mean-variance (MV) model of Markowitz,¹ the main aim of portfolio selection models was that of reducing the risk of an investment in the stock market through diversification while trying to achieve a satisfactory return.

However, an excessive diversification that requires investment in a large number of assets may be undesirable for several reasons. Indeed, starting from Markowitz and from the work of Evans and Archer,² it has been observed by several scholars that the decrease in in-sample volatility that one can obtain by increasing the number of securities in the portfolio rapidly declines with the size of the portfolio. In his fundamental book, Markowitz observed that: “To understand the general properties of large portfolios, we must consider the averaging together of large numbers of highly correlated outcomes. We find that diversification is much less powerful in this case. Only a limited reduction in variability can be achieved by increasing the number of securities in a portfolio.”³

This intuition was confirmed in the very influential work where, for any given size $K$ from 1 to 40, the authors randomly picked subsets of $K$ assets from a market of 470 securities and computed some statistics on the standard deviations of the equally-weighted (EW) portfolios formed with each subset of assets.⁴ They found that the average standard deviation for each size $K$ was decreasing and rapidly converging to an asymptote and they concluded that no more than around 10 assets were needed to almost completely eliminate the unsystematic variation of a portfolio return. Several authors have contributed to the debate about the right size of a portfolio that almost completely eliminates the diversifiable risk in a market.⁵ Furthermore, based on Evans and Archer’s work and on other similar findings, such magic size, or size range, has been recommended in a number of textbooks on investment management and corporate finance.⁶

On a practical side, it is clear that acquiring and monitoring a portfolio that includes many assets is not suitable for small investors. Furthermore, even big investors should consider the opportunity cost of holding large portfolios and should identify the threshold where the costs exceed the benefit of risk reduction.⁷ In addition, from a theoretical viewpoint, it has been shown that the estimation errors for the parameters needed to implement Markowitz’s and other portfolio selection models may rapidly increase with the size of the portfolio when the historical data is limited.⁸ Such errors in the parameters are typically amplified in the optimization process required by the models, thus leading to unreliable solutions.⁹

All these observations, which will be better illustrated in the following, provide a motivation for selecting a relatively small portfolio among those optimizing an appropriate criterion. We thus provide models for selecting such small optimal portfolios and perform an empirical analysis to determine the best-size ranges for the portfolios depending on the size of the market. We also provide further evidence of the benefits of small portfolios, both in terms of in-sample risk reduction and in terms of out-of-sample performance. We point out that our approach is significantly different from the mainstream approach pioneered by Evans and Archer. Indeed, we overcome one of the main weaknesses of their approach, which consists of stating results that are valid only on average. In other words, if one picks an arbitrary EW portfolio of a given size in a market, there is no guarantee that its risk will not be much larger than the average risk of all portfolios of the same size in that market.

The conceptually simple solution that we propose here is just to choose the best EW portfolio for each given size with respect to variance, and, furthermore, the optimal portfolios for each given size with respect to three different and complementary risk measures. In this way, for each size we clearly obtain a portfolio which has a risk not greater (and typically, quite smaller) than the average risk. The reason why this simple idea was not investigated before is probably due to the computational hardness of the models required to find such best portfolios. Indeed, some of these models have been solved exactly only recently for small to medium-size markets,¹⁰ and one model is solved here for the first time. Once we have obtained the optimal size of the minimum risk portfolio, we proceed with a sensitivity analysis that allows us to find the smallest size of a portfolio whose risk is not more than 1% larger than that of the

¹ Markowitz (1952, 1959)
² Evans and Archer (1968)
³ Markowitz (1959, 102)
⁴ Evans and Archer (1968)
⁵ E.g., Newbold and Poon (1993) and references therein.
⁷ Statman (1987)
⁸ Kondor et al. (2007)
⁹ Michaud and Michaud (1998)
¹⁰ E.g., Angeletti et al. (2008); Cesarone et al. (2015) and references therein.
minimum risk portfolio, thus finding even smaller portfolios with satisfactory risk level.

Another difference between our approach and that of Evans and Archer consists of the possibility of using general weights in the selected portfolio instead of equal weights only. For each portfolio size, this clearly allows one to find portfolios with even lower in-sample risk. However, since optimizing weights may also cause the maximization of estimation errors, this choice does not necessarily imply better out-of-sample performance. For both weighting schemes and for all risk measures, we find results comparable to those of Evans and Archer. More precisely, we identify some ranges of (typically small) sizes where the portfolio risks are minimized, and ranges of even smaller sizes where the portfolio risks do not exceed the minimum by more than 1%. The out-of-sample performance of the selected portfolios for each specified size is another important feature of our analysis which is rarely found in previous works on the subject. In our analysis, we find that the best performances are generally obtained by portfolios with no more than 15 assets.

As an interesting complement to our findings, we mention that, in a recent and detailed study on the empirical behavior of investors and on the performance of their portfolios, Ivković et al. show that portfolios of small investors with low diversification exhibit superior performance with respect to those with high diversification.12

### 2. Fixed-size portfolio models

In this section, we present a fixed-size version of several well-known portfolio selection models and provide an integer or a mixed-integer linear or quadratic formulation for all models. We first need to introduce some notation. Let \( T + 1 \) be the length of the in-sample period used to estimate the inputs for the models.

We use \( p_i(t) \) to denote the price of the \( i \)th asset at time \( t \), with \( t = 0, ..., T \); \( \ell_i(t) \) is the \( i \)th asset return at time \( t \), with \( t = 1, ..., T \); \( x \) is the vector whose components \( x_i \) are the fractions of a given capital invested in asset \( i \) in the portfolio we are selecting; \( y \) is a Boolean vector whose components \( y_i \) are equal to 1 if asset \( i \) is selected, and 0 otherwise. We assume that \( n \) assets are available in a market and, adopting linear returns, we have that \( \sum_{i=1}^{n} y_i x_i \) is the portfolio return at time \( t \), with \( t = 1, ..., T \).

The \( n \)-dimensional vector \( \mu \) is used to denote the expected returns of the \( n \) risky asset, while \( \Sigma \) denotes their covariance matrix, and \( u \) denotes an \( n \)-dimensional vector of ones.

#### 2.1 The equally weighted portfolio

The most intuitive way to diversify a portfolio is to equally distribute the capital equally among all stocks available in the market. In terms of relative weights, we have \( x_i = 1/n \). This is known as the EW (also called naïve or uniform) portfolio. Clearly, the choice of the EW portfolio does not use any in-sample information or involve any optimization approach. However, some authors claim that its practical out-of-sample performance is hard to beat on real-world data sets.13 Furthermore, from the theoretical viewpoint, Pflug et al. shows that when increasing the amount of portfolio model uncertainty, i.e., the degree of ambiguity on the distribution of the assets returns, the optimal portfolio converges to the EW portfolio.14 We will, thus use this portfolio as a benchmark to compare the performances of the portfolios obtained by the models.

#### 2.2 Fixed-size minimum variance equally weighted portfolios

As already observed, the EW portfolio is the most robust choice when there is a great uncertainty about the distribution of the asset returns. However, the EW portfolio has the drawback of using all available assets, which may be too numerous and not all desirable. A first proposal to overcome this drawback comes from Jacob, who proposes to select a small EW portfolio (with a specified number \( K \) of assets) that has minimum variance among all EW portfolios of the same size.15 The model by Jacob is a nonlinear 0-1 optimization model that has not yet been tested in practice, owing to its computational complexity. Thanks to the recent advances in solution methods and computing power, here, we can propose an empirical study of such a fixed-size minimum variance equally weighted (FSMVEW) model, formally described below.

\[
\begin{align*}
\min & \quad y^\top \Sigma y \\
\text{s.t.} & \quad u^\top y = K \\
& \quad y \in (0,1)^n
\end{align*}
\]

This is probably the simplest fixed-size portfolio model and has the advantage of not requiring the problematic estimates of the assets’ expected returns. Furthermore, the effects of the possible

\[12\] DeMiguel et al. (2009b); Michaud (1989)

\[13\] Pflug et al. (2012)

\[14\] DeMiguel et al. (2009b)

\[15\] Jacob (1994)
estimation errors of the covariance matrix \( \Sigma \) do not result in very large or small weights for some assets, but only influence the choice of the subset of selected assets in the portfolio. From the optimization viewpoint, it falls into the class of pseudo-Boolean quadratic programming problems, which are known to be theoretically hard to solve in the worst case (NP-hard). However, due to its special structure, practical problems of this type with several hundred variables can be solved fairly efficiently with available free or commercial codes. Note that the vector \( x \) of weights of the optimal FSMVEW portfolio selected by model (1) is obtained as \( x = \frac{1}{K} y \). When \( K = n \), the FSMVEW portfolio coincides with the EW portfolio.

2.3 Fixed-size minimum variance portfolios

Another model that does not require the estimates of the assets’ expected returns is the extreme case of the Markowitz model, where we seek only to minimize variance. Within our framework, we thus consider the following fixed-size minimum variance (FSMV) model, where only \( K \) assets are allowed in the selected portfolio:

\[
\begin{align*}
\min & \quad x^T \Sigma x \\
\text{s.t.} & \quad u' x = 1 \\
& \quad u' y = K \\
& \quad y \leq s x y, \\
& \quad y \in \{0, 1\}^n
\end{align*}
\]  

(2)

The first constraint above is the budget constraint; the second one represents the portfolio fixed-size constraint; \( u \) is an \( n \)-dimensional vector of ones; \( y \) is an \( n \)-dimensional vector of binary variables used to select the assets to be included in the portfolio; \( x \) is the vector of portfolio weights, and \( \ell \) is a minimum threshold (often called the buy-in threshold) for the weights of the selected assets which must be greater than zero (in our experiments we chose \( \ell = 0.01 \)). Without these thresholds, problem (2) could generate portfolios with less than \( K \) assets, which is equivalent to replacing the constraint \( y \leq s x y \). Note that problem (2) is a quadratic mixed integer programming (QMP) problem that falls again in the class of NP-hard problems. However, also in this case, problems with a few hundred variables can be solved fairly efficiently with available free or commercial codes. Furthermore, a recently proposed specialized algorithm can solve problems of this type with up to 2,000 variables.

2.4 Fixed-size minimum conditional value-at-risk portfolios

The fixed-size minimum conditional value-at-risk (FSMCVaR) model is a minimum risk model like the previous one, but instead of variance it measures risk with conditional value-at-risk at a specified confidence level \( \epsilon \) (CVaR(\( \epsilon \))), namely the average of losses in the worst \( 100\epsilon \% \) of the cases. In our analysis, losses are defined as negative outcomes, and we set \( \epsilon \) equal to 0.05. The FSMCVaR model can be written as follows:

\[
\begin{align*}
\min & \quad \text{CVaR}_\epsilon(x) \\
\text{s.t.} & \quad u' x = 1 \\
& \quad u' y = K \\
& \quad y \leq s x y, \\
& \quad y \in \{0, 1\}^n
\end{align*}
\]  

(3)

where \( \ell \) plays the same role as in (2).

Using a classical approach introduced by Rockafellar and Uryasev,\(^{19}\) problem (3) can be reformulated as a mixed integer linear programming (MILP) problem with \( n + T + 1 \) continuous variables, \( n \) binary variables and \( T + n + 3 \) constraints. Some recent computational experiences reporting on the solution of this model with state-of-the-art commercial solvers show that models with more than a few hundred variables are hard to solve with general-purpose solvers and would probably benefit from more specialized methods.\(^{20}\)

2.5 Fixed-size minimum mean semi-absolute deviation portfolios

The last risk measure that we take into account in our analysis is the downside mean semi-absolute deviation (semi-MAD or SMAD):

\[
\text{SMAD}(x) = E\left[\min(0, \sum_{i=1}^{n} (r_i - \mu) x_i)\right]
\]

(4)

This is a concise version of the more famous mean absolute deviation (MAD) risk measure, which is defined as the expected value of the absolute deviation of the portfolio return from its mean.\(^{21}\) Indeed, Speranza showed that semi-MAD leads to a portfolio selection model that is equivalent to the MAD model, but with half the number of constraints.\(^{22}\)

---

16 Boros and Hammer (2002)
17 Cesaroni et al. (2009, 2013)
18 Acerbi and Tasche (2002)
19 Rockafellar and Uryasev (2000); Cesaroni et al. (2014)
20 Cesaroni et al. (2015)
21 Konno and Yamazaki (1991)
22 Speranza (1993)
We thus consider the following fixed-size minimum semi-MAD (FSMSMAD) model:

\[
\begin{align*}
\min & \quad \text{SMA}(x) \\
\text{s.t.} & \quad u^T x = 1 \\
& \quad u^T y = K \\
& \quad \ell y \preceq x \preceq y, \\
& \quad y \in (0,1)^n
\end{align*}
\]

where \( \ell \) plays the same role as in (2). Using the linearization approach described, we can reformulate this problem as a MILP problem with \( n + T \) continuous variables, \( n \) binary variables and \( n + T + 3 \) constraints.

From the computational experiences reported, it appears that this model, although slightly easier than the previous one, also cannot easily be solved with general purpose state-of-the-art solvers when more than a few hundred variables are involved.

3. Large estimation errors for large portfolios

In the real-life asset management, an important issue concerns the problem of the estimation errors of the inputs required by the portfolio selection models, and the subsequent amplification of these errors in the phase of risk-gain optimization.

To analyze the impact of the estimation and model errors on optimized portfolio weights, we recall here the methodology. The authors study the sensitivity to estimation errors of several minimum risk portfolios by considering an artificial investment universe, represented by multivariate standard normal returns with different numbers \( n \) of assets and lengths \( T \) of the time series. They measure the effect of the estimation errors on portfolio selection with the following ratio:

\[
q_0 = \frac{x_{\text{opt}}^T \Sigma x_{\text{opt}}}{x_{\text{opt}}^T \Sigma x_{\text{opt}}} = \frac{1}{1 - \frac{\text{rank}(\Sigma)}{n}}
\]

where \( \Sigma \) is the identity matrix, \( x_{\text{opt}} \) is the minimum variance portfolio w.r.t. \( \Sigma \), \( x_{\text{opt}} \) is the minimum variance portfolio w.r.t. \( \Sigma \). Clearly, by construction, \( x_{\text{opt}} \) coincides with the EW portfolio both when short sellings are allowed and when they are restricted. Note that by optimality of \( x_{\text{opt}} \) w.r.t. \( \Sigma \), the ratio \( q_0 \) will never be smaller than 1. An interesting asymptotic result for the ratio \( q_0 \) is provided by Kondor et al., who show that when \( n, T \to \infty \) with \( n/T \) constant, the asymptotic average estimation error is:

\[
q_0 = \frac{1}{\sqrt{1 - \frac{\text{rank}(\Sigma)}{n}}}
\]

To test the impact on the estimation error of the constraints that limit the number of selected assets (the fixed-size constraint) in the optimal portfolios, we replicate the experiment of Kondor et al. considering four artificial markets with \( n = 100, 200, 400, 470 \) and four lengths of the time series \( T = 500, 1000, 1500, 2000 \). We analyze the sensitivity to estimation errors of the following three portfolio models:

- Minimum variance with short selling (MV-ss)
- Minimum variance without short selling (MV-Noss)
- Fixed-size minimum variance (FSMV)

For the FSMV model (2) and for each artificial market with \( n \) assets, we consider 10 sizes of the selected portfolio \( K = n \cdot (10\%, 20\%, \ldots, 100\%) \).

In Figure 1, we report the mean of \( q_0 \) on 20 simulations for different values of \( n \) and \( T \), as described above. The black solid line is the theoretical shape of \( q_0 \) as in (7), which is well fitted by the simulated black squares coming from the MV-ss model. The blue circles and the red crosses are the average values of \( q_0 \) obtained from the MV-Noss and FSMV models, respectively. It is clear that the addition of no short selling or size constraints to the portfolio selection models tends to reduce the influence of estimation errors, particularly when the portfolio becomes large, i.e., when \( n/T \) approaches the critical value 1, where the empirical estimate of the covariance matrix becomes singular. Furthermore, the size constraints seem to provide a slightly better reduction of the effects of the estimation errors on the models with respect to the no short selling constraints.

4. Empirical behavior of the models

In this section, we test the models described above on some publicly available data sets. The analysis consists of two parts. First, we examine the behavior of the portfolios selected by the

23 Speranza (1993)
24 Cesaroni et al. (2014)
25 Cesaroni et al. (2015)
26 Kondor et al. (2007)
models on the in-sample window, where we obtain the input parameters of the models. The second part consists of evaluating the out-of-sample performance of the portfolios, which is the aspect that matters most to investors. Since the markets are in continuous evolution, it seems appropriate to rebalance the portfolio from time to time in order to take new information into account. For this purpose, we use a rolling time window procedure (RTW), i.e., we shift the in-sample window (and consequently the out-of-sample window) all over the time length of each data set. More specifically, we consider a time window (in-sample period) of 200 observations for the data sets with weekly frequency, and of 120 observations for the data sets with monthly frequency. The choice of the lengths of the in-sample and the out-of-sample windows is based on typical settings of portfolio selection problems.\textsuperscript{28} Then we solve the selection problem for overlapping windows built by moving forward in time with step size 4 (for the weekly data sets) or 1 (for the monthly data sets). The optimal portfolio found with respect to an in-sample period is held for the following four weeks (out-of-sample period of the weekly data sets) or one month (out-of-sample period of the monthly data sets). The out-of-sample performances of the resulting portfolios are evaluated in different ways by computing some performance measures commonly used in the literature.\textsuperscript{29}

\begin{align*}
\hat{x}^* = (x^*_1, \ldots, x^*_n) & \text{ denote the allocation of the selected portfolio and } r_t = (r_{1t}, \ldots, r_{nt}) \text{ denote the assets returns at time } t. \\
\text{Then, in our analysis we consider:} \\
\bullet & \text{ The standard deviation of the selected portfolio return} \\
& \text{The Sharpe ratio as } \frac{E[\hat{x}^* r_t]}{\text{Std}[\hat{x}^* r_t]}, \text{ where } r_t = 0 \\
& \text{The Rachev ratio as } \frac{\text{Cov}[\hat{x}^* r_t]^{\alpha}}{\text{Var}[\hat{x}^* r_t]^\beta}, \text{ where } r_t = 0 \text{ and } a = \beta = 0.1 \\
& \text{The max drawdown as } -\min x^* r'_t, \text{ which is the maximum loss achieved by a portfolio during the holding period}
\end{align*}

In our analysis we use six data sets, summarized in Table 1. The monthly data sets (FF25, 48Ind, 100Ind) are taken from Ken French’s website.\textsuperscript{30}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Data set & \# of assets & Time interval & Frequency & Source \\
\hline
3 & 100Ind & 100 & 01/1969-12/2011 & Monthly & K. French \\
4 & Stox50 & 32 & 01/2007-05/2013 & Weekly & Yahoo Finance \\
5 & FtseMib & 34 & 01/2007-05/2013 & Weekly & Yahoo Finance \\
6 & Ftse100 & 63 & 01/2007-05/2013 & Weekly & Yahoo Finance \\
\hline
\end{tabular}
\caption{List of data sets analyzed.}
\end{table}

\subsection*{4.1 In-sample analysis}

For each model described in section 2, we study the behavior of its optimal value (minimum risk) when varying the number \(K\) of assets in the portfolio. One of our main empirical findings is the scarce effect of diversification in terms of risk reduction when the portfolio size \(K\) does not belong to a certain range of values. Indeed, in all analyzed markets, we find that the risk measures, representing the objective functions of the models, achieve minimum values for a range of portfolio sizes corresponding to a significantly limited number of assets with respect to the total. Furthermore, these risk measures tend to increase when increasing the

\textsuperscript{28} E.g., Bruni et al. (2012, 2013); Cesaroni et al. (2015); DeMiguel et al. (2009a)
\textsuperscript{29} Rachev et al. (2008)
\textsuperscript{30} The weekly data sets (Stox50, FtseMib, Ftse100) are downloaded from http://finance.yahoo.com, and are publicly available at http://host.uniroma3.it/~cesarone/DataSets.htm.
Why small portfolios are preferable and how to choose them

portfolio size, thus contrasting the paradigm that the larger the diversification, the lower the risk. In Figures 2 and 3, we report some empirical evidence of this phenomenon for monthly (48Ind) and for weekly (Stoxx50) data sets. However, this behavior is similar for each data set analyzed. Figure 2 exhibits the boxplots of the different risk measures with respect to all considered in-sample windows by varying the portfolio size $K$. This means that, e.g., in the case of the 48Ind data set, for a fixed $K$ we have 377 values of risk, one for each in-sample window (i.e., one for each rebalancing of the portfolio). Similarly, we obtain Figure 3, where we examine the Stoxx50 data set. Note, however, that in the cases of weekly data sets for a fixed $K$, we have 32 in-sample windows (i.e., 32 values of risk). As mentioned above, the boxplot of the in-sample volatility generated by the EW portfolios corresponds to that of the FSMVEW portfolios when $K = n$, and it generally presents the highest median volatility. This feature is common to all data sets, and it suggests that a greater diversification does not always imply a risk reduction, i.e., increasing the number of assets in the portfolio could worsen its in-sample performance in terms of risk.

The empirical results of the FSMVEW portfolios could be compared to the findings obtained by Evans and Archer, and by further influential experiments in the literature, such as the well-known Fama’s experiment.\(^{31}\) The author finds that, in a market with 50 stocks, the effect of naïve diversification determines a remarkable reduction of the portfolio in-sample volatility, but only when it is included in the portfolio of up to 20 stocks. We refer to naïve diversification as an EW strategy with a random selection of $K$ out of $n$ available stocks. Indeed, he observes that adding further stocks in the portfolio does not yield a considerable improvement. More precisely, Fama claims that approximately 95% of the possible reduction deriving from diversification is achieved passing from 1 to 20 assets. However, we point out that our approach is significantly different from that of Fama, as well as from that of Evans and Archer. Indeed, we overcome one of the main weaknesses of their approach, which consists of stating results that are valid only on average. In other words, if one picks an arbitrary EW portfolio of a given size in a market, there is no guarantee that its risk will not be much larger than the average risk of all portfolios of the same size in that market, whereas the results obtained by the FSMVEW portfolios are those corresponding to the best EW portfolios for each given size with respect to volatility. The findings on the FSMVEW model

---

\(^{31}\) Fama (1976)
also highlight that when the EW strategy is combined with risk minimization (instead of a random selection of $K$ out of $n$ available stocks), the selected small portfolios show an improvement in terms of both volatility and robustness of their values obtained on all in-sample windows.

In addition, once we have obtained the optimal size of the minimum risk portfolio for each in-sample window, we examine the range spanned by these optimal sizes. In Figure 4, we show for the 100Ind data set the distribution of the optimal portfolio sizes (i.e., corresponding to the global minimum risk) for all models analyzed with respect to all in-sample windows. We can see that the global minimum risk portfolio never exceeds 15 stocks for the 100Ind data set. However, this behavior is almost the same in all the other considered data sets, with the only exception of Ftse100, where the optimal portfolio size is seldom around 20 stocks. Furthermore, given the global minimum risk on an in-sample window, we detect the smallest size of a portfolio whose risk is not more than 1% larger than that of the minimum risk portfolio, thus finding even smaller portfolios with satisfactory risk level. Then, we repeat this procedure for each in-sample window and for each portfolio model. In Figure 5, we report the distribution of these 101% min-risk optimal portfolio sizes for each model analyzed with respect to all in-sample windows. More precisely, for each in-sample window we consider all the cardinalities for which the corresponding portfolio has a risk at most 1% greater than that of the minimum risk portfolio. As highlighted from the four sub-figures (one for each portfolio), the 101% min-risk portfolios generally show a significant risk reduction with 10 stocks for the 100Ind data set. Furthermore, in most of the cases we can achieve it with just six stocks. However, the 101% min-risk portfolio for the other data sets never exceeds a size of 15, and generally at most 10 stocks are needed.

The most compelling result emerging from the in-sample analysis is the existence of a portfolio size range (whose location could depend on the number of assets for each market), where one can generally find the lowest values of risk for all models considered. Indeed, we find that the smallest risk portfolios generally require no more than 15 assets. Furthermore, it is almost always possible to find portfolios that are just 1% more risky than the smallest risk portfolios and contain no more than 10 assets.
Why small portfolios are preferable and how to choose them

4.2 Out-of-sample analysis

The second part of our analysis concerns the out-of-sample behavior of the portfolios. Our main goal is to confirm the finding, emerging from the in-sample analysis, that we can improve performances without investing in a large number of stocks.

Again, we consider the EW portfolio as a benchmark and, instead of focusing only on volatility reduction, we also compute the performance indices described in section 4. We start by verifying the behavior of the out-of-sample standard deviation. More precisely, we check whether this performance measure reaches an optimal value, or at least a good value, for small-size portfolios. We can see in figures 6(a) and 7(a) that, both for monthly (48Ind) and for weekly (Stoxx50) data sets, the standard deviations of the portfolio returns reach their minima for small sizes. For larger sizes, the portfolio volatility tends to increase with different growth rates. These increases, except for the FSMVEW, are due to the buy-in threshold constraints. Without these constraints, we should expect nearly flat curves.
Table 2: Standard deviation of the out-of-sample returns.

<table>
<thead>
<tr>
<th>K = 5</th>
<th>FF25</th>
<th>4BInd</th>
<th>100Ind</th>
<th>Ftse100</th>
<th>FtseMib</th>
<th>Stoxx50</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSMVaR</td>
<td>0.0448</td>
<td>0.0378</td>
<td>0.0556</td>
<td>0.0134</td>
<td>0.0185</td>
<td>0.0169</td>
</tr>
<tr>
<td>FSMVEW</td>
<td>0.0431</td>
<td>0.0397</td>
<td><strong>0.0424</strong></td>
<td>0.0139</td>
<td>0.0181</td>
<td>0.0170</td>
</tr>
<tr>
<td>FSMV</td>
<td>0.0431</td>
<td><strong>0.0377</strong></td>
<td>0.0443</td>
<td>0.0141</td>
<td><strong>0.0169</strong></td>
<td><strong>0.0166</strong></td>
</tr>
<tr>
<td>FSMAD</td>
<td><strong>0.0430</strong></td>
<td>0.0377</td>
<td>0.0446</td>
<td>0.0136</td>
<td>0.0177</td>
<td>0.0172</td>
</tr>
<tr>
<td>EW</td>
<td>0.0509</td>
<td>0.0488</td>
<td>0.0512</td>
<td>0.0210</td>
<td>0.0296</td>
<td>0.0288</td>
</tr>
<tr>
<td>K = 10</td>
<td>FF25</td>
<td>4BInd</td>
<td>100Ind</td>
<td>Ftse100</td>
<td>FtseMib</td>
<td>Stoxx50</td>
</tr>
<tr>
<td>FSMVaR</td>
<td>0.0447</td>
<td>0.0376</td>
<td>0.0536</td>
<td>0.0133</td>
<td>0.0189</td>
<td>0.0175</td>
</tr>
<tr>
<td>FSMVEW</td>
<td>0.0441</td>
<td>0.0416</td>
<td><strong>0.0428</strong></td>
<td><strong>0.0125</strong></td>
<td>0.0203</td>
<td>0.0184</td>
</tr>
<tr>
<td>FSMV</td>
<td><strong>0.0432</strong></td>
<td>0.0372</td>
<td>0.0438</td>
<td>0.0126</td>
<td>0.0170</td>
<td><strong>0.0168</strong></td>
</tr>
<tr>
<td>FSMAD</td>
<td><strong>0.0432</strong></td>
<td><strong>0.0369</strong></td>
<td>0.0440</td>
<td>0.0128</td>
<td><strong>0.0167</strong></td>
<td>0.0174</td>
</tr>
<tr>
<td>EW</td>
<td>0.0509</td>
<td>0.0488</td>
<td>0.0512</td>
<td>0.0210</td>
<td>0.0296</td>
<td>0.0288</td>
</tr>
<tr>
<td>K = 15</td>
<td>FF25</td>
<td>4BInd</td>
<td>100Ind</td>
<td>Ftse100</td>
<td>FtseMib</td>
<td>Stoxx50</td>
</tr>
<tr>
<td>FSMVaR</td>
<td>0.0448</td>
<td>0.0378</td>
<td>0.0537</td>
<td>0.0134</td>
<td>0.0188</td>
<td>0.0179</td>
</tr>
<tr>
<td>FSMVEW</td>
<td>0.0458</td>
<td>0.0419</td>
<td><strong>0.0435</strong></td>
<td>0.0128</td>
<td>0.0218</td>
<td>0.0201</td>
</tr>
<tr>
<td>FSMV</td>
<td><strong>0.0433</strong></td>
<td>0.0371</td>
<td>0.0439</td>
<td>0.0127</td>
<td>0.0173</td>
<td><strong>0.0170</strong></td>
</tr>
<tr>
<td>FSMAD</td>
<td>0.0434</td>
<td><strong>0.0367</strong></td>
<td>0.0436</td>
<td><strong>0.0126</strong></td>
<td><strong>0.0171</strong></td>
<td>0.0175</td>
</tr>
<tr>
<td>EW</td>
<td>0.0509</td>
<td>0.0488</td>
<td>0.0512</td>
<td>0.0210</td>
<td>0.0296</td>
<td>0.0288</td>
</tr>
</tbody>
</table>

Table 3: Sharpe ratio of the out-of-sample returns.

<table>
<thead>
<tr>
<th>K = 5</th>
<th>FF25</th>
<th>4BInd</th>
<th>100Ind</th>
<th>Ftse100</th>
<th>FtseMib</th>
<th>Stoxx50</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSMVaR</td>
<td>0.244</td>
<td>0.279</td>
<td>0.133</td>
<td>0.334</td>
<td>0.124</td>
<td><strong>0.211</strong></td>
</tr>
<tr>
<td>FSMVEW</td>
<td>0.267</td>
<td>0.285</td>
<td><strong>0.249</strong></td>
<td>0.247</td>
<td>0.119</td>
<td>0.165</td>
</tr>
<tr>
<td>FSMV</td>
<td>0.268</td>
<td><strong>0.293</strong></td>
<td><strong>0.249</strong></td>
<td>0.257</td>
<td><strong>0.126</strong></td>
<td>0.207</td>
</tr>
<tr>
<td>FSMAD</td>
<td><strong>0.276</strong></td>
<td>0.293</td>
<td>0.247</td>
<td>0.244</td>
<td>0.086</td>
<td>0.198</td>
</tr>
<tr>
<td>EW</td>
<td>0.276</td>
<td>0.242</td>
<td>0.215</td>
<td>0.173</td>
<td>-</td>
<td>0.026</td>
</tr>
<tr>
<td>K = 10</td>
<td>FF25</td>
<td>4BInd</td>
<td>100Ind</td>
<td>Ftse100</td>
<td>FtseMib</td>
<td>Stoxx50</td>
</tr>
<tr>
<td>FSMVaR</td>
<td>0.248</td>
<td>0.278</td>
<td>0.142</td>
<td><strong>0.298</strong></td>
<td>0.120</td>
<td>0.175</td>
</tr>
<tr>
<td>FSMVEW</td>
<td>0.290</td>
<td>0.268</td>
<td><strong>0.256</strong></td>
<td>0.284</td>
<td>0.091</td>
<td>0.099</td>
</tr>
<tr>
<td>FSMV</td>
<td>0.27</td>
<td>0.289</td>
<td>0.242</td>
<td>0.276</td>
<td>0.112</td>
<td><strong>0.178</strong></td>
</tr>
<tr>
<td>FSMAD</td>
<td>0.276</td>
<td><strong>0.294</strong></td>
<td>0.253</td>
<td>0.266</td>
<td><strong>0.122</strong></td>
<td>0.162</td>
</tr>
<tr>
<td>EW</td>
<td>0.264</td>
<td>0.242</td>
<td>0.215</td>
<td>0.173</td>
<td>-</td>
<td>0.026</td>
</tr>
<tr>
<td>K = 15</td>
<td>FF25</td>
<td>4BInd</td>
<td>100Ind</td>
<td>Ftse100</td>
<td>FtseMib</td>
<td>Stoxx50</td>
</tr>
<tr>
<td>FSMVaR</td>
<td>0.250</td>
<td>0.274</td>
<td>0.143</td>
<td><strong>0.281</strong></td>
<td>0.124</td>
<td>0.161</td>
</tr>
<tr>
<td>FSMVEW</td>
<td>0.296</td>
<td>0.260</td>
<td>0.249</td>
<td>0.278</td>
<td>0.042</td>
<td>0.072</td>
</tr>
<tr>
<td>FSMV</td>
<td>0.272</td>
<td>0.289</td>
<td>0.242</td>
<td>0.269</td>
<td>0.098</td>
<td><strong>0.179</strong></td>
</tr>
<tr>
<td>FSMAD</td>
<td>0.280</td>
<td><strong>0.293</strong></td>
<td><strong>0.257</strong></td>
<td>0.260</td>
<td>0.109</td>
<td>0.152</td>
</tr>
<tr>
<td>EW</td>
<td>0.264</td>
<td>0.242</td>
<td>0.215</td>
<td>0.173</td>
<td>-</td>
<td>0.026</td>
</tr>
</tbody>
</table>
Why small portfolios are preferable and how to choose them

### Table 4: Rachev ratio of the out-of-sample returns.

<table>
<thead>
<tr>
<th>$K$ = 5</th>
<th>FF25</th>
<th>48Ind</th>
<th>100Ind</th>
<th>Ftse100</th>
<th>FtseMib</th>
<th>Stoxx50</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSMVaR</td>
<td>1.216</td>
<td>1.281</td>
<td>0.877</td>
<td>1.618</td>
<td>1.162</td>
<td>1.167</td>
</tr>
<tr>
<td>FSMVEW</td>
<td>1.201</td>
<td>1.260</td>
<td>1.049</td>
<td>1.614</td>
<td>1.048</td>
<td>1.101</td>
</tr>
<tr>
<td>FSMV</td>
<td>1.245</td>
<td>1.403</td>
<td>1.043</td>
<td>1.514</td>
<td>1.164</td>
<td>1.205</td>
</tr>
<tr>
<td>FSMSMAD</td>
<td>1.25</td>
<td>1.363</td>
<td>1.032</td>
<td>1.382</td>
<td>1.07</td>
<td>1.218</td>
</tr>
<tr>
<td>EW</td>
<td>1.150</td>
<td>1.200</td>
<td>0.992</td>
<td>1.007</td>
<td>0.861</td>
<td>0.910</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$K$ = 10</th>
<th>FF25</th>
<th>48Ind</th>
<th>100Ind</th>
<th>Ftse100</th>
<th>FtseMib</th>
<th>Stoxx50</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSMVaR</td>
<td>1.215</td>
<td>1.271</td>
<td>0.892</td>
<td>1.459</td>
<td>1.145</td>
<td>0.861</td>
</tr>
<tr>
<td>FSMVEW</td>
<td>1.199</td>
<td>1.284</td>
<td>1.047</td>
<td>1.477</td>
<td>0.963</td>
<td>1.007</td>
</tr>
<tr>
<td>FSMV</td>
<td>1.244</td>
<td>1.360</td>
<td>1.027</td>
<td>1.514</td>
<td>1.147</td>
<td>1.150</td>
</tr>
<tr>
<td>FSMSMAD</td>
<td>1.234</td>
<td>1.343</td>
<td>1.033</td>
<td>1.333</td>
<td>1.147</td>
<td>1.185</td>
</tr>
<tr>
<td>EW</td>
<td>1.150</td>
<td>1.200</td>
<td>0.992</td>
<td>1.007</td>
<td>0.861</td>
<td>0.910</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$K$ = 15</th>
<th>FF25</th>
<th>48Ind</th>
<th>100Ind</th>
<th>Ftse100</th>
<th>FtseMib</th>
<th>Stoxx50</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSMVaR</td>
<td>1.217</td>
<td>1.245</td>
<td>0.896</td>
<td>1.397</td>
<td>1.123</td>
<td>1.064</td>
</tr>
<tr>
<td>FSMVEW</td>
<td>1.195</td>
<td>1.261</td>
<td>1.052</td>
<td>1.347</td>
<td>0.861</td>
<td>0.907</td>
</tr>
<tr>
<td>FSMV</td>
<td>1.244</td>
<td>1.343</td>
<td>1.024</td>
<td>1.426</td>
<td>1.135</td>
<td>1.135</td>
</tr>
<tr>
<td>FSMSMAD</td>
<td>1.234</td>
<td>1.343</td>
<td>1.047</td>
<td>1.320</td>
<td>1.103</td>
<td>1.132</td>
</tr>
<tr>
<td>EW</td>
<td>1.150</td>
<td>1.200</td>
<td>0.992</td>
<td>1.007</td>
<td>0.861</td>
<td>0.910</td>
</tr>
</tbody>
</table>

### Table 5: Max drawdown of the out-of-sample returns.

<table>
<thead>
<tr>
<th>$K$ = 5</th>
<th>FF25</th>
<th>48Ind</th>
<th>100Ind</th>
<th>Ftse100</th>
<th>FtseMib</th>
<th>Stoxx50</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSMVaR</td>
<td>0.209</td>
<td>0.150</td>
<td>0.578</td>
<td>0.027</td>
<td>0.042</td>
<td>0.031</td>
</tr>
<tr>
<td>FSMVEW</td>
<td>0.214</td>
<td>0.175</td>
<td>0.201</td>
<td>0.029</td>
<td>0.045</td>
<td>0.036</td>
</tr>
<tr>
<td>FSMV</td>
<td>0.198</td>
<td>0.123</td>
<td>0.206</td>
<td>0.030</td>
<td>0.037</td>
<td>0.031</td>
</tr>
<tr>
<td>FSMSMAD</td>
<td>0.201</td>
<td>0.121</td>
<td>0.206</td>
<td>0.039</td>
<td>0.041</td>
<td>0.032</td>
</tr>
<tr>
<td>EW</td>
<td>0.261</td>
<td>0.259</td>
<td>0.262</td>
<td>0.052</td>
<td>0.064</td>
<td>0.061</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$K$ = 10</th>
<th>FF25</th>
<th>48Ind</th>
<th>100Ind</th>
<th>Ftse100</th>
<th>FtseMib</th>
<th>Stoxx50</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSMVaR</td>
<td>0.210</td>
<td>0.145</td>
<td>0.501</td>
<td>0.033</td>
<td>0.042</td>
<td>0.034</td>
</tr>
<tr>
<td>FSMVEW</td>
<td>0.219</td>
<td>0.201</td>
<td>0.201</td>
<td>0.029</td>
<td>0.041</td>
<td>0.041</td>
</tr>
<tr>
<td>FSMV</td>
<td>0.198</td>
<td>0.126</td>
<td>0.209</td>
<td>0.023</td>
<td>0.039</td>
<td>0.031</td>
</tr>
<tr>
<td>FSMSMAD</td>
<td>0.203</td>
<td>0.125</td>
<td>0.212</td>
<td>0.031</td>
<td>0.039</td>
<td>0.034</td>
</tr>
<tr>
<td>EW</td>
<td>0.261</td>
<td>0.259</td>
<td>0.262</td>
<td>0.052</td>
<td>0.064</td>
<td>0.061</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$K$ = 15</th>
<th>FF25</th>
<th>48Ind</th>
<th>100Ind</th>
<th>Ftse100</th>
<th>FtseMib</th>
<th>Stoxx50</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSMVaR</td>
<td>0.212</td>
<td>0.141</td>
<td>0.493</td>
<td>0.029</td>
<td>0.040</td>
<td>0.036</td>
</tr>
<tr>
<td>FSMVEW</td>
<td>0.235</td>
<td>0.206</td>
<td>0.213</td>
<td>0.027</td>
<td>0.050</td>
<td>0.045</td>
</tr>
<tr>
<td>FSMV</td>
<td>0.199</td>
<td>0.129</td>
<td>0.210</td>
<td>0.027</td>
<td>0.039</td>
<td>0.031</td>
</tr>
<tr>
<td>FSMSMAD</td>
<td>0.203</td>
<td>0.127</td>
<td>0.210</td>
<td>0.027</td>
<td>0.040</td>
<td>0.035</td>
</tr>
<tr>
<td>EW</td>
<td>0.261</td>
<td>0.259</td>
<td>0.262</td>
<td>0.052</td>
<td>0.064</td>
<td>0.061</td>
</tr>
</tbody>
</table>
However, the buy-in threshold constraints are necessary to eliminate unrealistically small trades that can otherwise be included in an optimal portfolio. In Figures 6(b) and 7(b), we show the values of two other performance measures, namely the Rachev and Sharpe ratios, for the same data sets. As for the standard deviation, each model generally tends to provide the best values of the latter performance measures for small sizes. Furthermore, these values almost always decay when the portfolio size approaches $n$. This behavior provides further support for the idea of improving the performances of a portfolio by limiting the number of its stocks.

In addition to the graphical evidence, where only the most representative results are shown, we performed an extensive comparative analysis on all data sets considered. Since describing the results for all data sets and for all portfolio sizes is impractical, we report here the out-of-sample analysis for only three fixed sizes: $K = 5, 10, 15$. This choice is based on the observation that $K = 5, 10, 15$ generally belong to the optimal ranges in which the various models achieve the in-sample lowest risk for each data set. In Table 2, we provide the standard deviation of the out-of-sample returns for $K = 5, 10, 15$ for each model and data set analyzed. It is remarkable that the EW portfolio has almost always the worst performance, with the single exception of the 100Ind market, where the FSMCVaR portfolios generate the highest standard deviation. In Table 3, we report the Sharpe ratio of the out-of-sample returns for the same portfolio sizes of the previous table and for each model and data set analyzed. Note that when the portfolio excess return is negative, some gain-to-risk ratios have no meaning, thus we report “-”. Again, we observe that the EW portfolio yields the worst performances compared with those of the other models, with the exception of the FSMCVaR portfolios for the 100Ind and FF25 markets. Similar considerations can be made about the Rachev ratio of the out-of-sample returns shown in Table 4. Indeed, again the EW portfolio tends to be the worst choice, with the only exception of the 100Ind data set. We also observe that for $K = 10$, the FSMV model seems to be preferable since it provides the best results for four data sets out of six, while for $K = 15$, it presents the best performances for five data sets out of six. The last performance measure considered in our analysis is the max drawdown, which is the worst out-of-sample loss achieved by a portfolio, as described in section 4. Table 5 shows that, again, the EW portfolio always has the worst performance for the prescribed sizes $K = 5, 10, 15$, with the exception of the 100Ind market, where the FSMCVaR portfolios provide the worst loss. On the other hand, although no single model is clearly superior, we observe that the FSMV portfolios present the best values for three data sets out of six for $K = 5$, and for four data sets out of six for $K = 10$ and for $K = 15$.

5. Conclusions

The concept of diversification is not well defined and the measures of diversification are continuously evolving. However, the qualitative idea of diversification is to not overly concentrate the investments in very few stocks. Indeed, the role of diversification is to reduce risk by diversifying it as much as possible. In this work, we investigated the possible benefits and disadvantages of enlarging the portfolio size in several portfolio selection models with respect to various measures of performance. Similarly to various previous findings, but with a substantially different approach, our empirical results show that in most cases limiting the size of the selected portfolio improves both the in-sample and the out-of-sample performance. We might call this a small portfolio effect. These results are somewhat in line with the tendency described by DeMiguel et al., where an improved out-of-sample performance is often observed for the 1-norm-constrained minimum-variance portfolios. The analogy is based on the observation that the 1-norm is often regarded as an approximation of the 0-norm, i.e., the size of the portfolio. Further studies are underway to investigate the validity of this small portfolio effect with respect to other risk and performance measures and in larger markets.

32 E.g., Fragkiskos (2013); Meucci (2009) and references therein.
33 DeMiguel et al. (2009a)
Why small portfolios are preferable and how to choose them

References
Markowitz H (1952) Portfolio selection. J Financ 7:77-91


116 The Journal of Financial Perspectives
Ring-fencing cross-border banks: an effective supervisory response?

Katia D’Hulster
Lead Financial Sector Specialist at the World Bank

Inci Otker-Robe
Division Chief, Financial Analysis, Monetary and Capital Markets Department, International Monetary Fund (IMF)“

Disclaimer: The findings, interpretations and conclusions expressed in this article are entirely those of the authors. They do not represent the views of the International Monetary Fund, the International Bank for Reconstruction and Development or the World Bank and its affiliated organizations.

This article was previously published in the Journal of Banking Regulation in 2015. This is an updated and edited version, published with permission.
Abstract
While international policymakers are making good progress on the important work of global resolution and the preparation of recovery and resolution plans, a growing number of supervisors, home as well as hosts, are resorting to territorial approaches. Higher capital ratios, dividend restrictions, restrictions on liquidity flows and even forced subsidiarization are gaining renewed popularity. The objective of these territorial approaches is to protect the interests of the domestic stakeholders of a foreign bank and to limit the effects of cross-border contagion. This type of “ring-fencing” has a negative connotation as it comes at a cost for banks and the efficiency of the overall global financial system. This article addresses the following questions: (1) What makes prudential supervisors more likely to ring-fence?; (2) Do all forms of ring-fencing really deserve this bad reputation?; (3) What are the risks that these measures are addressing and which instruments have been used?; and (4) What are the implications of ring-fencing for the banking group, financial stability in the home and host country, as well as global financial stability?
1. Introduction

"Ring-fencing" has become a fashionable word in regulatory circles. Federal regulators in the US are implementing the so-called Volcker Rule, requiring the ring-fencing of risky assets in the trading book of a bank from its retail banking operations.² It should be noted that the Office of the Comptroller of the Currency (OCC) is currently looking into amending the rule.³ Similarly, proposals for ring-fencing "core" banking activities in the UK have been put in place following the Vickers Commission report.⁴ ⁵ These structural measures primarily address the "too important to fail" problem of systemically important financial institutions (SIFIs) by reducing the risk of failure and minimizing the impact, if they do fail.⁶

But, there is another form of ring-fencing that is quietly gaining renewed popularity among supervisors. It is the geographical separation of part of a cross-border banking group from its parent or other affiliates on a permanent or temporary basis.⁷ This type of ring-fencing can take the form of higher capital and liquidity requirements, dividend restrictions, restrictions on liquidity flows, and even forced subsidiarization to make individual parts of the banking group self-sufficient and to protect the domestic assets of a bank from cross-border contagion. So, while international policymakers are making good progress on the important work of global resolution and the preparation of recovery and resolution plans, a growing number of banking supervisors, home as well as hosts, are resorting to these territorial approaches. Their objectives are to protect the interests of their domestic stakeholders of a foreign bank and to limit the effects of cross-border contagion in the intervening period. These ring-fencing measures are generally described by adverse terms, such as fragmentation, home bias,⁸ territorial approaches, balkanization, financial protection and nationalism, and even moves toward deglobalization.

Of course, ring-fencing is not the optimal outcome and there is broad agreement that it comes at a cost for banking groups and the efficiency of the overall global financial system. But, is doing nothing a better alternative for supervisors, many of whom are still facing the same weaknesses in "global resolvability" as during the global financial crisis? Why do supervisors ring-fence and which supervisors are more likely to do so? Do all forms of ring-fencing really deserve this bad reputation? What are the risks these measures are addressing and which instruments have been used? What are the implications of ring-fencing for the banking group, financial stability in the home and host country, as well as global financial stability? This article focuses on geographical ring-fencing measures, and assesses their potential benefits and costs as an instrument for crisis prevention and management. It also discusses the various motivations for adoption of ring-fencing measures.⁹

2. What is ring-fencing?

Few clear definitions of ring-fencing can be found in the literature. Schwarcz defines ring-fencing, in a general way, as "legally deconstructing a firm in order to more optimally reallocate and reduce risk."¹⁰ He states that this deconstruction can occur in three ways: (1) by separating risky assets from the firm; (2) by preventing the firm itself from engaging in risky activities or investing in risky assets; or (3) by protecting the firm from affiliate or bankruptcy risks, usually referred to as making a firm "bankruptcy remote". Building on this third theme, Schwarcz defines two related functions of ring-fencing that play a vital role in the cross-border context. The first is to ensure that a firm is able to operate on a stand-alone basis, even if its affiliated firms fail. The second is to protect a firm from being taken advantage of by its affiliated firms.

The instruments that prudential supervisors use to address these two related functions have been described by Cerruti et al. as "different restrictions on intra-group cross-border transfers imposed by the host or home country regulators."¹¹

---

² Board of Governors of the Federal Reserve System (2013)
³ Jopson (2017) The Financial Choice Act, which is now with the US Senate, also seeks to repeal the Volcker Rule. See Financial Services Committee (2017)
⁴ Independent Commission on Banking (2011)
⁵ Prudential Regulation Authority (2014); Prudential Regulation Authority (2017)
⁶ Chow and Surti (2011)
⁷ D’Hulster (2012)
⁸ This article refers to "territorial bias" in home and host countries; it is referred to as "home" bias, regardless of it being exercised by the "home" or the "host" supervisor. In theory, the term "host bias" should be used to refer to home bias exercised by the host supervisor.
⁹ This article focuses on prudential supervision in an international context. Consumer protection, and other aspects of regulation and supervisory arrangements in the EU lie outside of its scope.
¹⁰ Schwarcz (2013)
¹¹ Cerruti et al. (2010)
Song describes the objective and the most important tools more broadly: “ring-fencing involves isolating the bank from other companies in the group by taking several actions, for instance, by: (i) prohibiting or placing severe limits on the financial exposure of the bank vis-à-vis other companies in the group (ii) restricting the volume of funding the bank receives from companies in the group (iii) ensuring that directors and management of the bank can operate the bank independently of the group management”.12

For the purposes of this article, we will use the term “geographical ring-fencing” or the more general term “territorial approaches.” This type of ring-fencing is imposed unilaterally by prudential supervisors with the objective of protecting a bank’s domestic assets so that they can be seized and liquidated under local law in case of failure of the whole banking group or other entities of the group. By doing this, the interests of domestic stakeholders, such as taxpayers, depositors, shareholders, creditors and deposit insurers, are better protected in times of stress as the effects of cross-border contagion would be limited. The use of territorial approaches thus results in a measured delineation of the local operations from other cross-border affiliates of the banking group on a permanent or temporary basis. We refer to the degree of this separation of the local operations as “territorial bias” or “home bias.”

Applying this definition, the most drastic form of home bias is the Stand Alone Subsidiary Model (SAS model).13 Under this model, affiliates of a banking group operate independently from each other and the parent does not consider itself responsible for the rescue of individual entities that form part of the group. Complete separation means no intercompany transactions (hedging, loans, guarantees and outsourcing of critical activities to a group company or transfer of funds). Each entity would hold sufficient capital and liquidity to survive on its own, thus reducing interconnectedness and contagion risk between the entities in the group.

Geographical ring-fencing is usually applied by host supervisors. In recent years, many host supervisors have taken actions to ensure that the operations of foreign banks in their jurisdictions become better insulated and more resilient on a stand-alone basis. Many host supervisors have experienced the speediness and impact of cross-border contagion during the global financial crisis first hand.

That said, territorial bias in prudential regulation and supervision of cross-border banks has also been introduced by home supervisors. Countries with banking systems characterized by considerable asset overhangs abroad are particularly vulnerable to solvency, legal and operational problems in the event of financial stress.14 The measures introduced by the Austrian supervisory authorities are a clear example. In November 2011, the Austrian National Bank and the Financial Market Authority introduced a measure to make business models of Austrian banks more sustainable. Credit growth was made conditional on the growth of sustainable local refinancing (comprising mainly local deposits, but also local issuance activity and supranational funding by international financial institutions). Foreign subsidiaries that are particularly exposed were required to ensure that the ratio of new loans to local refinancing (i.e., the loan-to-deposit ratio including local refinancing) does not exceed 110%.15

The business model adopted by Spanish parent banks operating in Latin America and driven by pressures from the home country supervisor is another example of a territorial approach adopted by the home supervisor.16,17 This model minimized within group contagion and mitigated the transmission of the financial crisis in the US and Europe to Latin American banks.18

12 Song (2004)
13 Hoelscher et al. (2011)
14 Swiss Federal Market Supervisor Authority (2010)
15 Austrian National Bank (2011)
16 Imam et al. (2013)
17 Committee on the Global Financial System (2010)
18 Flechter et al. (2011)
3. Why authorities ring-fence: policy and structural reasons

Authorities introduce home bias in their prudential regulation and supervision for a variety of reasons. Broadly speaking, these can be split into policy reasons and more structural reasons.

3.1 Overarching policy reasons

Three policy drivers for ring-fencing can be identified: (1) the lack of a globally enforceable resolution mechanism; (2) the existence of information asymmetries; and (3) the desire to protect credit supply.

3.2 Lack of global enforceable resolution mechanism

A key reason for ring-fencing is that “global banking institutions are global in life and national in death.” 19 That is, when a crisis occurs, it is domestic central banks that have to provide lender-of-last-resort facilities to the ailing banks and national governments that have to provide public support. Without a clear understanding of what will happen when a bank fails or is likely to fail, the interests of the home and host supervisors will diverge at the very moment problems start to emerge. Home supervisors will want maximum transferability of liquidity and assets to reduce the probability and impact of group-wide problems, and host supervisors will wish to keep good-quality assets within their borders. In spite of their noble intentions for supervisory cooperation, home and host supervisors will each seek to minimize the losses accruing to their stakeholders in their own jurisdictions. It is, therefore, only natural for national authorities to be forward looking, to plan for the worst, and to try to minimize losses for their local banking system and taxpayers.

After all, that is exactly what prudential banking supervisors are paid to do. When there are uncertainties regarding resolution, ring-fencing brings immediate comfort and is legitimate from the perspective of the national supervisor. It is only when considering the bigger picture that ring-fencing, particularly sudden supervisory ring-fencing during an emerging crisis, has the potential to lead to a “scramble for assets,” put stress on the banking group as a whole and prevent good crisis management.

Policymakers have been working intensively on addressing this lack of enforceable and effective mechanisms for crisis resolution for cross-border banking groups at a global level. The public policy stance on resolution has been to shift the burden to the private sector (bail-in) instead of the public sector (bail-out) and to improve crisis preparedness and early intervention. Bail-in instruments will allow banks’ liabilities to be written off or converted into equity so that the institution can continue as a going concern, while giving the authorities time to reorganize or wind down parts of the business in an orderly manner.

Supervisors will also be better prepared for a crisis, as bank-specific resolution strategies and plans, cooperation agreements and resolvability assessments have been prepared for the Global Systemically Important Banks (G-SIBs). These have given supervisors a better insight into the structure and organizational setup of a banking group, including where liquid and good-quality assets are located, and what the implications of the failure of the banking group might be for each national supervisor. In other words, crisis preparedness makes the structural imbalances20 in the banking group visible before it enters into difficulty.

In practice, however, many obstacles still exist. Bail-in is unlikely to provide absolute protection for host supervisors in the future, for a number of reasons. First, there will always be a risk that public funds may be needed to preserve financial stability, even if only for a short period or only for liquidity reasons. For example, loss-absorbent funding instruments may be exhausted or not available, the instruments may be held by investors that are not fit and proper to become bank shareholders, or they may be located in another jurisdiction and not transferrable to the location where the losses are actually incurred. Thus, there still remains a risk, even though it is smaller than before the global financial crisis, that the taxpayer will end up bearing losses.

---

19 Marvyn King, Governor of the Bank of England (2009)

20 In the context of this article, supervisors are mainly concerned with structural geographical imbalances. These imbalances can occur in the asset-liability structure (asset quality, asset encumbrance and capital buffers can differ by country, etc.), in the liquidity and in loss absorbency (buffers of loss-absorbent instruments and capital can differ by country, etc.).
Second, these proposed international reform policies are still supported by nonbinding and non-enforceable contracts that are akin to statements of intent. Although the number of participants has been expanded to include resolution agencies and central banks, reliance has, once again, been placed on “soft law” as there are no binding mechanisms to enforce cooperation. In a crisis, there is always the risk that individual authorities make decisions driven by political expediencies and deviate from agreed mechanisms and frameworks. “Pulling the trigger” to allocate losses to the creditor, or to convert a financial instrument into equity, is to be decided by the relevant supervisory authority and thus depends on the ad hoc decision of a jurisdiction. For political or national financial stability reasons, it may well decide not to bail in. This could be the case when the holders of the capital instrument are retail investors, or the public pension fund or a supranational institution where the relevant jurisdiction is the major shareholder or investors that are not deemed politically acceptable to carry the burden. In those instances, the actual resolution may look very different from what was expected and this may leave other national authorities bewildered and the global financial system endangered.

Third, the lack of harmonization of creditor hierarchies between jurisdictions remains an important obstacle to the resolution of branches of international banks. Uninsured depositors of a branch in the host country may, for example, have a higher status in the hierarchy than other creditors, while in the home country, they are at the same level. As a result, the host country will oppose the pooling of assets.

Fourth, it should be kept in mind that these reforms are still work in progress and that they will be raising new implementation challenges for prudential supervisors, central banks and resolution authorities along the way. The proposed reforms are for G-SIBs only and whether they can be easily implemented for banks that are regional players or banks that are systemic in only a number of countries remains to be seen. Many of the proposed measures also require fundamental legal changes and thus come with long implementation timelines.

Finally, adequate loss-absorbing capacity is not in itself a sufficient condition for ensuring effective resolution. There are many other legislative reforms that still need to be addressed, such as the full implementation of the key attributes, the removal of obstacles to cross-border resolution, structural changes to ensure resolvability, enhanced cross-border information sharing and the prevention of large-scale termination of financial contracts in resolution. Therefore, at a global level, major international policymakers still assume that resolution regimes and powers are likely to remain national in the longer term.

3.3 Information asymmetries

A second important rationale for ring-fencing measures is information asymmetries in the absence of adequate arrangements for information exchange. As noted by the Basel Committee, actions taken in a crisis can be influenced by uncertainty about the way weak or failing institutions will be handled in a crisis. Ring-fencing measures may, therefore, be driven by lack of information, doubts about the reliability and timeliness of information, or uncertainty about how the resolution process will function in the jurisdictions where the financial institution operates and where it is headquartered. For example, at the height of the sovereign debt crisis in Europe, ring-fencing had been applied to protect banks in host jurisdictions from political or sovereign vulnerabilities in home countries of parent banks and risk of redenomination. Doubts about fiscal sustainability, quality of supervision, scope of financial regulation or political stability in the home country can all be grounds for ring-fencing. In these instances, supervisors made conservative assumptions and ring-fenced as a precautionary measure.

The simple threat of ring-fencing can also be used by host supervisors as a way to persuade home supervisors to enhance supervisory cooperation by stepping up information sharing or addressing particular host concerns.

---

23 Financial Stability Board (2011)
24 Basel Committee on Banking Supervision (2010)
25 Shortcomings in information sharing may sometimes be associated with difficulties and costs in obtaining, processing and analyzing information in a timely manner, and sometimes be a result of home authorities’ reluctance to provide complete information that they perceive as negative to host authorities, with the fear of spreading distress and prompting host authorities to take adverse measures. Host authorities may take action to protect local depositors and creditors and ring-fence assets in the absence of such information. From the host supervisors’ perspective, the home supervisor has the most reliable information about the parent’s group health and is thus better equipped to assess the consequences of ring-fencing (including obtaining liquidity from the central bank in the home country).
In cases, where the host operation is material to the whole banking group, this instrument may make the home supervisor more cooperative to avoid escalating ring-fencing actions by the host.

3.4 Protection of supply of credit
Finally, ring-fencing may also be applied to protect the supply of credit and prevent excessive deleveraging in the host country resulting from parents pulling out funding and capital from their host operations. The risk of faster bank deleveraging in host countries than in home countries was identified from surveys.26 Host supervisors have thus ring-fenced local liquidity and capital to support local lending, while international banking groups prepare to comply with higher capital and liquidity requirements under Basel III.

3.5 Structural characteristics of the financial system
In addition to these policy factors, there are a number of other factors associated more with structural characteristics of the financial system that can make home or host authorities more or less likely to resort to ring-fencing decisions.

3.6 Location of financial stress
Host supervisors have a strong incentive to ring-fence when problems in the parent bank or in the home country are suspected. The moves by a number of host country authorities to ring-fence the local operations of peripheral Euro Area countries during the escalation of the Euro Area crisis are a case in point. Host supervisors may have legitimate prudential concerns, and protect the safety and soundness of their domestic financial system through effective firewalls between the parent and the local operations. Accordingly, the home supervisor has an incentive to ring-fence when problems are suspected in the host operations.

3.7 Business model of the banking group
There is a wide spectrum of business models in international banking.27 Business models can be positioned along a continuum of increasing self-sufficiency. On one end of the spectrum, we find a typical branch structure, where key functions are centralized. This centralization can occur in various areas, including corporate control functions (setting of strategy), operational functions (sharing of back-office functions, IT infrastructure, accounting, information and telecommunications infrastructure), treasury functions and brand name. Centralized institutions also use funding surpluses in one entity to fill funding deficits in another. Affiliates are, thus, not independent entities from the parent, but they lend and trade using the capital base of the parent and operate under the jurisdiction of the home country.

On the other end of the spectrum is the typical stand-alone subsidiary structure. These are self-sufficient entities, operating nationally with effective firewalls between other parts of the group. No support from the parent company or other group members is assumed, and each entity has sufficient capital and liquidity to survive on its own. These banks generally rely heavily on local deposits for their funding and their retail-based business models are strongly anchored in the host country. Many institutions fall between these two extremes, with hybrid structures; however, centralized structures are more common.28-29

When dealing with more centralized business models, host supervisors will typically rely on cooperation with the home supervisor for the purposes of ongoing prudential supervision. As a second line of defense, however, host supervisors are likely to take a more territorial stance by imposing some restrictions on these centralized models or requiring additional local liquidity and capital buffers, particularly if the entity in their jurisdiction is allowed to take retail deposits. Indeed, the more one moves to a more centralized business model, the less likely it is that a foreign entity can survive, once funding, operational and managerial connections have been cut as a result of a problem in the global or home financial system. When the source of stress is at the core of the banking group, centralized models increase contagion risk and stand-alone business models help to reduce contagion risk. Conversely, centralized models can dampen contagion risk to the rest of the host’s financial system when the source of stress is in the host country or is an external factor, by providing resources locally.

26 International Monetary Fund (2012)
27 Nguyen (2010)
28 Basel Committee on Banking Supervision (2012)
29 The high-level analysis by the Joint Forum of 31 financial institutions headquartered in 31 jurisdictions found that the majority of institutions said that centralized capital and liquidity management systems were in place. Respondents also indicated that the way they managed capital and liquidity within the group are the key drivers in their decisions about intragroup transactions and the support they used.
In addition to centralization, the complexity of organizational structures of big banking groups is often a serious concern for both the host and the home supervisor. A global banking group can consist of thousands of legal entities and branches with complex financial flows and guarantees between them. This complexity makes it more difficult to introduce effective firewalls.

3.8 The geographical risk profile of the banking group
Herring distinguishes three dimensions linked to the geographical risk profile or the systemic dimensions of a banking group:³⁰ (1) whether the parent bank is considered to be of systemic importance in the home country; (2) whether the foreign operations are of significance to the solvency of the parent bank;³¹ and (3) whether the foreign operations are systemically important in the host country.

Foreign operations of a systemic nature in the host country will make the host supervisor more likely to ring-fence as failure of the banking group would have a bigger impact on the domestic financial system and the economy in this jurisdiction. Home supervisors, on the other hand, will be more likely to ring-fence if the bank is systemic in the home country, particularly if the banking group has significant foreign operations.

Tables 1 and 2 summarize the factors discussed for host and home supervisors, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Location of financial stress</th>
<th>Financial stress in the home country or banking group.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Location of financial stress</td>
<td>Includes safety and soundness concerns directly linked to the parent bank or banking group; political or sovereign vulnerabilities; doubts about quality of supervision; uncertainties related to political stability and fiscal sustainability and home country resolution process for failing institutions.</td>
</tr>
<tr>
<td>II.</td>
<td>Business model of the banking group</td>
<td>Operating with centralized business models.</td>
</tr>
<tr>
<td>III.</td>
<td>Geographical risk profile or systemic dimensions of the banking group</td>
<td>Foreign operations systemic in host country.</td>
</tr>
</tbody>
</table>

Table 1: Which dimensions make host supervisors likely to ring-fence?

<table>
<thead>
<tr>
<th></th>
<th>Location of financial stress</th>
<th>Financial stress in host country or foreign operations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Location of financial stress</td>
<td>Including safety and soundness concerns directly linked to the foreign subsidiary; political or sovereign vulnerabilities of host jurisdiction; doubts about quality of host supervision and regulation; uncertainties related to political stability, fiscal sustainability and local bank resolution process.</td>
</tr>
<tr>
<td>II.</td>
<td>Business model of the banking group</td>
<td>Banking group with significant asset overhang abroad.</td>
</tr>
<tr>
<td>III.</td>
<td>Geographical risk profile or systemic dimensions of the banking group</td>
<td>Foreign operations immaterial to banking group.</td>
</tr>
</tbody>
</table>

Table 2: Which dimensions make home supervisors likely to ring-fence?

³⁰ Herring (2007)
³¹ In this respect, it is important to note that one should consider not only the size of the host operations in terms of assets, but also the contribution of the subsidiary to the profitability of the banking group.
4. Risks addressed and ring-fencing instruments used

Ring-fencing measures vary widely in strictness and form, ranging from a narrow measure targeted at a particular risk category, such as liquidity buffers, to more blanket-type measures to address contagion risk in a more general way, such as forced subsidiarization. An overview of the risks and measures used to address them is provided below, in order of increasing reach.

- **Market and funding liquidity risk**: Higher liquidity buffers or the requirement to manage liquidity on a self-sufficient basis (i.e., without reliance on other parts of the banking group) and keep liquid funds in the jurisdiction can be used to make the foreign operations more self-sufficient and resilient to market and funding liquidity risks.³²

- **Market and funding liquidity risk, as well as safety and soundness risks**: Another common and simple geographical ring-fencing instrument is a restriction on upstream³³ liquidity flows, such as dividend restrictions to maintain higher than minimum capital buffers. These measures are intended both to mitigate funding and market liquidity risks, and to enhance the safety and soundness by building a more resilient banking system.

- **The risk of a shortfall in assets**: These risks are typically addressed by asset maintenance requirements or asset pledges that require banks to have — or to pledge — a minimum amount of safe assets tied to the host jurisdiction so that these assets can be easily monetized to compensate local depositors in the event of liquidation. Permanent asset maintenance requirements may be imposed, where a minimum proportion of domestic assets, generally expressed as a percentage of deposits, is to be maintained in the host country in the form of buildings, liquid assets or holdings at the central bank. Branches can also be subject to asset maintenance requirements or be required to hold an endowment account or a capital equivalent deposit, deposited in a separate account and overseen by the prudential supervisor. This amount can be expressed as a proportion of local deposits or assets.

- **Asset quality erosion**: For this concern, which is high on the supervisor’s agenda, a range of restrictions on intragroup cross-border asset outflows, or inflows of doubtful assets from other group members, can be imposed. These are intended to prevent the purchase of assets at inflated prices from group members or the sale of assets at lower than market prices. At the very least, intra-group asset inflows or outflows will be considered as a transaction with a connected party, which is subject to the application of the arm’s-length principle.

- **Contagion risk**: Stricter large exposure measures³⁴ or limits on intragroup funding can be used to reduce interconnectedness. Exposures to group companies can originate from a variety of sources: loans from the parent or group entities to the foreign subsidiary; intragroup shareholdings; central management of short-term liquidity within the group; and guarantees or commitments provided to, or received from, other entities in the group. This network of contractual obligations within a banking group can, from a host supervisor’s perspective, materially raise the likelihood of contagion to the local operations in case of distress in another group entity or in the group as a whole. Equally, intragroup exposures are likely to become avenues of intragroup support in times of stress.

- **Funding and solvency risk**: Imposing higher than minimum capital requirements is a very simple way to shield national banking systems from possible funding problems and capital adequacy concerns of the parent bank or the international banking group. It is also a way to encourage local lending and protect the domestic financial system from “deleveraging” when capital and liquidity may be transferred to other parts of the group.

- **Risks from conflicting interests**: Territorial approaches can be introduced in the form of firewalls or specific governance requirements to manage conflicting interests between the group and the subsidiary.³⁵ Company and banking laws diverge greatly in the extent to which parents can instruct their subsidiaries to engage in certain transactions, and hence make subsidiaries consent to transactions that are not necessarily in their best interest. In some countries, the notion of “group interest” is defined in company law and asset transfers are allowed against fair compensation. In a cross-border context, where it is easy to see how this notion could conflict with the interests of depositors, creditors and minority shareholders of the subsidiary. Although group and subsidiary interests are often not detachable, host

---

³² Market liquidity risk is the risk that the market liquidity worsens when the bank needs to unwind a position. Funding liquidity risk is the possibility that over a specific horizon a bank will become unable to settle obligations with immediacy.

³³ Upstream means from the subsidiary to the parent bank, downstream means from the parent bank to the subsidiary.

³⁴ A large exposure regime refers to capturing credit concentration risk, generally on a borrowers’ basis.

³⁵ European Commission (2010)
supervisors will act first in accordance with their core mandate, which remains the protection of local depositors in the host jurisdiction, regardless of the group interest.

Such conflicting interests can have important implications. In more general terms, all board members of a partly or fully foreign-owned subsidiary have a “duty of care” and a “duty of loyalty” toward the subsidiary bank itself, regardless of who appoints them. In other words, the board of the subsidiary needs to validate that its decisions are not detrimental to the sound and prudent management, the financial health, and the legal interests of the subsidiaries’ stakeholders. It will not always be straightforward to align this obligation with centralized cross-border banking structures and practices, certainly not when some of the members sitting on the subsidiary board also sit on the parent’s board. When financial difficulties arise, whether at the parent or at the subsidiary, a clear conflict of interest emerges. Moreover, for host supervisors, strong governance requirements can act as safeguards to enhance risk management in a local affiliate. For example, some jurisdictions have requirements for independent, nonexecutive and locally based board members and, to a lesser extent, for locally based senior management and internal auditors. This allows supervisors to better oversee the governance of foreign entities in their jurisdiction, and to make the board, senior management and the internal auditors accountable and liable to local authorities in case problems arise. Other supervisors go as far as requiring that most directors and key management of the bank operate independently from the group management and speak the local language. This would also allow them to sell the subsidiary more easily if problems arise.

Resolvability: Territorial approaches can also be introduced by restricting certain business models or requiring a specific legal form to ensure easier resolvability of the local operations. More and more host supervisors (e.g., Brazil, Mexico, India and New Zealand) now impose subsidiarization to protect local stakeholders from losses for SIBs or for banks where the home country applies depositor preference.

Intuitively, subsidiaries as independent legal entities under the control of the host supervisor may provide more protection for local depositors than branches. Forced subsidiarization by itself cannot, however, be seen as the silver bullet to “protect” a foreign subsidiary from contagion from the parent. Indeed, cross-border banks are generally organized along business lines, leading to operational structures that are very different from legal structures. In fact, some branches operate rather independently from their headquarters and some subsidiaries are very closely directed and managed by the parent bank. This trend has made it very hard for banks and banking supervisors alike to allocate activities to legal entities. The independent subsidiary view is thus increasingly undermined by the growing integration and centralization of key management functions, such as liquidity and funding, compliance, IT, auditing and internal controls. Common practices, such as issuing group-wide guarantees and cross-guarantees, raising of equity through entities in “cheaper” jurisdictions, and the onward direction of these funds to the jurisdictions where they generate a higher return, further blur legal separation from the parent bank.

On the other hand, subsidiary structures can allow more efficient resolution or wind up in a crisis provided that the business interconnections are well understood, there is a feasible and credible strategy to resolution and effective firewalls exist. Subsidiarization can probably also help with the calculation of burden sharing in case the crisis has to be resolved with public support. The setup of crisis management groups, and the drafting of cross-border cooperation and coordination agreements, as well as recovery and resolution plans are very useful tools in this respect. Nevertheless, the operations of international banks in many host countries are not always material to the banking group and host supervisors are thus not always actively involved in these exercises as they would like.

Some argue, therefore, that more radical approaches are needed, particularly in host countries dominated by foreign subsidiaries or branches of parent banks that are located in vulnerable sovereigns (e.g., countries with a dominant

---

36 The OECD defines “duty of care” as the duty of a board member to act on an informed and prudent basis in decisions with respect to the company. The OECD defines “duty of loyalty” as the duty of the board member to act in the interest of the company and shareholders.

37 Some home countries have general deposit preference which gives preference to all deposit liabilities irrespective of their deposit insurance eligibility, their covered status or the location where the deposits are booked or payable (whether in the home jurisdiction or at a foreign branch). This may lead to the home country depositors taking preference over the host country branch depositors in the case of resolution.
presence of Greek bank subsidiaries) or sovereigns with weak supervision. Requiring operations of foreign banks to be organized as stand-alone subsidiaries has been considered in those situations. Under this model, each entity would hold sufficient capital and liquidity to survive on its own, without support from the parent, reducing the connectedness and thus potential contagion risk between different components of the group.

The incidence of ring-fencing measures is not well described in the literature and no measurement of the extent of home bias can be found. Measuring home bias is very difficult, mainly because of the lack of transparency, as most decisions imposing capital, liquidity buffers or other restrictions on individual foreign banks remain confidential. While banking laws and regulations, which can also introduce territorial bias, are publicly available, comparing these laws and prudential regulations, thresholds and calculations across jurisdictions is a complicated exercise. Moreover, the existence of a supervisory power in the banking law or the prudential regulation does not automatically mean that it will actually be used in practice. D’Hulster has performed a survey of 22 prudential host supervisors of the Organisation for Economic Co-operation and Development (OECD), Euro and Eastern Europe and Central Asia (ECA) banking systems, and established a simple measure of territorial bias in prudential regulation and supervision.88 Appendix A provides some examples of national practices and the ring-fencing instruments used, the use of which seems to have accelerated recently in Europe as banks look to reduce their risks and regulators become more focused on firewalls.39 Appendix B looks into the implementation of the Banking Union and possible implications for ring-fencing measures introduced in the Eurozone.

5. Characteristics of ring-fencing decisions
Ring-fencing comes in many guises. We can distinguish several dimensions of ring-fencing decisions:

- The originator: home or host supervisor
- The scope: all banks in the jurisdiction, a group of banks with a similar risk profile or one bank
- The legal basis: legislation, prudential regulation or moral suasion
- The disclosure: public measure, confidential measure
- The timing: permanent ring-fencing or temporary ring-fencing (preventative, remedial and resolution stages)

5.1 The scope of the ring-fence
The scope of ring-fencing can be the whole banking population in a jurisdiction, a group of banks with a similar risk profile or a single bank. In the aftermath of the global financial crisis, some Central and Eastern European host supervisors of banking systems dominated by foreign banks have imposed dividend restrictions to ensure banks build stronger capital bases (see Appendix A). The scope of application of these dividend restrictions was based on a variety of indicators of the risk profile of a group of banks, such as Core tier 1 ratios, the supervisory risk rating score and the share of FX loans to unsecured borrowers.

5.2 The legal basis of the ring-fence
The legal basis for ring-fencing can be in legislation or prudential regulation or, in its absence, moral suasion can be used. Most banking laws or prudential regulations provide a legal basis for relatively straightforward ring-fencing decisions, such as higher capital ratios and liquidity buffers. In some instances, however, specific conditions, such as the requirement that capital adequacy must have fallen below a certain threshold, are included. In those cases, forward-looking supervisors may base ring-fencing decisions on their supervisory risk assessment and use moral suasion.

5.3 The disclosure of the ring-fence
Permanent territorial approaches incorporated in regulation and legislation are publicly available. A temporary ring-fencing decision on an individual bank that results from the supervisory risk assessments will generally not be disclosed. But temporary territorially biased decisions for a group of banks with similar risk profiles are usually made public as part of a Pillar 240 supervisory methodology or as an effort to build the resilience of the local banking system.

---

88 D’Hulster (2014)
39 Morgan Stanley (2012)
5.4 Timing and stages of application of ring-fencing: permanent versus temporary ring-fencing

Ring-fencing is often part of the overall policy framework in the sense that the requirements have been discussed before being implemented, they are permanent and they are publicly disclosed. Ring-fencing can also be temporary, in the form of an ad hoc decision of the prudential supervisor. During the global financial crisis, many host supervisors resorted to temporary ring-fencing decisions. When ring-fencing is temporary, the time for planning and preparation given to the bank as well as the health of the banking group or the overall economy become important factors in assessing its impact.

A closer look at stages of banking supervision will help to explain the complex interlinkages between temporary ring-fencing decisions and the timing of their implementation. Claessens et al. describe the three typical stages of banking supervision (preventative, remedial or recovery and resolution stages) and the instruments used during each stage. In reality, the first two stages are more of a continuum and banks can move back and forth between them. The first, or preventative, stage consists of licensing procedures; ongoing supervision, such as inspections and off-site analysis; and taking actions to address potential risks when required. In this stage, policy is developed and supervision is performed to address emerging risks. In the remedial or recovery stage, the bank has come under some financial stress, but is normally still viable. Supervisors take actions in response to the problems that have emerged. In the resolution stage, the bank, as a whole, has generally passed the point of viability and supervisory authorities act in concert with central banks, the Ministry of Finance, resolution authorities and bankruptcy courts, where applicable. The instruments range from private sector solutions, such as a takeover, restructuring and bankruptcy, to public sector solutions, such as public capital injections and nationalization.

Rosengren and Claessens et al. have found that divergences in incentives between home and host supervisors become magnified as problems get worse. As a consequence, incentives for ring-fencing will increase as uncertainty increases, the health of the bank deteriorates and problems in the home country or global financial stress worsen. In those instances, each jurisdiction tends to prioritize addressing its own problems instead of communicating with others. It is expected that the timing and the stage of introduction of ring-fencing actions matter for the impact on the banking group. Ring-fencing decisions made during the preventative banking supervision stage clearly tend to come with less risk for the financial health of the institutions than ring-fencing actions targeting an institution already under financial stress or operating in an already distressed financial environment. Similarly, ad hoc ring-fencing decisions, with little lead time for implementation, are also more likely to put the banking group under strain, as are decisions on a banking group that does not have robust liquidity contingency planning.

When looking at the third stage, there are two main strategies to the resolution of cross-border banks. First, there is the “single point of entry” resolution, in which resolution powers are applied to the top of a group by a single national resolution authority. Second, there is a “multiple point of entry,” in which resolution tools are applied to different parts of the group by two or more resolution authorities acting in a coordinated way. Neither of these approaches has actually been fully put in practice, but the reality has tended toward the second approach. This will almost automatically result in the application of ring-fencing approaches by host supervisors during a resolution since national authorities tend to pursue their own national interests in the management of a crisis.

6. Implications of ring-fencing on the banking group and financial stability

The potential impact of ring-fencing measures on the banking group and financial stability in the home and host country depends on a number of factors, including:

- Whether the home or the host supervisor is ring-fencing
- Whether the parent/home country or the affiliate/host country is experiencing financial stress
- The business model of the cross-border banking group
- The geographic risk profile of the banking group
- The level of development of the host country financial system
- Whether one host or multiple hosts are ring-fencing at the same time
6.2 Impact on the home country

In many respects, the interests of the home country are aligned with the interest of the banking group. Ring-fencing would be a blessing for a home supervisor with limited fiscal resources if the problem entity is not the parent bank. This will reduce the fiscal burden on the home country if the foreign operations become distressed, provided they are organized as a subsidiary. The home country supervisor will not be faced with the choice of either saving or resolving the banking group in its entirety. The arrangement can thus isolate a parent from a stand-alone subsidiary’s problem and potentially limit the latter’s problem from spreading to other parts of the banking group, thereby allowing for a selective resolution of the troubled subsidiary. A home supervisor may also welcome or even actively encourage, ring-fencing of host operations when faced with doubts about the adequacy of supervisory cooperation arrangements, the stability of the regime or the strength of the host country economy.

6.3 Impact on the host country

The host country financial system will be less vulnerable to contagion from other countries and will be better protected from parent bank problems at least in the short term. At the same time, credit availability for large international clients would be undermined and the banking system would become less diversified. For host countries with underdeveloped and shallow capital markets, and weak deposit bases, measures to ensure self-sufficiency could have significant implications for the availability and cost of lending. This may be reduced by direct cross-border lending, but this could complicate monetary policy decisions of the host country.

There is a risk that in the longer term, global banking groups, faced with higher costs imposed by ring-fencing requirements, may revise their business models, becoming more inward looking and perhaps ultimately leaving the host jurisdiction. Such home bias could result in retrenchment of international banks from host jurisdictions, including those where no territorial approaches have been adopted, creating more fragmented financial systems. Provision of credit to the private sector could thus be adversely affected in vulnerable regions or countries. This could contribute to existing deglobalization pressures that have emerged due to the need to shrink balance sheets, and shore up capital and liquidity for weak banks since the global financial crisis. Moreover, this trend could be exacerbated by some home regulators’ increased desire for banking systems that have more limited exposure across borders so as to preserve the available capital and funding for stressful times at home.
There is broad consensus that group-wide mobility of resources can help to dampen the impact of financial stability shocks and act as a stabilizer, a shock absorber or a source of systemic stability.

7. Conclusions
This paper has discussed why both home and host countries may find it beneficial to introduce restrictions and firewalls between parts of a cross-border banking group to protect their country from external shocks and minimize the fiscal costs of a failing bank. In so doing, they consider a number of factors, including: supervisory quality and cooperation, sovereign credit risk (home and host), uncertainty about the fiscal capacity and willingness of the home authority to support the affiliates in stress, how the resolution process will function in different jurisdictions, and the systemic importance of a foreign affiliate for the domestic banking system.

While ring-fencing may provide immediate benefit for host supervisors, it comes at a cost that is positively correlated with the severity of the ring-fencing measures. These costs have implications for the resilience of the banking group and the financial system as a whole; the availability and cost of capital and funding; the ability to manage liquidity and funding risks; the real sector consequences in host countries with less well-developed financial markets to provide alternative sources of funding; and a potential trend toward deglobalization.

The negative connotation surrounding ring-fencing comes mainly from its detrimental effects when it is abruptly implemented during a crisis. It is, therefore, important to distinguish between home bias as part of the policy framework (permanent ring-fencing) of the host country and sudden supervisory ring-fencing decisions (temporary ring-fencing) when problems start to emerge in a banking group or the system as a whole. While the former is common and does not necessarily put additional stress on an ailing banking group, the latter certainly would. Hence, well-calibrated and well-designed ring-fencing approaches may be better introduced when the banking system is healthy and has the ability to adapt and embed them in its centralized business practices. This gradual approach could well be considered a second-best option for many host supervisors, while the international policy work on resolution and burden sharing continues.

The ultimate solution to ensuring financial stability, and making nationalistic policy responses by home and host authorities redundant, lies in:

- The establishment of compatible and enforceable international mechanisms, which ensure effective oversight and orderly resolution of banks at both national and global levels
- The development of a combination of national and international arrangements to ensure that cross-border banking groups fully internalize the costs associated with their failure. These should address: (1) better risk management by banking groups; (2) effective supervisory oversight; (3) adequate information sharing and supervisory coordination mechanisms; and (4) workable cross-border resolution regimes and burden-sharing agreements.

In this context, addressing information asymmetries, within and across borders, would go a long way toward achieving these key elements and avoiding mutually harmful nationalistic policy responses. Clear rules that allocate responsibilities to home and host jurisdictions, and that address the lack of comprehensive information about how the resolution process functions in relevant jurisdictions, would allow stakeholders to plan more efficiently for potential consequences of insolvency. Globally enforceable information sharing and resolution mechanisms could reduce the need for ring-fencing.

In practice, however, there always remains the risk that during a crisis, authorities make decisions driven by political expediencies and deviate from agreed mechanisms and frameworks.\(^{(44)}\)

\(^{(44)}\) For example, Iceland did not pay for UK and Dutch deposits, while under European Economic Area (EEA) arrangements it had to support foreign deposits of the Icelandic branch. Also, institutions of relatively small magnitude have time and again been classified by politicians as too big to fail, e.g., Hypo Alpe Adria and German State banks.
Appendix A

<table>
<thead>
<tr>
<th>Country</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>A “systemic” threshold above which regulation of subsidiaries applies to branches that become systemic. The banking supervisor also has the power to require a foreign branch to become a subsidiary owing to “systemic risk.”</td>
</tr>
<tr>
<td>Australia</td>
<td>Asset maintenance requirements: a bank’s assets in Australia must be equal to or greater than the amount of its deposit liabilities in Australia.</td>
</tr>
<tr>
<td>Austria</td>
<td>In late 2011, Austrian rules limited new lending in Central and Eastern Europe (CEE) above 110% loan-to-deposit ratio, introducing limits on parent-funded credit expansion in CEE subsidiaries.</td>
</tr>
<tr>
<td>Brazil</td>
<td>The Brazilian banking supervisor has specific regulation banning the application of funding mobilized domestically (deposits/securities/equity issuances) to fund the acquisition of assets abroad, including lending to headquarters or buying portfolio/assets.</td>
</tr>
<tr>
<td>Croatia</td>
<td>During the 2007-08 crisis, bank regulators recommended the non-distribution of profits by the subsidiaries of foreign banks despite relatively strong bank fundamentals. The Croatian National Bank (CNB) Governor said that “the CNB would not look favorably upon attempts to withdraw capital, deposits, or pay out total accumulated profits because that would destabilize the domestic banking system. In such a case, the CNB would be forced to undertake protective measures, regardless of thus connected risks.”</td>
</tr>
<tr>
<td>India</td>
<td>In a statement issued on 6 November 2013, the Reserve Bank of India required foreign banks to establish a wholly owned subsidiary if the following conditions are satisfied: A: (1) Banks incorporated in a jurisdiction having a legislation giving a preferential claim to deposits of home country in a winding-up procedure; (2) banks that do not have adequate disclosure requirements in their home country; (3) banks with complex structures; (4) banks which are not widely held; (5) the Reserve Bank of India is not satisfied with the adequacy of supervisory arrangements (including disclosure arrangements) and market discipline in the country of incorporation; (6) for any other reason that the Reserve Bank of India considers necessary for subsidiary for or presence of the foreign bank; Or B: if a foreign bank, which has set up its presence in India through a branch mode, is considered by the Reserve Bank of India as being systemically important by virtue of its size of business.</td>
</tr>
<tr>
<td>New Zealand</td>
<td>In January 2013, the Reserve Bank of New Zealand issued a “Statement of Principles” on bank registration and supervision. The principles detail that foreign bank operations will have to take the form of a locally incorporated subsidiary if a foreign bank fails, or is expected to fall within the next five years, in one of the following four categories: (1) SIBs, that is, banks whose New Zealand liabilities, net of amounts due to related parties, exceed NZS15 billion. (2) Retail deposit takers incorporated in a jurisdiction that has legislation, which gives deposits made, or credit conferred, in that jurisdiction a preferential claim in a winding up. Australia and the US are examples of countries with such legislation. (3) Retail deposit takers which do not provide adequate disclosure in the home jurisdiction. (4) Applicants other than those listed above may also be required to incorporate locally if the Reserve Bank is not satisfied that supervisory arrangements (including disclosure arrangements) and market disciplines in the country of incorporation are adequate.</td>
</tr>
<tr>
<td>Poland</td>
<td>The Polish prudential supervisor, the Polish Financial Supervision Authority or Polish KNF, has imposed risk-based dividend restrictions for the earnings of 2011 and 2012. In a letter sent to all banks, the Chairman of the KNF expressed an expectation that all banks take action aimed at strengthening their capital base due to heightened risks in global financial markets. He stated that should be done by retaining earnings from 2011 if the capital adequacy ratio is below 12% or the tier 1 ratio below 9%. Capital ratios were not the only criterion influencing the ability to distribute earnings for 2011. Other areas taken into consideration included: the score a bank received in the Supervisory Review and Evaluation (SREP) process, the share of FX loans to unsecured borrowers in the bank’s loan portfolio and the financial situation of the parent company. The level of dividend payouts was not allowed to exceed 50% of total 2011 earnings.</td>
</tr>
<tr>
<td>Poland and Hungary</td>
<td>In November 2012, the Polish KNF issued another statement on dividend policy. The level of earnings retention was relaxed and banks were allowed to pay up to 75% of the 2012 earnings. The criteria were, however, tightened; the prescribed levels of capital ratios of 12% and 9% tier 1 capital should be maintained by banks under the adverse scenarios of the Polish KNF’s bottom-up stress test exercise. The SREP score criteria referring to the capital level and internal were also maintained, while the criteria for the desirable level of FX lending in the credit portfolio were removed.</td>
</tr>
<tr>
<td>Slovakia</td>
<td>More recently, Polish and Hungarian regulators have started to closely monitor intragroup liquidity flows and localizing business models with the intention that with negative ratings migration tighter parent liquidity should be offset by local ring-fenced funding and high capital buffers.</td>
</tr>
<tr>
<td>Slovakia</td>
<td>In 2012, the National Bank of Slovakia recommended that banks restrict the distribution of profits that have not yet been included in equity capital, in accordance with the table below.</td>
</tr>
<tr>
<td></td>
<td>Value of Core tier 1 capital when the decision of profit distribution was taken</td>
</tr>
<tr>
<td></td>
<td>1. Less than 9.625%</td>
</tr>
<tr>
<td></td>
<td>2. At least 9.625%, but less than 10.25%</td>
</tr>
<tr>
<td></td>
<td>3. At least 10.25%, but less than 10.875%</td>
</tr>
<tr>
<td></td>
<td>4. At least 10.875%, but less than 11.5%</td>
</tr>
<tr>
<td></td>
<td>5. More than 11.5%</td>
</tr>
</tbody>
</table>
The National Bank of Slovakia also warns those banks that need to strengthen their capital adequacy ratio. It states that they need to do so by means of current earnings or share capital increases. Banks should avoid tightening lending conditions as a means of bringing their capital ratios into line with the regulatory requirements or with the level recommended in the decision.

**Switzerland**

The Swiss authorities have introduced regulation that relates to the global operations of SIBs and foreign banks operating in Switzerland. A set of criteria was presented using a scorecard approach called the “Resolution Effectiveness Test” to determine whether the banks are organized in such a way that resolution in a crisis is facilitated while protecting the systemically critical functions of these banks. Based on the outcomes of the assessment, the authorities can propose territorial measures that require changes in these entities so that intragroup exposures are limited, and domestic and foreign operations are legally separated.

**Turkey**

Bank regulators recommended the non-distribution of profits by the subsidiaries of foreign banks despite relatively strong bank fundamentals. The head of the banking regulation agency stated in December 2009 that, “it is our natural right to expect those profits generated in this country to be invested and used in credit extension again in this country.” Banks in Turkey were expected to consult the regulator before distributing dividends during the crisis.

**UK**

The UK FSA imposed a liquidity regime that requires stand-alone liquidity requirements on both subsidiaries and branches of foreign banks.

**US**

- In February 2014, the Federal Reserve approved a final rule strengthening supervision and regulation of large US bank holding companies and foreign banking organizations. The final rule establishes a number of enhanced prudential standards for large US bank holding companies and foreign banking organizations to help increase the resiliency of their operations. These standards include liquidity, risk management, and capital. It also requires a foreign banking organization with a significant US presence to establish an intermediate holding company over its US subsidiaries, which will facilitate consistent supervision and regulation of the US operations of the foreign bank. The final rule was required by section 165 of the Dodd-Frank Wall Street Reform and Consumer Protection Act. For US bank holding companies with total consolidated assets of US$50 billion or more, the final rule incorporates the previously issued capital planning and stress testing requirements as an enhanced prudential standard. It also requires such a US bank holding company to comply with enhanced risk management and liquidity risk management standards, conduct liquidity stress tests, and hold a buffer of highly liquid assets based on projected funding needs during a 30-day stress event. These requirements will help ensure that these firms can continue to lend to households and businesses even in times of financial stress. In addition, the final rule requires publicly traded US bank holding companies with total consolidated assets of $10 billion or more to establish enterprise-wide risk committees. The new requirements for US bank holding companies complement the stress testing and resolution planning requirements for large bank holding companies that the board previously finalized.

- Banking regulation in New York State notes that in the case of insolvency of the parent bank, creditors with residence in New York State have preference over other creditors with respect to the assets of the foreign branch.

**Latin American countries**

- Some hosts (e.g., Brazil, Mexico) strongly encourage subsidization of local business units.

- In Argentina, Bolivia, Brazil, Chile and Ecuador, branches face local capital and liquidity charges identical to those applied to subsidiaries and require local representation on their boards.

- Spanish subsidiaries in some Latin American countries have many elements of SAS, generally relying on customer deposits to finance loans, but excess liquidity is managed at the group level.

**Croatia, Mexico, Australia**

Restrictions exist on branches of foreign banks accepting deposits.

<table>
<thead>
<tr>
<th><strong>Table 3: Territorial approaches in selected countries</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source:</strong> Authors' compilation of regulations and supervisory decisions.</td>
</tr>
</tbody>
</table>

---

132 The Journal of Financial Perspectives
Appendix B

Ring-fencing in the EU and the Banking Union

There is a general consensus that the Banking Union consists of four interconnected building blocks: common prudential regulation, common banking supervision materialized by the SSM, common deposit insurance, and common resolution to be implemented by the SRM, including the Single Resolution Fund (SRF). The SSM was fully accomplished by 4 November 2014 when it had taken over the direct supervision of all “significant” credit institutions in the Eurozone, 120 banking groups, including approximately 1,200 supervised legal entities. This represents about 85% of the banking assets in the Eurozone. The SRM became “fully operational” on 1 January 2016 with the SRF being “built up over a period of eight years with 'ex-ante' contributions from the banking industry.”

The SSM is likely to lead to the elimination of many of the ring-fencing measures that were introduced during the global financial crisis in Eurozone countries, particularly the liquidity buffers and the dividend restrictions. Common prudential regulation has been achieved by the implementation of the fourth Capital Requirements Directive (CRDIV) and the Capital Requirements Regulation (CRR). The latter is a directly applicable minimum harmonization regulation; this means that the range of regulatory ring-fencing instruments, such as capital buffers, has become more limited and more difficult to implement for all EU Member States.

In 2014, the EU Commission performed a confidential survey on geographical ring-fencing measures in 27 member states. The survey covered Pillar 2, large exposures and domestic liquidity frameworks. The Commission takes the view that duly justified and proportionate ring-fencing measures are not restrictions on free movement of capital. It is only when these instruments are used for a purpose not in line with EU legislation that they may be considered. The Commission recognizes that only the further integration of the regulatory and supervisory framework can eliminate ring-fencing measures. In other words, the cause of ring-fencing measures, not the symptoms, should be eliminated.

45 European Commission (2015)
Ring-fencing cross-border banks: an effective supervisory response?

References


The broken buck stops here: embracing sponsor support in money market fund reform

Jill E. Fisch
Perry Golkin Professor of Law, University of Pennsylvania Law School

1 This article is a shorter and updated version of an article originally published in the North Carolina Law Review (93). It is published with the permission of the Law Review. I am grateful to Jeff Gordon, Laura Merianos, Heidi Stam, David Skeel, Jennifer Taub and participants at Georgetown’s Financial Regulatory Camp for helpful comments on earlier drafts.
Abstract
Since the 2008 financial crisis, money market funds (MMFs) have been the subject of ongoing policy debate and targeted for regulatory reform. The US Securities and Exchange Commission (SEC) eventually adopted reforms that have the potential to increase MMF fragility, while, at the same time, reducing MMF utility for many investors.

The shape of the SEC's reforms was substantially influenced by bank regulators who framed the narrative of MMF fragility in terms of nefarious “shadow banking.” This rhetoric failed to acknowledge a critical difference between the asset segregation in MMFs and in banks. Unlike banks, MMF sponsors have assets and operations that are separate from the assets of the MMF itself. This difference caused the SEC to mistake sponsor support as a weakness and to adopt reforms that burden sponsor support instead of encouraging it.

As this article explains, required sponsor support offers a novel and simple regulatory solution to MMF fragility. Accordingly, this article proposes that the SEC requires MMF sponsors explicitly to guarantee the US$1 share price. Taking sponsor support out of the shadows, embraces, rather than ignores, the advantage that MMFs offer over banks through asset partitioning. At the same time, sponsor support harnesses market discipline as a constraint against MMF risk taking and moral hazard.
1. Introduction

On 16 September 2008, the Reserve Fund “broke the buck,” reducing its net asset value (NAV) to 97 cents per share after the announcement of the Lehman Brothers bankruptcy filing. The filing caused the board to write down the value of the fund’s US$785m in Lehman Brothers debt to zero.

By September 16, 2008, the financial markets had already experienced substantial turbulence. Contributing to this turbulence were the bailout of Bear Stearns, the federal government’s decision to put Fannie Mae and Freddie Mac into conservatorship, the Federal Reserve Board’s announcement of its decision to support AIG financially and the Lehman bankruptcy. Many money market fund (MMF) sponsors provided support to their MMFs to maintain the stable US$1 share price by taking actions such as buying debt holdings that had declined in market value. Nonetheless, over the next several days, investors redeemed substantial amounts of money from MMFs. According to the SEC: “[During the week of September 15, 2008 [Lehman week], investors withdrew approximately US$300b from prime [MMFs]].”

Widespread redemptions put more MMFs at risk of breaking the buck, but they also had a broader effect on the economy. Fund manager efforts to conserve cash reduced the availability of short-term credit. Tightening credit conditions caused businesses to reduce capital expenditures and lay off workers. These effects led policymakers to view the fragility of MMFs as a substantial cause of the financial crisis and to target them for regulatory reform. The central goal of reform proposals was to prevent MMFs from breaking the buck in the future because of the concern that doing so would generate “runs,” which would, in turn, generate a contagion effect across MMFs and destabilize the economy. As Treasury Secretary Timothy Geithner explained: “[T]he financial crisis of 2007–08 demonstrated that MMFs are susceptible to runs and can be a source of financial instability with serious implications for broader financial markets and the economy.”

After a contentious six-year debate among policymakers about the appropriate shape of MMF reform, on July 23, 2014, the SEC adopted a rule requiring a floating NAV for institutional prime MMFs and implementing a variety of additional regulatory requirements for other MMFs, including a complex structure of gates and fees for retail MMFs. As this article will explain, neither aspect of the reform responds in a meaningful manner to the problem of large redemptions in a time of financial distress. The structure of the reform was heavily influenced by political pressure, pressure that led the SEC to take a flawed approach, and to overlook a simple and superior regulatory solution.

The rhetoric that describes MMFs as shadow banks and analogizes MMF redemptions to bank runs has obscured a critical structural difference between MMFs and bank accounts: the separation of MMFs from the securities in the MMF’s portfolio. As this article will explain, “runs,” which would, in turn, generate a contagion effect across MMFs and destabilize the economy. As Treasury Secretary Timothy Geithner explained: “[T]he financial crisis of 2007–08 demonstrated that MMFs are susceptible to runs and can be a source of financial instability with serious implications for broader financial markets and the economy.”

After a contentious six-year debate among policymakers about the appropriate shape of MMF reform, on July 23, 2014, the SEC adopted a rule requiring a floating NAV for institutional prime MMFs and implementing a variety of additional regulatory requirements for other MMFs, including a complex structure of gates and fees for retail MMFs. As this article will explain, neither aspect of the reform responds in a meaningful manner to the problem of large redemptions in a time of financial distress. The structure of the reform was heavily influenced by political pressure, pressure that led the SEC to take a flawed approach, and to overlook a simple and superior regulatory solution.

The rhetoric that describes MMFs as shadow banks and analogizes MMF redemptions to bank runs has obscured a critical structural difference between MMFs and bank accounts: the separation of MMFs from the securities in the MMF’s portfolio. As this article will explain, “runs,” which would, in turn, generate a contagion effect across MMFs and destabilize the economy. As Treasury Secretary Timothy Geithner explained: “[T]he financial crisis of 2007–08 demonstrated that MMFs are susceptible to runs and can be a source of financial instability with serious implications for broader financial markets and the economy.”

After a contentious six-year debate among policymakers about the appropriate shape of MMF reform, on July 23, 2014, the SEC adopted a rule requiring a floating NAV for institutional prime MMFs and implementing a variety of additional regulatory requirements for other MMFs, including a complex structure of gates and fees for retail MMFs. As this article will explain, neither aspect of the reform responds in a meaningful manner to the problem of large redemptions in a time of financial distress. The structure of the reform was heavily influenced by political pressure, pressure that led the SEC to take a flawed approach, and to overlook a simple and superior regulatory solution.

The rhetoric that describes MMFs as shadow banks and analogizes MMF redemptions to bank runs has obscured a critical structural difference between MMFs and bank accounts: the separation of MMFs from the securities in the MMF’s portfolio. As this article will explain, “runs,” which would, in turn, generate a contagion effect across MMFs and destabilize the economy. As Treasury Secretary Timothy Geithner explained: “[T]he financial crisis of 2007–08 demonstrated that MMFs are susceptible to runs and can be a source of financial instability with serious implications for broader financial markets and the economy.”

After a contentious six-year debate among policymakers about the appropriate shape of MMF reform, on July 23, 2014, the SEC adopted a rule requiring a floating NAV for institutional prime MMFs and implementing a variety of additional regulatory requirements for other MMFs, including a complex structure of gates and fees for retail MMFs. As this article will explain, neither aspect of the reform responds in a meaningful manner to the problem of large redemptions in a time of financial distress. The structure of the reform was heavily influenced by political pressure, pressure that led the SEC to take a flawed approach, and to overlook a simple and superior regulatory solution.
This article argues for a dramatic shift in the regulation of MMFs. Rather than trying to make them operate more like banks by requiring capital buffers or sacrificing their viability by mandating a floating NAV, reform should formalize the role of sponsor support. Specifically, this article proposes that regulators require MMF sponsors to stand behind their MMFs by committing to maintain the stable US$1 NAV. In a time of crisis, sponsors could provide such support by buying distressed assets from the fund, reducing management fees or subsidizing the fund with other business revenues. Sponsors could also privately insure their obligation.

Mandatory sponsor support offers several advantages over the 2014 rule and other reform proposals. It would both prevent MMFs from breaking the buck and harness market discipline to provide fund sponsors with appropriate incentives to limit risk-taking. Required sponsor support would eliminate market uncertainty about the extent to which a sponsor would voluntarily support its fund in a time of crisis — uncertainty that contributed to the turmoil surrounding the events at the Reserve Primary Fund. Sponsor support would substitute sponsor financial stability for the need for investors to monitor the quality of MMF assets directly, a task that the SEC has highlighted with its new and unworkable disclosure requirements. Most importantly, sponsor support would address MMF fragility, while allowing MMFs to continue to meet investor demand for a liquid, stable-value, cash-management option.

The article continues in section 2 by briefly describing the background to MMF reform. Section 3 evaluates the SEC’s 2014 rule. In section 4, the article introduces its proposed alternative: mandatory sponsor support for a US$1 NAV.

2. Background

2.1 MMFs and the financial crisis

MMFs are a type of mutual fund regulated by the SEC under the Investment Company Act of 1940. MMFs offer investors access to a pool of short-term debt securities. The critical feature that distinguishes MMFs from other short-term investment funds is that, while the price of most mutual funds fluctuates on a daily basis in accordance with the funds’ NAV, MMF shares are bought and sold at a stable US$1 share price. The price that MMFs pay for authorization to trade at US$1 per share is in compliance with SEC Rule 2a-7, which imposes a variety of constraints on the safety and liquidity of the assets in which MMFs are permitted to invest. See Jill E. Fisch, Rethinking the Regulation of Securities Intermediaries, 158 U. PA. L. REV. 1961, 1975 (2010).

Facilitates the role of MMFs as a cash-management tool because the frequent investment and redemption of MMF shares does not result in a gain or loss for tax or accounting purposes. As a result, MMFs offer investors both liquidity and stability. MMFs have traditionally offered investors higher returns and greater diversification than traditional bank accounts, as well as features like check writing and debit card access, although, unlike bank deposits, they are not protected by government insurance.

The investments of MMFs, unlike most mutual funds, are conservative securities that are typically held to maturity. MMFs rarely sell their portfolio holdings unless compelled to do so by redemption requests. As a result, while the value of the MMF’s portfolio may fluctuate on a daily basis, the value of any particular asset will approach its face amount as the instrument nears maturity. Since 1982, SEC rules have reflected this fact by allowing MMFs to value their portfolio assets using amortized cost accounting rather than market price. If, however, an MMF’s share price drops below US$0.995, it is required to price its shares at 99 cents (or less). This is described as “breaking the buck.” Only two MMFs have broken the buck: the Reserve Primary Fund in 2008 and the Community Bankers US Government Fund, a small institutional fund, in 1994.

MMFs rarely break the buck because MMF sponsors have traditionally been willing to support the US$1 share price. Sponsor support may take the form of capital support agreements, letters of credit, waiving management fees or purchasing distressed assets from the MMF at amortized cost.

On the supply side, MMFs provide a major part of the market for short-term debt securities, including commercial paper. MMFs are also an important source of financing for state and local governments, and government entities.

In the early morning of September 15, 2008, Lehman Brothers filed for bankruptcy. The Reserve Primary Fund held approximately 1.2% of its portfolio, US$785m, in short-term Lehman debt. On September 15, 2008, the Reserve Fund received redemption requests of US$25b, reflecting more than 40% of the fund’s value. The next day the fund broke the buck...
and, following the announcement, suspended redemptions.

Other MMFs also received substantial redemption requests during Lehman week, leading fund managers to increase their cash holdings and, in turn, reducing the supply of short-term credit. In addition, the prospect that additional financial firms might fail increased the pressure on MMFs. As the then Executive Vice President of the Markets Group at the New York office of the Federal Reserve, William Dudley, observed in the meeting of the Federal Reserve Board on September 16, 2008: “The risk here, of course, is that, if AIG were to fail, money funds have even a broader exposure to them than to Lehman and so, breaking the buck on the money market funds is a real risk.”

The US Government responded to concerns about MMFs by establishing two programs. The Treasury Department Temporary Guarantee Program for MMFs was designed to reduce redemption requests by assuring investors that they would not lose the value of money invested in MMFs. The program provided for a federal government guarantee – federal insurance – of MMF assets, which the government provided in exchange for a premium paid by fund sponsors. The Asset-Backed Commercial Paper Money Market Mutual Fund Liquidity Facility provided a market for asset-backed commercial paper. In so doing, it provided a source of liquidity for MMFs that were subject to redemption requests. In addition, it supported the price of commercial paper. Reducing the need for distressed sales would depress the value of short-term credit instruments. The two programs were, together, quite effective in reducing the volume of outflows from MMFs and supporting the price of commercial paper.

In January 2010, the SEC initially responded to the Reserve Primary Fund situation and the broader turmoil in the MMF industry by amending Rule 2a-7 to add requirements that MMF portfolios have higher investment quality, shorter maturities, and greater liquidity. MMFs were also required by the rule to engage in periodic stress testing so as to determine how likely the fund was to break the buck in the event of adverse economic developments or heavy redemption requests, and were subjected to increased disclosure requirements.

When it adopted the 2010 rule changes, the SEC noted that more fundamental changes to MMFs might be needed. Chair Mary Schapiro identified several possible regulatory alternatives, including, most prominently, a requirement that MMFs shift to a floating NAV rather than trading at a stable US$1 share price. At the same time, bank regulators were advocating additional reforms to MMF regulation. The Treasury Department identified specific reforms that it believed the SEC should adopt. The President’s Working Group on Financial Markets (PWG) published a report indicating that, despite the SEC’s rule changes, MMFs remained vulnerable to runs and highlighting the destabilizing effect of runs on the economy.

The SEC then began to develop an additional rulemaking proposal, but that proposal did not command the support of a majority of the commissioners.

When Schapiro announced that the SEC would not move forward with additional reforms, the Treasury Secretary Timothy Geithner wrote a letter urging the Financial Stability Oversight Council (FSOC), which he chaired, to become involved. At his prompting, the FSOC took the unprecedented step of releasing its own MMF reform recommendations for public comment.

12 The members of the PWG included the Secretary of the Treasury Department (as Chairman of the PWG), the Chairman of the Board of Governors of the Federal Reserve System, the Chairman of the SEC and the Chairman of the CFTC, id. at 1 n.1.
14 Id. at 4–6.
16 Letter from Timothy F. Geithner, Sec’y of the Treasury, to Members of the Fin. Stability Oversight Council (Sept. 27, 2012), available at http://link.coremotivesmarketing.com/c/306/c9c08c9c7eb64ed07c9d05c3c47da44efb1d1713c48f31b665172e5b0c3d9fbeb6.
The broken buck stops here: embracing sponsor support in money market fund reform

The FSOC described its proposal as pursuant to its authority under section 120 of Dodd-Frank18 to “provide for more stringent regulation of such financial activity or practice by issuing recommendations to a primary financial regulatory agency to apply new or heightened standards or safeguards.”19 Significantly, the FSOC’s involvement created political pressure for the SEC to implement additional reforms.20

2.2 The case for MMF vulnerability

The case for MMF reform was based on the argument that the structural characteristics of MMFs made them vulnerable to runs, which have a destabilizing effect on the economy.21 Reform is, therefore, deemed necessary to reduce or eliminate the risk of future runs.

The structure of this argument is questionable because the term “run” is typically associated with banks, not mutual funds.22 Banks hold money that depositors are entitled to withdraw on demand. Banks, in turn, lend that depositor money to borrowers. If too many depositors demand their money at the same time, the bank cannot repay them all. By contrast, MMFs hold high-quality short-term assets that typically can be liquidated at, or near, par value. This is known as the first-mover advantage.

It is somewhat misleading to characterize heavy investor withdrawals from MMFs as a “run.” It is common for the financial markets to experience large and rapid movement of assets without the claim that such movements warrant regulatory intervention.25 Rather, it seems that two attributes distinguish a run from normal movement of assets: panic trading and a first-mover advantage. As Schapiro put it, the redemptions from the Reserve Fund and other MMFs during Lehman week were made by “panicked investors.”26 Panic trading implies a degree of irrationality – redemptions that are motivated by fear rather than genuine financial weakness. Moreover, panic selling is not enough. Runs are also characterized by a first-mover advantage, meaning that investor behavior itself causes a shortage or diminution in value of the remainder. This leads investors, who would not otherwise have traded to take action in order to avoid losing out entirely.27 As a result, a run can induce scarcity by creating an abnormal level of investor demand.28

The mechanics of a run operate differently for an MMF than for a bank. A bank holds long-term illiquid assets that cannot readily be converted to cash. As a result, if many of a bank’s depositors demand their money at the same time, the bank cannot repay them all. By contrast, MMFs hold high-quality short-term assets that typically can be liquidated at, or near, par value.

As a result, under normal market conditions, heavy redemption requests do not create a first-mover problem for an MMF because the MMF can satisfy those requests by liquidating assets.29

The situation in 2008 was distinctive for three reasons. First, the bankruptcy of Lehman generated substantial losses in the value of MMF assets. A loss of this size in money market assets was highly unusual.30 Second, the economic climate during the fall of 2008 put many MMF investors under economic pressure and, in particular, liquidity pressure because of the freeze in the short-term credit markets.31 This led MMF investors to withdraw

18 Id. at 69.455; Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. 111-203, § 1200a, (c)(2), 124 Stat. 1376, 1408-09 (2010).
21 For example, Douglas W. Diamond and Philip H. Dybvig, Banking Theory, Deposit Insurance, and Bank Regulation, 59 J. Bus. 55, 63-64 (1986).
22 Scholars generally agree that a key component of banks’ susceptibility to runs is the maturity mismatch: the bank’s assets are tied up in long-term investments, but it offers immediate
23 Id. at 69.455; Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. 111-203, § 1200a, (c)(2), 124 Stat. 1376, 1408-09 (2010).
24 Significantly, the FSOC’s involvement created political pressure for the SEC to implement additional reforms.20
26 Schapiro Statement
28 Diamond and Dybvig, at 63.
31 The loss would have been greater if, as was possible, other financial institutions had been
funds to meet their cash flow needs. Third, non-Lehman events, including the bailout of Bear Stearns and the trouble at a number of other financial institutions, including AIG, Wachovia and Citigroup, created widespread concern about the quality (and possible default) of money market debt from other issuers. As a result, the MMFs that experienced a high volume of redemptions could not readily find buyers for their assets.

Both attributes of a run — panic trading and the first-mover advantage — are relevant to the causal relationship between breaking the buck and a run on MMFs. First, the act of breaking the buck may increase the salience to the market of the fact that MMFs do not guarantee the US$1 share price. Once the risk of losing money becomes salient, investors may panic and withdraw their funds from financially stable MMFs. These withdrawals can deplete MMF's liquidity so that those who run slowly may be unable to withdraw their money.

In the case of an MMF, a stable US$1 NAV aggravates the situation. So long as the MMF’s NAV is sufficiently high, investors can redeem at the US$1 share price.

But, as the MMF’s NAV falls below US$1 per share, investors can continue to redeem at US$1, even if their share of the fund’s assets is somewhat less, as long as the NAV is above US$0.995. These redemptions deplete the fund’s assets because redeeming investors are receiving more than their entitlement — the difference between the fund’s actual NAV and US$1 — and leaving even less for subsequent investors. These redemptions reflect the “arbitrage opportunity” created by the US$1 share price. Importantly, absent heavy redemption pressure, these small deviations between US$1 and the fund’s actual NAV do not deplete fund assets because the effect is offset by simultaneous purchases that also take place at US$1 per share. As a result, the gap has a meaningful effect on fund value only in situations in which redemption demand significantly outpaces purchases.

The Reserve case exemplified this effect; early redeemers got out at US$1 per share and other investors received only 99 cents.

The third problem with MMFs is contagion effect. According to AIG, Wachovia and Citigroup, created widespread concern about the quality and possible default of money market debt from other issuers. As a result, the MMFs that experienced a high volume of redemptions could not readily find buyers for their assets.


The shadow NAV reflects the value of the MMF’s portfolio using a market-based valuation rather than amortized cost. SEC Investment Company Act of 1940, 17 C.F.R. § 270.2a-7(c) (Revised 2014).


For example, Gordon and Gandia at 328 n.35.


For example, Gordon and Gandia at 328 n.35.

For example, Gordon and Gandia at 328 n.35.

For example, Gordon and Gandia at 328 n.35.

For example, Gordon and Gandia at 328 n.35.

For example, Gordon and Gandia at 328 n.35.

For example, Gordon and Gandia at 328 n.35.

For example, Gordon and Gandia at 328 n.35.

For example, Gordon and Gandia at 328 n.35.
The broken buck stops here: embracing sponsor support in money market fund reform

3. The 2014 reform

3.1 The SEC’s 2014 MMF rule

On July 23, 2014, the SEC approved a final rule reforming MMF regulation by a divided 3-2 vote.40 Chair Mary Jo White explained that the reforms would “significantly mitigate the risks of a run in money markets funds and limit further contagion should a run occur.”41 The other commissioners were less sanguine.42 Commissioner Aguilar, who voted in favor of the reform, described the rulemaking process as “one of [the] most flawed and controversial” ever undertaken by the SEC.43

The new rule requires prime institutional MMFs to implement a floating NAV, but it exempts retail and government funds from this requirement.44 The rule authorizes boards of retail funds to implement gates and fees to discourage redemptions, and provides that the power to use these tools is triggered by declines in fund liquidity.45 Finally, the rule adopts a new and narrow definition of government MMFs and exempts such MMFs from both the floating NAV, and gates and fees provisions.46

In addition to these structural changes,48 the new rule includes important new disclosure requirements. MMFs are required to provide extensive additional information on their websites.49 These requirements are supplemented by additional disclosures in the MMF prospectus and marketing materials, in Form N-CR, and in the statement of additional information (SAI).50 The requirements include disclosure of the fund’s current and historical market-based NAV, calculated on a daily basis and rounded to four decimal points – the nearest ten thousandth of a cent.51 Funds are required to disclose any past use of gates and fees and historical sponsor support.52

Funds must also disclose current and historical information about the percentage of daily and weekly liquid assets in their portfolios, as well as current and historical information about net shareholder inflows and outflows.53

Because of the potentially substantial effects of the new rules, the SEC provided that compliance dates for both the gates and fees provision, and the floating NAV would not occur until two years after the effective date of the rule.54 According to Chair White, the new rule “will fundamentally change the way that most MMFs operate.”55

3.2 Evaluating the reforms

The 2014 rule became effective in November 2016. To date, its adoption of a combination of liquidity gates and fees, and a floating NAV – two regulatory approaches that had been the subject of many prior reform proposals – has not been tested by a liquidity crisis. Initial market reactions to the rule have been significant, however. Specifically, the rule has led to a substantial outflow of funds from Prime MMFs into government MMFs and banks.56

---

44 Retail funds are defined as funds that have “policies and procedures reasonably designed to limit all beneficial owners of the fund to natural persons.” 2014 Final Rule at 47,794. This was a change from the proposed rule, which would have defined retail funds as those that did not allow shareholders to redeem more than US$1m in a single business day; 2013 Rule Proposal at 36,856.
45 In the final rule, the fee and gate provisions are explicitly discretionary. A fund may impose such provisions, however, only when a fund’s weekly liquid assets drop below 30% of its total assets. See 2014 Final Rule at 47,747. The 2010 amendments already required MMFs to maintain 30% of their portfolios in weekly liquid assets. 17 C.F.R. § 270.2a-7(c)(5)(iii) (2012); see 2014 Final Rule at 47,738; 2010 Final Rule at 10,113–14.
46 2014 Final Rule at 47,794 (“We therefore are redefining the definition of a government fund to require that such a fund invest at least 99.5% (up from 80% in the proposal) of its assets in cash, government securities, and/or repurchase agreements that are collateralized by cash or government securities.”).
47 Id. at 47,791-94. The exemption for government MMFs preserves such funds as a cash-management option for institutional investors. At the same time, it expands the number of funds that remain potentially vulnerable. As the SEC noted in the adopting release, government MMFs experienced substantial outflows in connection with the 2013 debt ceiling impasse. See 2014 Final Rule at 47,746. In addition, the capacity of government MMFs to absorb the quantity of assets that will potentially migrate from institutional prime funds is unclear. See letter from Wells Fargo Funds Mgmt., L.L.C., to Elizabeth Murphy, Sec’y, US Sec. & Exch. Comm’n (Apr. 23, 2014), available at http://www.sec.gov/comments/87-03-13/ s70313-340.pdf.
48 The rule contains several additional features, including heightened diversification requirements and stress testing. See 2014 Final Rule at 47,736.
49 2014 Final Rule at 47,827-34.
50 Id. at 47,815 (prospectus and marketing materials); Id. at 47,838 [for N-CR, statement of additional information (SAI)].
51 Id. at 47,829. The requirement that funds calculate and disclose a current market-based NAV is not limited to floating NAV funds; Id. at 47,830.
52 Id. at 47,832 [past use of gates and fees]; Id. at 47,833 [historical sponsor support].
53 Id. at 47,827-29.
54 The SEC itself acknowledged the concern that the regulatory change could itself trigger a run on MMFs. See 2014 Final Rule at 47,790-91.
55 White Statement.
Numerous commentators have weighed in on the feasibility and effectiveness of the 2014 rule, and this article will not reexamine those comments in detail. Instead, it will highlight several reasons why the rule is likely to be ineffective in addressing the SEC’s identified concerns about MMF fragility. In addition, it will identify key problems with the new disclosure requirements that commentators have largely overlooked.

3.2.1 The floating NAV

A floating NAV has been a central component of proposals for MMF reform since the financial crisis. Under the new rule, the floating NAV requirement will only apply to a portion of existing MMFs: MMFs estimated by the SEC to hold almost US$1.3t in assets. The costs of moving to a floating NAV are substantial.

The SEC’s rule was predicated on accounting and tax concessions to simplify compliance issues created by a floating NAV, but it is nonetheless likely that many, if not most, institutional investors will be unwilling or unable to use a floating NAV product. Because each purchase and redemption in a floating fund will occur at a different price, investors will face the prospect of negative yields on a regular basis, making funds unsuitable for many types of investor. The SEC itself has noted the questionable efficacy of a floating NAV in reducing redemptions (“a floating NAV is a targeted reform that may not ameliorate all [the factors leading to heavy redemptions]”).

First, on a theoretical level, it is important to recognize that the arbitrage opportunity created by a fixed NAV only exists during the period in which the fund’s NAV has fallen below US$1, but remains above US$0.995. Once the fund must, by virtue of penny rounding, reduce its trading price to 99 cents, the arbitrage opportunity is reversed because purchasing shareholders can obtain, at a cost of 99 cents per share, assets valued at more than that. Of course, this effect is purely theoretical because only two MMFs in history have ever broken the buck, and no MMF has done so, but continued to operate as a going concern.

Second, existing empirical evidence does not support the claim that a floating NAV reduces redemption pressure in a time of crisis. In 2008, ultrashort bond funds — the floating NAV alternative to MMFs — experienced comparable levels of redemptions to MMFs. Indeed, the total assets invested in ultrashort bond funds declined by more than 60% from their

In addition, institutional investors may shift their money into unregistered and potentially less stable investment alternatives.

Thus, the floating NAV requirement is likely to impose substantial costs. The question for regulators is whether it generates corresponding benefits in terms of improving MMF stability. Advocates of a floating NAV argue that a stable NAV creates an incentive for early redemption. They argue that a floating NAV addresses this problem because redemptions always take place at the fund’s true NAV.

Even defenders of a floating NAV recognize, however, that it reduces the first-mover problem only to a limited degree. The SEC itself has noted the questionable efficacy of a floating NAV in reducing redemptions (“a floating NAV is a targeted reform that may not ameliorate all [the factors leading to heavy redemptions]”).


67 2014 Final Rule at 47,896.

68 FSOC Proposal at 69,467.

69 2013 Rule Proposal at 36,850.

70 Letter from Samuel Hanson, Assistant Professor of Finance, Harvard Bus. Sch. et al., to Elizabeth Murphy, Sec’y, US Sec. & Exch. Comm’n 2 (16 September 2013); Fisch and Roiter at 1036.
peak in 2007 to the end of 2008.71 Similarly, Gordon and Gandia studied the difference in run rates in European MMFs during the financial crisis and found that none of the difference is explained by whether the NAV is fixed or floating.72 A likely explanation for these findings is that the same economic factors that cause investors to redeem from an MMF cause them to redeem from a floating NAV fund. Critically, redemption requests create an analogous first-mover advantage for floating rate funds, as early redemptions can be satisfied through sales of the funds’ most liquid assets.

The key factor contributing to redemption pressure is the stale pricing of mutual fund assets. When an investor redeems mutual fund shares, that redemption request must be honored on the basis of the current value of the MMF’s portfolio, calculated as of the 4 p.m. close.73 As noted, however, the very fact of redemption may require a fund to sell assets at distressed prices, prices that will reduce the fund’s NAV.

Because the typical fund will maintain a certain liquidity level in order to meet redemption requests, the sale of those assets will generally not take place until after the redemption and will therefore not be reflected in the price at which the redemption occurs.74 Thus, the claim that the investor in a floating rate fund exits at the fund’s true value is misstated; in times of heavy redemption, all funds face a first-mover advantage.

Floating the NAV is likely to increase this redemption pressure because investors will then, on a regular basis, expect share prices to decline in response to various economic factors.75 In a stable-value fund, MMF managers face pressure to maintain the US$1 NAV, and, anticipating that, investors do not expect an arbitrage opportunity to materialize. The empirical evidence indicates that this pressure is effective because the NAVs of MMFs fluctuated very little, even in periods of substantial economic turmoil.76 The historical stability of MMFs’ NAVs belies the claim that a stable NAV is misleading or the result of a regulatory dispensation.77 MMFs trade at a US$1 share price because their sponsors manage the portfolios in a way that minimizes any discrepancy between the underlying share value and a dollar.

To eliminate the first-mover advantage, MMFs must do more than float their NAVs; they must satisfy redemption requests at fair value. For reasons described in further detail below (section 4), it is difficult to price MMF assets accurately. As a result, the floating NAV will require MMFs to sell and redeem shares on the basis of “noisy guesstimates of true value.”78 These transactions have the potential to generate far greater unfairness between shareholders than the arbitrage opportunity to which the floating NAV is addressed.

3.2.2 Gates and fees

As with the floating NAV, the SEC’s “gates and fees” alternative appears poorly suited to addressing the central problem identified by regulators as justifying further reform, i.e., run risk. Indeed, gates and fees potentially present a greater threat to MMF investors than a loss in principal, because they jeopardize the investors’ immediate access to their funds. Immediate access is a key factor motivating investor use of MMFs.79

One problem with the gates and fees alternative is its complexity. The final rule empowers fund boards to impose liquidity fees of up to 2% or suspend redemptions (impose gates) for up to 10 business days if a fund’s weekly liquid assets fall below 30% of its total assets.80 At the same time, boards are required to impose a 1% liquidity fee if the fund’s weekly liquid assets fall below 10% of its total assets unless the fund board decides that such a fee is not in the best interests of the fund.81 Thus, the rule provides an opt in for 2% gates and fees and an opt out for 1% fees. The SEC failed to provide meaningful guidance on the “best interests” standard,82 leaving it difficult for investors to estimate or price the }

---

72 Gordon and Gandia at 350.
73 17 C.F.R. § 270.22c-1 (2014).
74 Moreover, in a time of crisis, the inability to value distressed assets makes the fund’s calculation of NAV inherently unreliable. The court observed as much in the case of the Reserve Fund, finding that the Fund’s calculation of NAV on September 15 and 16 2008, was unreliable, even after the Fund broke the buck. See SEC v. Reserve Mgmt. Co. (In re The Reserve Fund Sec. & Derivative Litig.), 673 F. Supp. 2d182, 198 (S.D.N.Y. 2009).
75 Piwowar Statement at 4.
77 See William A. Birdthistle, Breaking Bucks in Money Market Funds, 2010 Wis. L. Rev.. 1155, 1161.
79 Goldman Letter at 1-3.
80 2014 Final Rule at 47,747.
81 Id. (emphasis added).
82 Id. at 47,761.
risk that boards will impose a gate or fee.

In addition, as Commissioner Kara Stein noted, gates and fees are likely to be counterproductive in addressing run risk and the greater problem of systemic contagion. Superficially, of course, gates and fees increase the cost of redemptions. If investors must pay a 1% or 2% fee to redeem their shares, they will be less willing to redeem. Similarly, while a gate is in effect, it completely prevents redemptions. Nonetheless, both gates and fees exacerbate run risk near the point of the trigger. Specifically, if investors are aware of the prospect of a draconian fee or complete bar on withdrawals, they may seek to redeem as the fund approaches the trigger point for the imposition of the gate or fee. The result would be precisely the type of first-mover advantage that this article has identified as a critical component of a run.

The incentive to run under a system with gates and fees would be more powerful than the arbitrage opportunity associated with penny-rounding, because the liquidity fees authorized under the rule are far greater than the 0.5% differential that the SEC identified as a concern under the status quo. Importantly as well, the fund board’s discretion as to whether to impose a gate or fee would generate uncertainty about any particular board’s willingness to do so. Under the final rule, the board’s power to impose gates and fees is triggered if a fund’s liquidity drops to twice the limit legally required by Rule 2a-7, suggesting that boards will potentially be able to exercise this power with some frequency. This uncertainty could lead investors to redeem well in advance of any fund distress, in which case, the investor redemptions, rather than economic developments, could cause the fund to fail.

Moreover, a single fund’s imposition of a gate or fee could scare investors in other funds into redeeming to avoid facing a similar restriction. This would cause one MMF to generate a spillover effect on the industry. Given that the risk of panic may be highest among individual investors—who would also be least able to evaluate the fund’s disclosures in an effort to ascertain the likelihood that a gate or fee will actually be imposed—the use of gates and fees for retail MMFs is particularly problematic. As Eric Rosengren, President of the Federal Reserve Bank of Boston, warns, “As this represents a new run mechanism that does not exist under the status quo, the fees and gates alternative may actually increase run risk relative to not enacting further reform.”

Put differently, Sheila Bair, former FDIC Chair, observed that gates and fees create a new source of uncertainty—the type of uncertainty that generates a run—uncertainty by investors about their ability to withdraw their money.

The biggest problem with gates and fees, however, is that mutual fund boards face powerful disincentives to use them. Although the circumstances under which the imposition of a gate or fee is warranted are likely to be extremely rare, imposing a gate or fee would irreparably damage the reputation not just of the MMF itself, but of its sponsor too. Investors who have been subjected to a gate or fee are unlikely to continue to invest with that fund family in the future. Prospective investors will be wary of investing in a fund that has implemented such restrictions in the past and, under the new rule, MMFs will have to disclose any use of a gate or fee for the next 10 years. In a highly competitive industry, there are reasons to believe that the use of gates or fees will limit a fund sponsor’s ability to attract investments to a degree that makes the survival of the sponsor questionable. This will be a major concern for a board considering the exercise of these powers. Although gates and fees may facilitate the liquidation of an irreparably damaged MMF, they are unlikely to be implemented for funds that are not terminal. As such, their value in enhancing MMF stability is questionable.
3.2.3 The new disclosure requirements

True to the disclosure orientation of the federal securities laws, the new rule adopts an extensive menu of additional required disclosures that offer independent reasons for concern. First, voluminous disclosure requirements may overwhelm investors and limit their ability to ascertain useful information about their investments. The problem of information overload is particularly apparent in mutual fund disclosure; commentators have observed for years that mutual fund regulation mandates too many disclosures that are of questionable value to investors.

Second, and perhaps more problematically, the disclosure is designed to make MMF portfolios, redemption requests and liquidity levels more transparent, ostensibly to enable more effective investor monitoring. Yet, active investor monitoring of MMFs is of uncertain value. Apart from the question of whether MMF investors have the necessary skill set to evaluate MMF risk on the basis of the required disclosures, the private money aspect of MMF is in tension with a high level of information sensitivity. As Tri Vi Dang and others have argued in the context of bank secrecy, it may be desirable to maintain a level of information opacity for financial institutions that produce private money or money equivalents.

Of the new disclosure requirements, two are of particular concern. The first is the requirement that all MMFs calculate and disclose a market-based NAV on a daily basis. For floating-value MMFs, this is the price at which the fund issues and redeems shares; for stable-value MMFs, it is a shadow NAV. Importantly, the rule requires that, in both cases, the calculation be made to four decimal places or to the nearest ten thousandth of a cent. This high level of precision is explicitly designed to create an artificial appearance of volatility in a fund’s NAV. The SEC rejected imposing a precision requirement analogous to that used by other mutual funds, a NAV rounded to three decimal places, on the basis of empirical data showing that only with the more stringent disclosure requirement would MMF prices appear to fluctuate.

As a result, the disclosure conveys a false degree of price fluctuation. The Investment Company Institute (ICI) describes the SEC’s proposal as “an artificially sensitive pricing scheme to force ‘movement’ in the NAVs of the funds.” More troubling is the fact that the use of four decimal places suggests a scientific degree of accuracy to the valuation process that simply is not present.

As noted in section 2.1, many of the assets held by MMFs rarely trade; they are held to maturity and rolled over, which means that when a fund calculates its NAV, current market prices for the securities may not be readily available. In the absence of an available market price, funds are required to determine the “fair value” of the assets that they hold. Fair value determinations are required for all investment funds, but the valuation methodology has a greater impact on funds that hold a large proportion of assets for which market prices are not readily available. Fair valuation methodology incorporates models, predictions and multifactor tests. As a result, although MMF prices will be calculated to four decimal places, they will incorporate valuations that are not scientific, but subjective and imprecise.
The SEC’s new rule also imposes troubling requirements with respect to the disclosure of sponsor support, reflecting the SEC’s view that sponsor support contributes to the fragility of MMFs. The provisions seek to reduce investor reliance on the possibility of sponsor support and to increase the disincentive for sponsors to provide such support. These objectives are troubling, because sponsor support has historically been a key factor enhancing MMF stability.

To achieve the first objective, the rule requires MMFs to inform prospective investors that “[t]he Fund’s sponsor has no legal obligation to provide financial support to the fund and you should not expect that the sponsor will provide financial support to the fund at any time.”

The SEC explains that this language is designed to “emphasize to investors that they should not expect a fund sponsor to provide financial support.” Although, under current law, the statement is certainly factually accurate, given the historical willingness of sponsors to provide such support, it is not clear what message investors are to take from this emphasis.

The message is particularly confusing in the context of the additional new requirement that sponsors disclose all prior instances in which they have provided support over the past 10 years. One possible reading of the disclosures is that investors should ignore the statement about legal obligation because this sponsor has historically gone beyond its obligations and voluntarily provided support. Another possible implication is that, despite the sponsor’s past practice of providing support when necessary, the support is not to be trusted.

Beyond these mixed messages is the question of how investors should interpret a sponsor’s prior practice of providing support. On the one hand, the disclosure might mean that the sponsor has stood behind its fund and has the financial wherewithal to do so. Alternatively, the fact that a fund required prior sponsor support might signal that it is poorly managed or takes excessive risks. Absent some meaningful indication of what sponsor support means, it is difficult to understand how investors can use this information to make informed investment choices.

Regardless of the effect of the signal, requiring sponsors to disclose financial responsibility for a fund’s NAV may reduce their willingness to assume such responsibility. Once a sponsor is forced to tell investors that it need not provide support, it may be unwilling to provide such support voluntarily. Similarly, in the face of a detailed disclosure requirement that will extend for the next 10 years, sponsors may be less willing to provide support in the face of weakness or may delay providing support in hopes that it will prove unnecessary, rather than acting promptly before investors become concerned. Either way, MMF stability is reduced.

4. Mandatory sponsor support: a new approach to MMF reform

4.1 A proposal for mandatory sponsor support

As noted, the SEC’s long-awaited reforms were unlikely to increase MMF stability and, in fact, are counterproductive. This article offers a new approach to MMF reform, i.e., mandated sponsor support of the US$1 share price: the SEC should amend Rule 2a-7 to require sponsors of stable-value MMFs to support the US$1 share price. Sponsors would be required to commit their support as a condition of offering a stable-value NAV MMF. Put differently, the proposal would provide MMF sponsors with a choice. Sponsors could continue to offer a stable NAV MMF, but if they did so, they would be required to commit to maintain the US$1 share price. Alternatively, sponsors could offer a floating NAV MMF, which could be regulated in accordance with the 2014 rule.

The article’s rationale for embracing sponsor support is the critical structural difference between MMFs and banks. MMFs, like other investment funds, consist of a pool of assets that are segregated from the assets of their sponsors. Redemptions from an MMF are made from the MMF’s assets, not from the sponsor’s funds at any time.”


110 2014 Final Rule at 47,816. The statement also adds language warning investors of the possibility that the fund will impose gates and fees. Id. at 47,817.

111 Id. at 47,817.

112 Id. at 47,824.

113 The SEC itself indicated some confusion about this point. See id. at 47,825 (“[t]he disclosure of affiliate sponsor support could have additional effects on capital formation, depending on whether investors interpret financial support as a sign of money market fund strength or weakness”).

114 Although this article does not favor a floating NAV, it retains it to broaden the pool of potential MMF sponsors beyond those who could commit credibly to provide adequate support.
assets. Although the MMF sponsor manages the fund, its financial structure is linked to the MMF only to the extent that it receives fees for the services provided to the MMF.

As John Morley explained, this separation of investments and management is an important and efficient feature of mutual funds because it critically changes the risk exposure of mutual fund investors, who are not exposed to the general operational risks of the mutual fund sponsor. The separation also means that, as a general rule, sponsor assets are not available to MMF investors.

Sponsor support is an exception to this traditional separation because it makes sponsor assets available to MMF investors in the event of MMF distress. Critically, sponsor support is conceptually possible only because the assets of the sponsor are an independent resource rather than part of the MMF’s portfolio value.

Banks, by contrast, lack this separation of investments and management. Bank deposits are a loan from the depositor to the bank and depositors look to the general assets of the bank to satisfy this obligation. The bank’s financial fragility therefore poses a risk to depositors and this risk is the source of bank runs. Banks lack an analogous option of sponsor support because the bank’s resources already stand behind its obligations and there is no additional pool of assets to supplement those resources.

Experience has demonstrated the effectiveness of sponsor support for MMFs that faced substantial redemptions or other forms of financial distress, and prior instances of sponsor support have been frequent, allowing the vast majority of funds to avoid breaking the buck. Fee waivers do not require approval from the fund’s board or the SEC. These waivers have the effect of shifting capital from the investment advisor to the fund itself. Fee waivers have been common for many years; the practice long predates the financial crisis. Susan Christoffersen found that, for example, between 1991 and 1995, over half of retail MMFs and nearly 80% of institutional MMFs waived all or part of their fees.

MMFs dramatically increased their use of fee waivers in the wake of the financial crisis. Since 2008, MMFs waived a total of US$24b in fees. In 2013 alone, MMF fee waivers totaled US$5.8b. A key reason for the fee waivers was the low interest rates that were available on money market assets: absent fee waivers, the funds would generate negative returns. The widespread use of fee waivers demonstrated the willingness of fund sponsors to forgo profits and to absorb virtually all the expenses of operating the funds. By waiving their fees, managers were transferring the waived amount to the funds to support their NAVs and absorbing the funds’ losses on behalf of fund sponsors to forgo profits and to absorb virtually all the expenses of operating the funds. 119

Sponsors have provided support to their MMFs in multiple ways. One of the most common, and the most frequently overlooked, is through discretionary fee waivers. Although MMF advisory fees are set by contract, fund managers have the discretion to waive the fees in whole or in part, and fund managers regularly do so. Fee waivers do not require approval from the fund’s board or the SEC. These waivers have the effect of shifting capital from the investment advisor to the fund itself. Fee waivers have been common for many years; the practice long predates the financial crisis.

The broken buck stops here: embracing sponsor support in money market fund reform

115 Morley at 1228, 1258-59.
117 McCabe at 1-2.
118 Id. at 8.
119 Id. at 35.
120 The fact that even the SEC has overlooked the importance of fee waivers is demonstrated by the fact that its new disclosure requirement for sponsor support does not include fee waivers. For example, Susan Christoffersen, Why Do Money Fund Managers Voluntarily Waive Their Fees?, 56 J. Fin. 1117, 1119–25 (2001).
121 Id. at 1119.
122 Id. at 1139.
123 Id.
124 Id. at 1139.
125 Tim McLaughlin, US Stock Fund Costs Fall; Money Market Fee Waivers Hit US$5.8b, Reuters (May 14, 2014), http://www.reuters.com/article/2014/05/14/funds-stocks-fees-idUSL1N0NZ13320140514.
126 Id. By way of comparison, total fee waivers in 1995 were US$348m. See Christoffersen at 1120.
128 McLaughlin.
of the investors.\textsuperscript{129} Importantly, by structuring sponsor support as a fee waiver, in which the fund manager has a contractual entitlement to a payment, and then voluntarily and discretorily waives that payment, MMFs avoided the accounting and regulatory complications that would accompany an explicit guarantee or ex-ante commitment. Although recent fee waivers were informal, discretionary and largely clandestine, they demonstrated the viability of sponsor support in enhancing MMF stability. It should be noted that many MMF sponsors reduced or eliminated fee waivers in the US and across Europe in 2016–17.\textsuperscript{130}

In addition to fee waivers, sponsors provide support by purchasing distressed assets from a fund at amortized cost, providing direct injections of capital or liquidity, or providing letters of indemnity or other types of guarantee.\textsuperscript{131} As history demonstrates, sponsors have a strong incentive to support their MMFs.

Breaking the buck could irreparably damage a sponsor’s reputation and make it unable to continue to operate.\textsuperscript{132} Importantly, for the vast majority of sponsors, MMFs represent only a small proportion of their overall business and the spillover effect could destroy the sponsors’ other operations as well.\textsuperscript{133} Sponsor support is not costly in the context of most sponsors’ overall operations. MMFs generally constitute a small percentage of the sponsor’s assets under management\textsuperscript{134} and the potential cost of furnishing support to the funds is a tiny portion of the sponsor’s independent value. BlackRock, for example, manages approximately US$300b in MMF products, out of a total of more than US$4t in assets under management.\textsuperscript{135}

BlackRock has an independent value of US$51b, meaning that US$51b of value from the sponsor’s shareholders is available to meet the potential demands of its MMFs.\textsuperscript{136} In essence, BlackRock’s market capitalization provides a capital buffer of 17%.

In addition, MMF sponsors receive a percentage of the total assets under management in the form of regular and highly liquid advisory fees.\textsuperscript{137} These fees provide a ready source of liquidity upon which the sponsor can draw to meet its support obligations.\textsuperscript{138} Indeed, even in an era in which MMF yields have plummeted, MMF sponsors continue to receive substantial fee income.\textsuperscript{139}

In addition, sponsor support is flexible, in that sponsors have multiple options as to the form of support provided.

Under current law, sponsor support is always discretionary.\textsuperscript{140} This is widely characterized as a weakness of MMFs because of the fear that the sponsor may fail to provide support in time of crisis and allow the MMF to fail.\textsuperscript{141}


130 For example, Aliya Ram, Fees for money market funds reinstated, Fin. Times, Oct. 9, 2016, https://www.ft.com/content/3a20252f-8ade-11e6-bcb7-e74a11d133b1 (describing reduction in availability of fee waivers). It is notable that institutional investors, at least in Europe, appear willing to use cash management vehicles that provide negative returns, perhaps because of the fees associated with bank deposits. See id. (noting that MMFs with negative returns offer improved diversification over bank deposits).

131 Shilling at 3.

132 For example, Kacperczyk and Schnabl at 3; McCabe at 6.


136 Id.


138 For example, the US$116b Fidelity Cash Reserves Fund generated US$200m in income for its sponsor in each of the three years from 2010 to 2012. Thus, the fund’s annual income greatly exceeded the degree of fluctuation in value targeted by the SEC’s rules as a cause for concern. See id.

139 McLaughlin (2012).

140 Various legal rules prevent sponsors from guaranteeing fund value in advance, including limitations on affiliate transactions, requirements for reporting contingent liabilities, consolidation requirements and, for regulated entities, a concern about extending the federal safety net to a nonbank entity. See SEC Staff Issues Clarification on Consolidation Issues Relating to Bank Support for Money Market Funds, Ernst & Young (Sept. 19, 2008), http://goo.gl/0VbMWG. In September 2008, the SEC staff issued guidance to clarify that banks were not required to consolidate the fund on their balance sheet if they provided discretionary support in connection with the financial crisis. US Sec. & Exch. Comm’n, Sec Issues Clarification on Accounting Issues Relating to Bank Support for Money Market Mutual Funds, Ernst & Young (Sept. 17, 2008), http://www.sec.gov/news/press/2008/2008-205.htm. Similarly, accounting rules limit the ability to segregate assets to cover potential losses in the form of a reserve because the sponsor has no obligation to cover those losses. Although accounting rules would allow reserves if the sponsor guaranteed its MMF losses in advance, such a guarantee would subject the sponsor to disclosure obligations and possible consolidation. See PWG report at 10. In addition, an explicit guarantee could arguably be treated as an asset of the fund, with the resulting requirement that it be reflected in the fund’s NAV.

This article’s proposal would modify Rule 17a-9 to require sponsor support rather than making such support voluntary.\textsuperscript{142}

Mandating sponsor support would address former Chair Schapiro’s concern about the unreliability of sponsor support in a time of crisis.\textsuperscript{143} To address the related concern about transparency, this article proposes several complementary disclosure requirements. MMFs would be required, on a real-time basis, to disclose the extent and form of support provided.\textsuperscript{144} MMFs would also be required to disclose any conditional forms of support, including insurance coverage, contingent purchases and third-party guarantees.

Importantly, these disclosures would not have the potentially adverse consequences of the requirements included in the 2014 rule because, in a regulatory environment in which sponsor support is mandated, disclosure that the sponsor has provided such support would not be a confusing signal about the need for support or the sponsor’s willingness to provide support in the future.

In addition to this disclosure, MMF sponsors would have to provide disclosures about their financial condition. For sponsors that regularly provide current financial information to the public, either through capital markets disclosures or through publicly available filings with regulators, such information would be sufficient. Private sponsors that are not otherwise subject to mandated financial disclosure, such as the Reserve Management Company or Fidelity, would be required to provide analogous periodic disclosures to allow investors to evaluate their capacity to meet the support requirement.

4.2 Advantages of mandatory sponsor support over current law

Mandatory sponsor support is a better approach than the 2014 rule for three reasons. First, sponsor support is effective.

As the SEC and others have documented, sponsors supported the NAVs of their MMFs for years.

Sponsor support has enabled hundreds of MMFs to weather the turmoil of the financial crisis of 2008, the European debt crisis, uncertainty about the US debt ceiling, the structured investment vehicle issue and more without breaking the buck. Notably, a commitment to sponsor support reduces run risk because it eliminates the pressure for investors to redeem. As a result, in most circumstances, the guarantee alone will be sufficient to provide stability without requiring the sponsor to incur substantial cost.

Second, sponsor support provides appropriate incentives for sponsors to minimize portfolio risk.\textsuperscript{145} One of the ongoing concerns about MMFs is the potential that sponsors will take excessive risk to increase yield and obtain a competitive advantage.\textsuperscript{146} Although the SEC’s 2010 MMF reforms reduce the degree of permissible risk taking, so long as the sponsor does not bear the full costs of its risk taking, it will have an incentive to take excessive risk.

This appetite for risk is likely to be concentrated in those MMF sponsors that are financially fragile or those that lack independent business reasons for maintaining a sound MMF.

Third, sponsor support does not create the moral hazard problem associated with external financial support, such as the reasons — supporting the fund’s NAV and supplying liquidity — are implicit in the regulatory mandate.

\textsuperscript{142} 17 C.F.R. § 270.17a-9 (2014). Critically, in order to allow a sponsor to commit to these forms of support up front, the SEC would need to modify the rules on affiliate transactions, calculation of NAV and consolidation, where necessary.


\textsuperscript{144} MMFs would be required to disclose all forms of support, including fee waivers. Unlike the current rule, the proposal would not require sponsors to disclose the reason for providing support, as the reasons – supporting the fund’s NAV and supplying liquidity – are implicit in the regulatory mandate.

\textsuperscript{145} Importantly, unlike capital buffers, sponsor guarantees would not provide a discontinuity with respect to sponsor incentives. With required capital buffers, a sponsor’s incentive to take risk increases as potential losses approach the size of the buffer, because the sponsor will not bear the cost of losses beyond the amount of the buffer. Thus, capital buffers create a distortion analogous to that created by low capital requirements for banks.

\textsuperscript{146} For example, commentators described the Reserve Fund’s risk taking as excessive, noting that, in September 2008, the Reserve Fund’s 12-month yield was “the highest among more than 2,100 money funds tracked, according to Morningstar.” Steve Stecklow and Diya Gullapalli, A Money-Fund Manager’s Fateful Shift, Wall St. J. (Dec. 8, 2008), http://online.wsj.com/news/articles/SB122869788400386907. This yield made the Fund an attractive investment—“the fund’s assets tripled in two years to US$62.6b.” Id.

\textsuperscript{147} Indeed, Patrick McCabe (at 34) finds that investors are capable of distinguishing among MMF sponsors and that the MMFs associated with risky sponsors experienced a higher level of institutional redemptions during recent economic crises.
as a private liquidity facility or an industry-wide insurance or guarantee system, because each sponsor is individually responsible for the stability of its own funds. Sponsors that take excessive risk with their MMF portfolios cannot draw upon resources contributed by more conservative sponsors. Similarly, because MMFs look to their individual sponsors for support, rather than to a common pool, the contagion effect of individual MMF fragility would be contained. Even if a particular sponsor experienced financial distress, that distress would have a limited effect on investors’ expectations about the stability of other funds.

4.3 Possible objections and responses

The most likely objection to this article’s proposal might be: if mandatory sponsor support is such a good idea, why hasn’t someone proposed it? The answer to this question is that the viability of sponsor guarantees has been masked by the political dynamic in which MMF reform has been debated. Commentators have described MMFs as shadow banks, termed the 2008 redemptions a run akin to bank runs, and proposed reforms, such as capital buffers, designed to make MMFs more like banks.

As noted above, mandatory sponsor support is only possible because of the unique separation of management and investments in MMFs, a separation that does not exist in a traditional bank.

Politically, sponsor guarantees are also an unattractive option for both key interest groups: banks and mutual fund sponsors. From the perspective of banks, sponsor guarantees highlight the difference between banks and MMFs by tapping a source of financial stability that banks cannot replicate.

To the extent that MMFs offer an attractive, competitive product, explicit sponsor guarantees would allow them to continue to offer that product without facing the regulatory burdens of banks. Mutual fund companies, which sponsor roughly half of MMFs, would likely also oppose explicit sponsor guarantees because of their potential cost. In light of the ability of several of the most powerful mutual funds to avoid a substantial regulatory burden through the exemption for retail funds, a support commitment would obviously be a less attractive option.

A second objection might be that explicit sponsor guarantees are too costly. As noted, one advantage of this article’s proposal is that, because it offers sponsors a variety of mechanisms for meeting their obligations and does not mandate an explicit set-aside of capital, it will be less expensive to implement than an alternative such as a capital buffer.

Nonetheless, the contingent liability associated with a guarantee, a standby letter of credit or the purchase of illiquid assets (even if those assets will trade at par on maturity) becomes more costly as the size of an MMF grows. As a result, sponsor support may require a sponsor to limit the size of its MMFs to reduce its liability exposure.

The possibility that requiring sponsor support might lead sponsors to limit MMF size may well be an additional advantage of the proposal. Such an effect would reduce the systemic importance of any single MMF. If it is problematic for banks and other financial institutions to get too big, regulation that indirectly places a practical limit on MMF size seems, at worst, benign. Admittedly, MMFs, like other mutual funds, do enjoy economies of size and scale, and funds would sacrifice those economies if they were limited in size.

In addition, although financially sound sponsors with other substantial business operations, such as large mutual fund companies, could likely fund any support obligation through operating capital or other assets, sponsor support may be particularly burdensome for smaller sponsors or those that lack other businesses. Thus, explicit guarantees might have the effect of precluding certain types of sponsor from offering MMFs, either because they lack the assets to guarantee fund value or because the market would be skeptical of their ability to meet the support obligation. In retrospect, investors

149 Twenty years ago, Howell Jackson proposed a conceptually similar alternative to increased regulation of financial holding companies. Howell E. Jackson, The Expanding Obligations of Financial Holding Companies, 107 Harv. L. Rev. 507, 583 (1994). Much of Jackson's reasoning can be applied to the context of MMFs.
150 FSOC Proposal at 69, 456.
151 Notably, the cost of fees and gates would be borne by MMF investors, not by fund sponsors. 2013 Rule Proposal at 36, 907.
152 Part of the contagion effect generated by the Reserve Fund’s failure was due to its size: US$62.6b in assets. See Stecklow and Gullapalli.
might be skeptical that the Reserve Management Company, “a stand-alone fund company with almost no other funds under management,”155 would provide support to its MMFs. By contrast, investors might be more confident in relying on support from a company, such as Fidelity, which, in 2006, “sponsored 252 non-money market mutual funds with US$814b in assets under management.”156

Sponsors with other businesses face the most spillover risk if their MMF is fragile and, accordingly, are likely to take steps to prevent that by reducing the riskiness of their MMF assets.157 Similarly, funds with substantial non-MMF assets are in a better position to provide support.

Again, to the extent that this reform has the effect of reducing the ability of financially compromised sponsors to offer MMFs, this article views that effect as an advantage, not a weakness. Notably, explicit sponsor support changes the focus of market discipline by properly focusing investors on sponsor financial stability in evaluating MMFs rather than on the MMF portfolio. As discussed above, the structure of MMFs makes direct investor monitoring of portfolio assets problematic. The portfolio’s assets are extremely short term, meaning that the quality of the portfolio can change rapidly. Portfolio assets are, in many cases, thinly traded and difficult to price. Finally, the MMF’s shadow prices tend to be stale, incorporating future economic developments incompletely. By contrast, investors can readily monitor the financial condition of MMF sponsors and identify the business practices that provide economic incentives for sponsors to meet their support obligations. This more efficient investor monitoring is a distinctive advantage of this article’s proposal.

At the same time, this article’s proposal would not preclude smaller and less stable sponsors from offering MMFs. Sponsors could address investors’ concerns about incentives and solvency through a variety of mechanisms, including explicit guarantees, standby letters of credit or purchasing insurance, to cover their support obligations.158 In particular, insurance offers another mechanism for monitoring MMF risk, as insurance providers have an economic incentive to understand an MMF sponsor’s risk profile, and tailor the cost and scope of coverage accordingly.159

Importantly, insurance for an individual sponsor’s MMF obligations would be quite different from industry-wide mandated insurance, and would not create the same concerns about cost and moral hazard.160

On the other hand, the sponsor support requirement could operate as a barrier to entry for smaller potential sponsors, reducing competition in the industry. Critics may also question whether sponsor support is sufficiently reliable. What happens, under this approach, if a sponsor defaults, as the Bents did?161 While mandatory sponsor support cannot eliminate the possibility, such a default would be no different from the failure of any financial institution to meet its obligations. Unlike the current system, however, investors would not face unpredictability about whether a sponsor would provide support because such support would be required rather than voluntary. In addition, an MMF sponsor that failed to meet its obligations would face the prospect of an enforcement action, not just the uncertain penalty of market discipline.

As a result, solvent and financially responsible sponsors are unlikely to default, and, as noted, investors should have adequate information to identify and avoid sponsors that cannot credibly commit to support their funds.

Finally, this article’s proposal may be criticized on the basis that it would undermine the efficient asset partitioning that is a key component of the MMF structure.162 Sponsor support would make sponsor assets available to meet MMF shortfalls. Importantly, the

---

155 Kacperczyk and Schnabl at 11.
156 Id.
157 Id. at 3.
161 It is unclear whether the Reserve Management Company was unable or unwilling to provide support for the Reserve Fund, but, as detailed in the SEC’s enforcement action, the defendants issued a number of public statements indicating that they intended to provide sponsor support; support that never materialized. See Complaint at 2–3, SEC v. Reserve Management Co. Inc., 732 F. Supp. 2d 310 (S.D.N.Y. May 5, 2009) (No. 09 CV 4346). The SEC’s fraud case was based on the claim that, at the time the Bents made these promises, they had no intention of providing such support. Id.
162 Morley, at 1240-41.
interference with asset partitioning would be both limited and operating only in one direction.

Sponsor assets would only be available to the extent necessary for the MMF to maintain a stable US$1 NAV. The added liability exposure of fund sponsors could, of course, be viewed as another justification for classifying investment managers as systemically important financial institutions (SIFIs).163

Whether it is appropriate for the FSOC to designate asset managers as SIFIs164 is a controversial topic and beyond the scope of this article,165 although it is worth noting that for the vast majority of asset managers, MMFs constitute a small percentage of their total assets under management.166

In addition, MMF assets would not be available to meet the needs of a financially distressed sponsor. This segregation of MMF assets from sponsor assets provides the key value of asset partitioning in the mutual fund structure.167

The segregation is illustrated by the failure of Lehman. Notably, although Lehman's bankruptcy brought down the Reserve Fund, it did not bankrupt Lehman's own mutual funds. The assets of those funds were segregated by law and out of the reach of Lehman's creditors.168

5. Conclusion
Since 2008, MMFs have been targeted for broad-based regulatory reform. In 2014, the SEC adopted a rule that may have draconian consequences for some types of MMF, while failing to address core concerns about MMF stability.

The limitations of the rule can largely be attributed to the flawed process by which it was produced and, in particular, the politics of MMF reform. In particular, by analogizing MMF redemptions to bank runs and debating proposed reforms on the basis of whether they will reduce a run risk to zero, policymakers have set an unrealistic objective for MMF reform and imposed a risk-reduction requirement far beyond that applicable to the banking industry.

In addition, by painting MMFs as part of the shadow banking system, critics have overlooked the critical attribute that distinguishes MMFs from bank deposits – the structural separation of MMFs from their sponsors. This attribute provides the key to increased MMF stability. The solution for reducing MMF fragility lies within MMFs themselves, in the form of explicit sponsor support. Although existing law prevents sponsors from committing to maintain a US$1 share price, this article argues that such a commitment is desirable. It therefore proposes that MMF sponsors should be required to support the US$1 share price of their fixed NAV MMFs.

Importantly, however, sponsor support should not be mandated through a rigid and costly vehicle such as capital buffers or mandatory insurance. MMF sponsors come in a variety of different shapes and sizes, and this variety offers a range of possible support mechanisms that take advantage of the sponsors’ reputation, outside assets, and overall business plans.

As a result, sponsor support should be permitted through the range of mechanisms that have been used successfully throughout the history of the MMF, including fee waivers, guarantees, capital infusions and the purchase of MMF securities at par. By mandating sponsor support, regulators can formalize existing support practices that have proved valuable in maintaining MMF stability, while increasing the transparency of sponsor support to the market.

163 In September 2013, the Office of Financial Research released a report produced at the request of the FSOC to enable the FSOC to consider whether asset managers should be considered for enhanced regulation as SIFIs. Office of Fin. Research, Asset Management and Financial Stability 1 (2013), http://www.treasury.gov/initiatives/oh/research/Pages/AssetManagementFinancialStability.aspx [hereinafter OFR report]. The OFR concluded that asset managers can “introduce vulnerabilities that could pose, amplify, or transmit threats to financial stability.” Id.

164 It is also conceivable that the FSOC could designate MMFs themselves as SIFIs. For a detailed analysis of why such a designation would not be appropriate, see Eric D. Roser, Should Money Market Funds Be Designated as “SIFIs”? 31 Rev. Banking & Fin. L. 749, 760 (2012).

165 Both the report and the prospect that the FSOC could designate certain asset managers as SIFIs have generated considerable controversy. See, e.g., Emily Stephenson and Sarah N. Lynch, US Senators Slam Study on Systemic Risks Posed by Asset Managers, Reuters (Jan. 24, 2014), http://www.reuters.com/article/2014/01/24/us-financial-regulation-asset-idsUSBREAOG1220140124.

166 OFR report, supra note 172, at 20. Federated Investors is an exception; its business consists primarily of institutional money market funds. Id.

167 Id. at 1.

The broken buck stops here: embracing sponsor support in money market fund reform

References


Sheila Bair, Chair, Systemic Risk Council, “Letter to Elizabeth Murphy, Secretary, US Securities and Exchange Commission,” 16 September 2013.


US Senate, Perspectives on Money Market Mutual Fund Reforms: Hearing Before the Senate Committee on Banking, Housing, and Urban Affairs, 112th Congress 75, 2012 (statement of Paul Schott Stevens, President and Chief Executive Officer, Investment Company Institute).


Good and bad risk: regulation and loan monitoring

Norvald Instefjord
Reader in Finance, Essex Business School, University of Essex

Hiroyuki Nakata
Senior Lecturer in Finance, School of Management, University of Leicester
Abstract
Separating good and bad borrowers is a key role of banks. To do this, banks need monitoring systems and they need to monitor risky loans. We show that the investment in monitoring systems encourages risk taking, which leads to higher regulatory costs for the bank. This effect is so strong that it not only discourages investment in monitoring systems, but also banks can profitably dismantle their existing systems, because the savings in terms of regulatory compliance costs are greater than what they lose to less efficient loan monitoring. A bank regulator who controls bank risk in an indiscriminate way, therefore, can distort the loan-monitoring activity of banks, which can be harmful to the cost of loans. A more sophisticated approach, where the regulator discriminates between the good risk that arises from loan-monitoring activity and the bad risk that arises in other contexts, can mitigate this effect.
1. Introduction
Monitoring and screening of loans are central functions of banks, which can generate social values. Diamond, for instance, provides a theoretical model of the role of banks as monitors, which is instrumental in minimising the cost of producing the information that is necessary to facilitate efficient lending and borrowing. Bernanke argues that the monitoring function of banks was affected by the 1930s financial crisis and led to inefficiencies in bank lending in the US, which contributed to the 1930s US depression. Thus, it appears that efficient monitoring of borrowers to ensure credit quality is important for a well-functioning financial system, and consequently for a well-functioning real economy.

However, the effects of regulations on banks’ incentives for investment in monitoring systems are not straightforward. One aspect that complicates the issue is that loan sales have become substitutes for monitoring activity, as argued, for instance, by Parlour and Plantin. Some are critical of increasing the regulatory burden for banks, particularly during a crisis. For example, the former CEO of Deutsche Bank Josef Ackerman has argued that more stringent capital requirements put upon banks “would restrict [banks’] ability to provide loans to the rest of the economy. This reduces growth and has negative effects for all”. In addition, Dermine argues that monitoring activity is adversely affected by heavy-handed regulation. In contrast, Admati and Hellwig argue strongly against lifting the regulatory requirements on banks. That regulation reduces the incentive to take risk (through lending), and therefore, also indirectly, the incentive to monitor loans is, however, relatively uncontroversial. As such, the debate is primarily about whether the risk of lending and the monitoring of loans are affected by regulation in a socially harmful way.

This paper studies the issue by focusing on the link between risk and monitoring. If we take an intuitive starting point, the link appears simple. As long as the banks benefit from the separation of good and bad borrowers, one would think that the monitoring systems have private value to the banks’ shareholders because the banks can make more informed decisions about whether to keep the loan or sell the loan. And while the banks use these systems to avoid bad lending – which is not implausible given the fact that regulation typically penalizes credit risk – they have social value to the banks’ regulators. Loan monitoring is, therefore, an apparent win-win for the shareholders and society. We argue that this view is too simplistic because it applies to the screening activity of loans ex ante prior to the lending decision being made, but not to the monitoring of loans that have been granted ex-post.

Once we study monitoring of loans that have already been made, the above simplistic logic breaks down. The bulk of this paper elaborates on this point in particular, by focusing on the dynamic nature of monitoring. People would like the bank to monitor loans that are risky. The investment in monitoring technology gives banks the incentive to hold a higher quantity of risky loans under monitoring, which increases the loan risk of banks. This risk leads to higher regulatory compliance costs for the bank, with the regulator needing to impose extra costs to create a disincentive for loan monitoring. The risk associated with the increase in monitoring activity has, however, social value as it facilitates and lowers the cost of loans to risky borrowers, such as young and small enterprises, which may yield higher expected returns. There is, therefore, a regulatory trade-off to be considered.

The empirical findings regarding the effects of bank regulation are mixed. By analyzing the bank performance across different countries during the financial crisis of 2007-08, Beltratti and Stultz show that strong regulation is likely to reduce risk but active regulatory intervention to respond to market changes is likely to reduce performance. There is, therefore, some evidence that regulatory intervention can harm shareholder value. Baker and Wurgler show that increasing the stringency of bank regulation could increase significantly the cost of lending. Barth et al. find that regulation that puts restrictions on bank activity has a negative impact, both in terms of bank performance and stability. They do, however, find that more stringent capital

1 Diamond (1984)
2 Bernanke (1983)
3 Parlour and Plantin (2008)
4 Admati and Hellwig (2013)
5 Dermine (2013)
6 Admati and Hellwig (2013)
7 Beltratti and Stultz (2009)
8 Baker and Wurgler (2013)
9 Their findings should however be viewed in the light of similar findings for the stock market as a whole, which suggest a general “low-risk anomaly”: Baker et al. (2011) and Baker et al. (2013)
10 Barth et al. (2004)
requirements are associated negatively with nonperforming
loans, which suggests that banks with more strict capital
requirements cut back on risky lending. Cebenoyan and Strahan\textsuperscript{11} find that banks that adopt a more active risk management policy
(buy and sell loans more actively) also hold more risky loans,
have lower risk, and higher profits than other banks. This study
does not, however, give information about whether these banks
are subject to less stringent capital requirements, although more
active risk management is associated with lower bank capital.
Laeven and Levine\textsuperscript{12} find that the relationship between capital
requirements/bank restrictions and risk-taking is positive when
the bank has a strong equity holder, and negative without such
a strong equity holder. Furthermore, Klomp and de Haan\textsuperscript{13} find
that bank regulation has an impact on the risk of high-risk banks,
but that this effect is not uniform across the various levels of risk.
Therefore, empirically it is very hard to talk about the effect
of regulation on bank risk.

Our paper is a version of a more technical paper,\textsuperscript{14} to which
the reader should turn for technical details and proofs, as the
exposition here emphasizes intuition rather than rigor. We
proceed as follows: in section 2, we outline the basic framework.
The main results are reported in section 3. We discuss the
regulatory implications in section 4. Section 5 concludes the
paper.

2. Framework

2.1 Ex-post monitoring decision rules

In this section, we describe a framework for ex-post monitoring of
loans that have already been granted. The framework is stripped
down into its most simplistic form. We assume, therefore, that all
loans are perpetual loans whose value depends on the coupon
flow of the loan and the losses related to the credit risk of the
loan. The coupon flow is \( c \) and we assume a constant risk-free
rate \( r \) and a risk-neutral probability measure, which implies that
all asset values are simply expected discounted cash flows. If a
loan is of high credit quality, we assume it is risk-free and carries
no credit risk. The value is consequently the discounted coupon
flow:

\[
V(c|\text{good}) = \int_{0}^{\infty} e^{-rt} c dt = \frac{c}{r}
\]

If the loan carries credit risk, there are shortfalls in the cash flow
of the loan. If the loan is of low credit quality, these losses equal
the expected present value \( k \), and the value of the loan is:

\[
V(c|\text{bad}) = \frac{c}{r} k
\]

Loans are generally perceived as good with some probability \( \pi \)
and bad with probability \( 1-\pi \), so the market value of a loan of
quality \( \pi \) is:

\[
V(c|\pi) = \frac{c}{r} (1-\pi) k
\]

The expected loss due to credit risk, \( k \), cannot be greater than
the risk-free value of the loan, \( k \leq \frac{c}{r} \), since the worst-case scenario
is that the lender receives no payments from the borrower. The
bank, however, is regulated and has a lower ability to bear credit
risk, which is represented by \( k+K \) with \( K > 0 \), where \( K \) is set by
the regulator. Thus, if the bank seeks to place a loan of quality \( \pi \n in-house, the value of the loan is:

\[
V(\pi) = V(c|\pi) - K(1-\pi) = \frac{c}{r} (k+K)(1-\pi).
\]

The decision the bank has to make is: (a) to sell the loan to the
investors, (b) to keep monitoring the loan or (c) to stop monitoring
and to keep the loan in-house. Thus, the bank keeps monitoring
the loan until it decides either to sell the loan or keep it in-house,
and such decisions will be made on the basis of the information
the bank acquires through ex-post monitoring.

When the bank sells a loan to the investors, it incurs a signaling
cost \( g \), so the net value that can be recovered by the bank of a
loan of quality \( \pi \) sold to the market is \( V(c|\pi)-g \). Note that if the
signaling cost \( g \) were zero, the bank would sell all loans to the
investors. Also, if the risk-bearing ability were the same for the
bank as for the investors, i.e., \( K=0 \), the bank would keep all loans
in-house since loan sales are costly. Thus, there is a trade-off —
the bank seeks to balance the gains from trade against the cost of
the loan sale transaction. For high-risk loans, the gains from trade
will typically dominate the signaling cost, whereas for low-risk
loans, the reverse will be true. We study the risk management
decision for the bank as a function of its access to monitoring
technology.

\textsuperscript{11} Cebenoyan and Strahan (2004)
\textsuperscript{12} Laeven and Levine (2009)
\textsuperscript{13} Klomp and de Haan (2012)
\textsuperscript{14} Instefjord and Nakata (2015)
A bank with screening technology makes an initial credit scoring of loans to generate a probability $\pi_0$ at the time the loan is considered. The bank has access to ex-post monitoring technology, which gradually reveals the true quality of the loan which is either good or bad, and the monitoring technology yields a posterior belief $\pi_t$. Instefjord and Nakata\textsuperscript{15} demonstrate that if a noisy Brownian signal about loan quality can be observed, where the diffusion parameter is $\sigma$ and the drift parameter is $\mu$ if the loan is good and zero if it is bad, then it leads to a posterior belief process that follows the stochastic differential equation:

$$d\pi_t = \frac{\mu}{\sigma} \pi_t (1 - \pi_t) dB_t$$

The ratio $\frac{\mu}{\sigma}$ is the signal-to-noise ratio of the signal produced by the ex-post monitoring system and $B_t$ is a standard Brownian motion that represents the fact that the signal is one. As the signal-to-noise ratio increases, the posterior beliefs will converge more quickly to a signal of good credit quality, $\pi \rightarrow 1$, or to a signal of bad credit quality, $\pi \rightarrow 0$. We assume that all new information about the loan is captured by the information signal, which implies that the cash flow from good and bad loans are observationally equivalent during the monitoring period.\textsuperscript{16}

Instefjord and Nakata (2015) show that the decision rules of the bank concerning ex-post monitoring are described as follows:

(a) The bank stops monitoring and keeps the loan in-house if $\pi_t \leq \pi^*$

(b) The bank keeps monitoring if $\pi_t \in [\pi^*, \pi^+]$

(c) The bank stops monitoring and sells the loan if $\pi_t > \pi^*$

The two trigger values $\pi^*$ and $\pi^+$ are determined by the parameters of the monitoring system, the signal-to-noise ratio and the monitoring cost $m$, and if they exist, they are unique.

2.2 Risk measure and intermediation surplus

The natural risk measure for a loan of quality $\pi$ is the discount of a risky loan relative to a risk-free one. Specifically, the risk associated with the loan is simply the discount $\frac{\pi}{1 - \pi}$. For a loan portfolio with loans of quality $\pi$ uniformly distributed over $[\pi^-, \pi^+]$, the loan portfolio risk is:

$$\pi \int k(1 - m) \pi = k(\pi - m) \left(1 - \frac{\pi - m}{2}\right)$$

Since the bank holds loans of type $\pi$ varying uniformly from $\pi^*$ to 1 by assumption, the risk of the bank's loan portfolio is: $\frac{\pi}{2(1 - \pi^* - \pi)}$.

Note that the risk measure $\frac{\pi}{2(1 - \pi^* - \pi)}$ is a market-based measure of the need to hold risk capital, since it depends on the discount of the market value of loans relative to a corresponding risk-free loan, i.e., the risk is measured relative to the market's risk-bearing ability $k$. If we made the risk measure on the basis of the internal risk-bearing ability $k + K$, it would depend on the regulator's actions and would mean that a more strict regulation of a bank would itself lead to a higher risk in the bank's loan portfolio without any action on the part of the bank. To prevent this from happening, we make the risk measure a market-based risk measure, broadly in line with the risk assessment models used in practice. An implication of the risk measure is that the regulator requires a higher cost of bearing risk, for instance, more strict capital requirements of the banks in which the barrier for holding risky loans is low (i.e., $\pi^*$ is low).

We consider the following two cases for the determination of the regulatory cost of capital $K$:

(a) $K$ is fixed regardless of the risk level of the bank's portfolio.

(b) $K$ is chosen so that the risk level of the bank's portfolio meets a target level $\Psi$.

Note that in the second case, $K$ is set so as to satisfy $\frac{\pi}{2(1 - \pi^* + K + \Psi)}$, and the expression $\pi^*(K^*)$ reflects the fact that the bank chooses $\pi^*$ in accord with the bank's decision rules described above, given $K^*$. Namely, $\pi^*$ is really a function of $K^*$, and $K^*$ is set so that the risk level is matched with $\Psi$, which

\textsuperscript{15} Instefjord and Nakata (2015)

\textsuperscript{16} This is an assumption that probably seems more unrealistic in theory than in practice, as borrowers typically seek to make their loan repayments as long as they can, because when they start missing their repayments the contractual penalty can be harsh. From the point of view of the lenders, they typically benefit from managing the risk of the loan before the repayments fail to materialise, as afterwards the options available for such management can be limited.
has a one-to-one relationship with $\pi^{**}$ given $k$. Thus, setting $\Psi$ is equivalent to setting $\pi^{**}$, and we can interpret that the regulator imposes the regulatory cost of risk capital $K^*$ on the bank so that the bank will set its lower threshold $\pi^{**}$ at a level targeted by the regulator. In other words, $\pi^{**}$ is fixed in this case.

Now, we turn our attention to a measure of the value-added of the bank through its monitoring capability and the keeping of the loans in its own book, which we call the “intermediation surplus.” A loan of quality $\pi$ has a market value of $V(\pi) - g$, conditional on loan quality $\pi$ and net of signaling costs. The bank can improve on the market value in two ways. The bank can either hold the loan in-house at the value $V_M(\pi)$, or keep the loan under observation in its monitoring system with the aim of delaying the decision to hold or to sell, at the value $V_I(\pi)$. The first option saves the signaling cost $g$, but incurs a higher regulatory cost of risk capital as determined by the regulator’s choice of $K$. The intermediation surplus is, therefore, simply the maximum of these three values minus the market value, i.e.:

$$W(\pi) = \max(V(\pi) - g, V_M(\pi), V_I(\pi)) - (V(\pi) - g)$$

$$= \max(0, V_M(\pi) - (V(\pi) - g), V_I(\pi) - (V(\pi) - g))$$

### 3. Main result

The main result is twofold. First, we demonstrate that an improvement in monitoring technology leads to greater loan values, provided the extra cost of capital $K$ imposed by the regulator is fixed. Second, we demonstrate that a win-win between the bank’s shareholders and the regulator from investments in monitoring systems is not feasible when the risk level the bank is allowed to bear is fixed by the regulator. As long as the regulator controls risk, the bank’s shareholders will create greater shareholder value from dismantling the bank’s monitoring systems. This is owing to a reduction in regulatory compliance costs, than they will lose owing to a reduction in the quality of risk management.

The first result is that loan values increase as the monitoring technology improves, provided the regulator does not respond to the increase in bank risk that follows improved monitoring. Improved monitoring has implications for loan values in the monitoring region, but will not change the loan values outside the monitoring region. This is illustrated in figure 1. Figure 1 shows the loan value for the base case where the bank uses a monitoring technology of a certain standard, and two alternative cases where the loan technology is more expensive than in the base case and where the signal-to-noise ratio is lower than in the base case, respectively. Both alternative cases carry lower loan values in the monitoring region as we would expect, the intuition being that the bank stops the monitoring activity sooner when the cost of monitoring increases or the signal-to-noise ratio decreases. As a consequence, the value of the loans in the monitoring region drops. This effect carries over to the average value of bank loans also, as illustrated by figure 2. Figure 2 shows a three-dimensional (3D) plot of average loan values against variations in the monitoring cost and signal-to-noise ratios. The average loan value increases as the monitoring cost is reduced and/or the signal-to-noise ratio is increased. The main implication of this is that the bank has incentives to make the necessary investments in loan-monitoring technology because it translates into higher loan values, which in turn leads to a competitive advantage relative to other banks.
The second result relates to the effect of regulation when the regulator fixes the target risk level the bank can bear. When the bank invests in improvements in the monitoring technology, the bank will expand the monitoring region. Therefore, the bank will take more loan risk in the intermediate period when loans are monitored and awaiting a risk management decision. This extra risk will be detected by the bank regulator and, as a consequence, the bank regulator will impose higher regulatory costs for the bank. The following assumes that the bank imposes regulatory costs to the point that all banks carry the same risk. To illustrate the effect of this, it may be of interest to study the monitoring region as a function of the cost of monitoring and the signal-to-noise ratio of the monitoring signal.

What happens when the regulator responds to the optimal risk-taking by the bank? As argued above, the amount of risk-taking must be the same across banks with different monitoring technologies. This implies that the marginal opportunity cost of selling a loan of the poorest quality under monitoring must be the same for the bank regardless of the bank’s monitoring technology. If the bank has access to high-quality monitoring systems (cheap monitoring and/or high signal-to-noise ratio of its monitoring efforts), then the bank will monitor more high-quality loans than low-quality loans. The only way this can be done is for the bank to have a high opportunity cost of placing a loan in-house, which means that the regulator imposes high compliance costs of carrying risk. If the bank only has access to low-quality monitoring systems, then the bank will monitor fewer high-quality loans, and the regulator consequently imposes lower regulatory compliance costs. This has knock-on effects on the loan values of the intermediated loans. Figure 3 shows the loan values that correspond to those in figure 1, with the exception that the regulator adjusts the compliance cost such that the bank takes exactly the same level of risk with more costly monitoring and with higher signal-to-noise ratio in their monitoring, as the bank takes in the base case. We see here that more costly monitoring leads to greater loan values and higher signal-to-noise ratio leads to lower loan values for the bank, which is exactly the opposite of the finding in figure 1.

The average value of the bank’s loan portfolio will also reflect the finding in figure 3. Figure 4 shows the average value in a 3D plot against the monitoring cost \( m \) and the drift rate \( \mu \) (for fixed \( \sigma \)), which is a proxy for the signal-to-noise ratio \( \mu/\sigma \), when the regulator imposes strict capital controls to prevent risk-taking. The finding is clear that any improvement in monitoring technology is penalized by the regulator to the extent that the average value of the bank’s loan is reduced. By dismantling its monitoring systems, therefore, the bank can compete more aggressively in the loans market.
What is the relationship between loan sales and loan monitoring? Our model does shed some light on this issue, but it is complicated by the fact that bank regulation may create incentives for banks to sell loans, simply to reduce their exposure to credit risk. If the compliance cost of regulation remains constant (i.e., the parameter K remains constant), the banks that obtain cheaper ex-post monitoring systems (lower cost m) or more effective ex-post monitoring systems (higher signal-to-noise ratio \(\mu/\sigma\)), the bank becomes less inclined to place loans in-house but also less inclined to sell loans to the secondary market. The bank will, however, spend more time monitoring loans that await a final decision about whether to be held or sold. Therefore, we should expect a negative relationship between loan sales and loan monitoring. This picture changes when the bank is subject to active bank regulation, where higher compliance cost is imposed on the banks that have cheaper or more effective monitoring technology, which allows them to monitor more loans of higher credit risk. In this case, the amount of loan sales is constant, and there is no relationship between monitoring activity and loan sales.

4. Discussion

The main conclusion from our analysis in the previous section is that bank regulation can have harmful effects on bank activities that arise from a regulatory trade-off. The regulator can control the risk that arises from lending, but this will be harmful to the stability of the banking system. We discuss the existing evidence that sheds light on the effects of bank regulation. Some of these studies are based on cross-country comparisons of regulation. We should make a note of how difficult it is to make precise inference from such studies. Regulation is only one factor that varies across countries. There are many other factors that also vary systematically across different countries. A direct comparison of the effects of regulation is, therefore, very difficult.

A key part of bank regulation is the value put on the investments that banks make in loan-monitoring systems. The Basel Committee's report on credit risk assessment and loan valuation sets out the basic principles for assessing a bank's management of credit risk. This report asserts that:

"... [A] significant cause of bank failures is poor credit quality and credit risk assessment. Failure to identify and recognise deterioration in credit quality in a timely manner can aggravate and prolong the problem. Thus, inadequate credit risk assessment policies and procedures, which may lead to inadequate and untimely recognition and measurement of loan losses, undermine the usefulness of capital requirements and hamper proper assessment and control of banks' credit risk exposure."18

It goes on, therefore, to formulate principles, including principle 2 that promotes investment in loan monitoring systems: “A bank should have systems in place to reliably classify loans on the basis of credit risk.”19

The credit risk that originates in the banks' lending activities should in turn be followed up by a strategy for making provisions for the credit risk that originates through the banks' lending operations: “Principle 5: A bank's aggregate amount of individual and collectively assessed loan loss provisions should be adequate to absorb estimated credit losses in the loan portfolio.”20

These principles form the core of the link between the investments in loan monitoring systems and the regulatory costs incurred in order to mitigate the risk created by bank...
lending. This link, which is assumed by our model, is therefore, a fundamental part of the current framework for bank regulation.

Despite the clear line set out by the Basel Committee, the evidence that investments in monitoring systems are a priority for banks is not clear-cut. Berger et al.\textsuperscript{21} finds, for instance, that large banks rely more and more on loan screening in the form of credit-scoring systems that are used ex ante when the lending decision is made, and less on monitoring of loans that have already been granted ex-post. There is further casual evidence to support this. Stacy Mitchell\textsuperscript{22} argues, for instance, that though small banks are not subjected to the stringency of bank regulation that the larger banks are, both do not rely on credit-scoring systems and also lend disproportionally to smaller businesses. Moreover, in a recent Financial Times article, Amir Sufi\textsuperscript{23} argues that the ability of the banking sector as a whole to discriminate between good and bad borrowers has been impaired during the financial crisis, and that banks are no longer “savvy information-gathering lenders, but take instead leveraged bets on real estate.” There is also empirical evidence that regulatory compliance costs affect loan pricing. Chen et al.\textsuperscript{24} finds that the banks that are subject to more stringent regulation charge more for their loans.

The results of our paper imply that the Basel framework may be contributing to the disappearance of relationship banking, where the bank establishes a close relationship with the borrower and picks up signals of changes in the borrower’s status, which may be of relevance to the ability to repay the loan. The fact that relationship banking provides early signals of deterioration in loan quality makes it cheaper for the bank to manage the credit risk of the loan, and therefore, the bank is also more inclined to lend in the first place. As such, relationship banking leads in general to a greater accumulation of bank risk. Mester et al.\textsuperscript{25} describes a similar monitoring system for individual borrowers on the basis of monitoring of the borrower’s checking account activity. It is shown that the practice of monitoring checking accounts is not only common, but also that borrowers who can be monitored in this way obtain cheaper finance, which indicates that the early signal of a deterioration of credit risk leads to improved risk management of the loan. The fundamental problem with this kind of monitoring activity is that it leads to greater willingness to take risk ex ante, and therefore, increased bank risk is a by-product of the monitoring activity. At the same time, the bank is under regulatory scrutiny and must pay compliance costs for taking increased bank risk. The shareholders expect, therefore, that increased monitoring activity will lead to greater compliance costs.

\textbf{5. Conclusions}

We have presented the argument that regulation can harm socially beneficial loan-monitoring activity in banks. Our suggestion is that there is a regulatory trade-off between risk and lending efficiency, but that there are solutions to this trade-off. The first, and obvious, solution is to improve the quality of the credit-scoring technology that will not be affected by the regulatory trade-off. Recently, we have witnessed some interesting innovations in this area linked to the use of big data. This includes the use of social network information to improve the credit assessment of personal borrowers. A recent example cited in the SmartData Collective is the use of loan applicants’ job history and their connections on LinkedIn in determining whether to grant loans.\textsuperscript{26} A problem is that innovations of this type will not improve naturally the credit assessment of corporate borrowers where, in particular, young and small firms can be challenging in terms of credit risk assessment, because there is a lack of history of any kind. A second problem is that once it is known that untraditional data is being used by credit-scoring systems, they may become subject to “sanitization” by the prospective borrower with the aim of maximizing the chances of getting through the credit-scoring process. Either way, credit scoring will have its limits.

The second solution is to adapt bank regulation to the trade-off. This solution builds on bank regulators being more tolerant to bank risk that is generated through ex post monitoring of loans than to bank risk that is generated from other bank activities. This requires a regulatory framework that emphasizes the monitoring technology available to the banks as well as the credit risk that arises through the banks’ credit risk. The entire bank risk should be controlled by controlling the risk that arises from non-lending activities. This approach will avoid the downside associated with increased total risk and will also capture the upside associated with less stringent regulation of lending activities.

\textsuperscript{21} Berger et al. (2005)\textsuperscript{22} Stacy Mitchell (2010)\textsuperscript{23} Amir Sufi (2014)\textsuperscript{24} Chen et al. (2000)\textsuperscript{25} Mester et al. (2007)\textsuperscript{26} Srinivasan (2015)
A couple of wider points should be mentioned. This paper is primarily about the loan-monitoring activity of borrowers. Banks spend resources on assessing their lending at portfolio level to analyze whether they have the right balance in their lending policy, for instance, to avoid concentration of lending to certain segments of the loan market.

This type of monitoring activity can be subject to criticism that it is likely to increase portfolio risk in the banks' lending operations. Banks that have good systems to analyze and make adjustments to their loan portfolio ex-post will be more willing to accept risk at the portfolio level ex ante. Therefore, by committing not to take such risks, the bank may simply restrict its ability to monitor lending at the portfolio level.

A second point relates to a potential non-monotonic relationship between the stringency of bank regulation and the credit risk of the banks' lending operations. Consider the following, which relies on the link between monitoring activity in the banking system and the real economy. Less stringent bank regulation can lead to greater investment in monitoring systems, which leads to more efficient lending and higher economic growth. This, in turn, reduces the credit risk of the banks' initial lending operations. Therefore, lenient bank regulation can potentially reduce bank risk. Conversely, more stringent bank regulation can prevent banks from investing in monitoring technology, which leads to less efficient lending and lower economic growth. More stringent bank regulation can increase the credit risk of the banks' loan portfolio. Neither point has been rigorously studied to our knowledge. However, we reiterate that our point in this paper is not that banks should be subject to less stringent bank regulation in general, rather that the regulator should rebalance the stringency of bank regulation to take the social value of bank risk into account.
References


Sufi, Amir, 2014. “Bernanke’s failed mortgage application exposes the flaw in banking,” Financial Times, October 13. Available at: https://next.ft.com/content/b02a24c2-4d60-11e4-bf60-00144feab7de
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.

EY is a leader in serving the financial services industry
We understand the importance of asking great questions. It's how you innovate, transform and achieve a better working world. One that benefits our clients, our people and our communities. Finance fuels our lives. No other sector can touch so many people or shape so many futures. That's why globally we employ 26,000 people who focus on financial services and nothing else. Our connected financial services teams are dedicated to providing assurance, tax, transaction and advisory services to the banking and capital markets, insurance, and wealth and asset management sectors. It's our global connectivity and local knowledge that ensures we deliver the insights and quality services to help build trust and confidence in the capital markets and in economies the world over. By connecting people with the right mix of knowledge and insight, we are able to ask great questions. The better the question... The better the answer... The better the world works...

© 2018 EYGM Limited. All Rights Reserved.

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