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

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

gfsi.ey.com
The Journal of
Financial Perspectives

Part 1: Strategic

13 Narratives of the Great Financial Crisis (GFC): Why I am out of step
Charles A. E. Goodhart, Emeritus Professor, Financial Markets Group, London School of Economics and Political Science

19 Intelligent ERM: Evolving risk management
David Baxter, Head of ERM Enabling Programme, Swiss Re
Phil Vermeulen, Partner, EY LLP

29 The shifting terrain of risk and uncertainty on the liability insurance field
Tom Baker, William Maul Measey Professor of Law, University of Pennsylvania Law School

39 Principles and policies for in-house asset management
Gordon L. Clark, Halford Mackinder Professor of Geography, and Director of the Smith School of Enterprise and the Environment, Oxford University
Ashby H. B. Monk, Executive Director, Stanford University’s Global Projects Center, Stanford University

49 Why foreign life insurers did not achieve their ambitions in China: Structural and operational obstacles
Maoqi Wang, School of Economics and Management, Tsinghua University
Bingzheng Chen, Professor of Finance, School of Economics and Management, Tsinghua University
Joseph Qiu, Senior Portfolio Manager, Global Property, AIG

61 Financial policymaking in the context of the known, the unknown and the unknowable
Richard J. Herring, Jacob Safra Professor of International Banking, The Wharton School, University of Pennsylvania

Part 2: Tactical

79 Japanese financial institutions expanding abroad: Opportunities and risks
W. Raphael Lam, International Monetary Fund

93 Safe to fail
Thomas F. Huertas, Partner, EY LLP

111 Market risk of real estate: Using indirect data to understand direct risks
Felix Schlumpf, Head of Strategic Asset Allocation, Investment Management, Zurich Insurance Company Ltd.
Genene Tessera, Senior Quantitative Analyst, Investment Management, Zurich Insurance Company Ltd.
Catalina Martínez, Research Assistant, Investment Management, Zurich Insurance Company Ltd.

121 Insurance risk transfer and categorization of reinsurance contracts
Alexander Itigin, Consultant, World Bank, and Senior Actuary at a global reinsurance company
Renate Wiechert, Consultant, World Bank, and Accounting Policy Consultant at a global reinsurance company

143 Superior Information and compensation fees of active mutual funds
Chekib Ezzili, Equity Derivatives, NATIXIS
Patrice Poncet, Distinguished Professor of Finance, ESSEC Business School
**Narratives of the Great Financial Crisis (GFC): Why I am out of step**

by Charles A. E. Goodhart, Emeritus Professor, Financial Markets Group, London School of Economics and Political Science

Economics is not a hard science, subject to controlled experiments. In their absence, economists construct narratives to explain events, often involving key sound bites. Among the sound bites of the recent financial crisis are the following: “separate the casino from the utility bank”; “protect the taxpayer; bail-in the bank creditor”; “moral hazard is ever present”; and “liquidate zombie banks and zombie borrowers.” This paper examines the accuracy of each of these sound bites and argues that they give a misleading picture of a far more complex reality. It suggests that most of the reason that the blame for the recent crisis was placed squarely at the doorstep of the investment bankers is that they are highly paid, which, in turn, has blinded many of the commentators on the real causes. The crisis was brought about by a massive and synchronized, but otherwise standard, retail bank-led, property bubble and bust. The author proposes that public funds should in fact be made available to rescue investment banks, which are significantly more interconnected and more difficult to separate into good and bad banks than retail banks.

**Intelligent ERM: Evolving risk management**

by David Baxter, Head of ERM Enabling Programme, Swiss Re, and Phil Vermeulen, Partner, EY LLP

Enterprise Risk Management (ERM) is a familiar concept in management theory. However, sometimes its implementation can fall short of delivery on the potential benefits to the business. The need for effective ERM is becoming increasingly evident for companies that want to ensure that their practices are agile enough to respond intuitively to today’s environment. Implementing an enhanced, more intelligent risk management capability is exactly what global reinsurer, Swiss Re, has looked to do in a current and ongoing phase of work conducted in collaboration with EY. This paper presents some of the initial findings from this program of work, some of the areas where enhancements to the existing risk management practice have added most value, as well as some of the more challenging aspects of implementing what is effectively a long-term strategy of cultural change across the business.
The shifting terrain of risk and uncertainty on the liability insurance field
by Tom Baker, William Maul Measey Professor of Law, University of Pennsylvania Law School
This paper looks at the limitations of quantitative models for evaluating risks in the liability insurance business, and suggests that insurance risks very often are not reliably calculable except in hindsight, at which point the risk has already been transformed into an all-too-measurable loss. Insurance is an “uncertain business,” characterized by competition for premiums that pushes insurers into the unknown. Even life insurance, the part of the business that manages the oldest and best-understood insurance risk — mortality — operates just beyond the limits of knowledge. The paper also investigates whether liability insurance premiums in the U.S. market have grown at a much higher rate than, say, property insurance premiums, which many attribute to the highly litigious U.S. market, and finds that in fact both have grown by about the same percentage since the 1960s. The reason that many commentators believe that liability insurance premiums have increased disproportionately is because they do increase by a greater amount during hard market periods, which is when the press seems to be paying more attention. These findings suggests that if liability insurance premiums simply kept up with inflation, rather than fluctuating during hard and soft markets, insurance companies would not come under the kinds of attacks in the media that we have all become accustomed to.

Principles and policies for in-house asset management
by Gordon L. Clark, Halford Mackinder Professor of Geography, and Director of the Smith School of Enterprise and the Environment, Oxford University, and Ashby H. B. Monk, Executive Director, Stanford University's Global Projects Center, Stanford University
Spurred on by the recent financial crises, a growing number of institutional investors are working to bypass traditional financial intermediaries, agents and centers through the development of in-house teams of investment professionals. As such, the institutional investment community, which is often characterized by broadly diversified and outsourced organizations, is becoming much more involved in the day-to-day asset management of their portfolios. Research shows that this new path offers a variety of important benefits, including higher net-of-fee returns. And yet, there remain significant pitfalls as well. This paper outlines the challenges facing would-be “in-sourcers” and offers a series of principles and policies for effective in-house asset management. Drawing on 20 case studies, it concludes that successful in-house asset management requires the institution of many of the systems and processes, including significantly higher compensation than is prevalent today, that third-party managers have had to develop over the years, and that the plan sponsors should only consider managing those assets in-house that can be done more efficiently and/or cost-effectively.
Why foreign life insurers did not achieve their ambitions in China: Structural and operational obstacles

by Maoqi Wang, School of Economics and Management, Tsinghua University, Bingzheng Chen, Professor of Finance, School of Economics and Management, Tsinghua University, and Joseph Qiu, Senior Portfolio Manager, Global Property, AIG

China's economy has experienced rapid growth in the past 30 years. A highlight is in the growth of its insurance industry, whose premiums increased at an annual rate of around 30%. Naturally, people would think that the foreign insurance companies that entered the Chinese market during this period should have achieved impressive performances, especially given that they are competing with the domestic insurance companies that do not have the advantage of experience and technology. However, the reality is that foreign insurers, especially foreign life insurers, often found it difficult to expand their market shares in this emerging market. Many foreign life insurers experienced frequent management turnover and some even withdrew from the Chinese market. This paper aims to analyze this interesting market phenomenon from two different perspectives: (1) structural obstacles, including the regulatory environment and the influence of the recent financial crisis and (2) operational issues faced by foreign life insurers, including operational efficiency, shareholding structure and distribution channels. On the basis of these analyses, the paper offers several suggestions to market participants structured around the need for foreign life insurers operating in China to employ a customized approach. The unique regulatory environment and market characteristics, such as the importance of bancassurance, must be considered when forming business strategies for the Chinese market. These can be achieved by accumulating more local knowledge, employing more experienced Chinese managers, showing the willingness to deviate from the standard business model, and finally being patient because the regulatory environment and market conditions will not change in one day.

Financial policymaking in the context of the known, the unknown and the unknowable

by Richard J. Herring, Jacob Safra Professor of International Banking, The Wharton School, University of Pennsylvania

This paper examines the challenge of financial policymaking from the perspectives of what is known, unknown and unknowable about the financial system. It focuses on policy challenges during the Great Recession and changes in regulations in response to lessons drawn from the global financial crisis. The paper suggests that while improving the effectiveness of prudential policy should be an important priority, it requires an honest assessment of what we actually know about the risks assumed by systemically important institutions and, indeed, what the managers themselves know. Since many of these risks are unknown and some are unknowable, supervisors need to place much greater emphasis on increasing the resilience of the system. Rather than imposing increasingly prescriptive regulations, a wiser course may be to devise simpler rules that can be defined and monitored more easily.
Japanese financial institutions expanding abroad: Opportunities and risks
by W. Raphael Lam, International Monetary Fund
Overseas activities of Japanese financial institutions have risen, mainly in Asia, since the global financial crisis. Stagnant growth and low interest margins in Japan have added to incentives to seek opportunities abroad. This paper explores the determinants of Japanese banks’ overseas expansion and assesses whether these cross-border activities will continue under the new macroeconomic policies often referred as “Abenomics.” The analysis finds that Japanese banks are well positioned to scale up foreign exposures, thanks to their relative resilient balance sheets and the robust growth in the region. Stronger domestic growth in Japan could mitigate the pace, but is unlikely to reverse a long-standing trend because empirical estimates suggest that global and regional factors play a more prominent role in the growth of cross-border claims. The increasing cross-border activity would pose funding risks and supervisory challenges that require continuous close monitoring. An incomplete set of domestic policies that fails to raise growth and exit deflation could, however, undermine the prospects of Japanese banks in expanding abroad.

Safe to fail
by Thomas F. Huertas, Partner, EY LLP
Banks cannot be made fail-safe. But they can be made safe to fail, so that the failure of a bank need not disrupt the economy at large nor pose cost to the taxpayer. In other words, banks can be made resolvable, and “too big to fail” can come to an end. To do so, the authorities, banks and financial market infrastructures need to prepare in advance for what amounts to a pre-pack reorganization of the bank that the resolution authority can implement over a weekend, if the bank reaches the point of non-viability in private markets (fails to meet threshold conditions). This pre-pack consists of two principal elements: (i) a recapitalization of the bank through the bail-in of investor instruments and (ii) the provision of liquidity to the bank-in-resolution. Creating such a pre-pack solution should form the core of the resolution plans that authorities are developing for global systemically important financial institutions. This paper sets out the conditions that must be met for a bank to be resolvable, the “safe-to-fail” test and the banking structures required in order to meet this test. How banks are organized matters less than what banks, authorities and financial market infrastructures do to prepare for the possibility that resolution may be required. The paper concludes with an agenda for action to ensure that too big to fail is not too tough to solve.
Market risk of real estate: Using indirect data to understand direct risks

Even if the market capitalization of direct real estate is comparable to that of equities and fixed income, the data on direct real estate is very poor. It is, therefore, difficult to estimate the market risk of this important asset class. Moreover, risk systems from most vendors cover equities and fixed income, but do not cover direct real estate. This paper proposes a simple methodology that uses widely available data on indirect real estate to estimate the market risk of direct real estate. In particular, we use data on Real Estate Investment Trusts (REITs) returns, determine their factor exposures to other asset classes and deleverage these exposures according to REITs' balance sheets. The paper shows that direct real estate can be considered as a portfolio of equities, fixed income and credit, combined with idiosyncratic risk. The authors find that the existing direct indices understate the risk of the real estate market. In addition, with the proposed methodology, the correlations to other asset classes become materially different and higher.

Insurance risk transfer and categorization of reinsurance contracts

Despite the existence of numerous quantitative approaches to the categorization of financial reinsurance contracts, insurance regulators may often find the practical implementation of the task to be technically challenging. This paper develops a simple, affordable and robust regulatory method that can help insurance regulators categorize whether financial reinsurance contracts classify as reinsurance. By reviewing real examples of different categorization methods, this paper explains how the proposed method, the Standardized Expected Reinsurer's Deficit (SERD), standardizes such categorization. It also summarizes the existing pertinent literature on the subject with the view to helping insurance regulators to first apply some simple indicators to flag the main issues with financial reinsurance contracts that may need further reviews.
Superior information and compensation fees of active mutual funds
by Chekib Ezzili, Equity Derivatives, NATIXIS, and Patrice Poncet, Distinguished Professor of Finance, ESSEC Business School

This paper looks at the cost of information for a client on using the knowledge of an asset management company. The manager possesses more information than the investor and charges for access to such information. How much the investor is willing to pay depends on the benefits that can be derived from such information. The investor will delegate portfolio decisions to the manager if and only if the expected utility of their wealth after fees is larger than the expected utility they can achieve by directly investing in the market. This paper allows us to characterize compensation fees charged by asset management companies in terms of information differential. The greater the difference, the higher the fees that clients are willing to pay. Hence, success for the managers depends on their ability to possess information that is superior to their peers.
Part 1: Strategic

Narratives of the Great Financial Crisis (GFC): Why I am out of step

Intelligent ERM: Evolving risk management

The shifting terrain of risk and uncertainty on the liability insurance field

Principles and policies for in-house asset management

Why foreign life insurers did not achieve their ambitions in China: Structural and operational obstacles

Financial policymaking in the context of the known, the unknown and the unknowable
Narratives of the Great Financial Crisis (GFC): Why I am out of step

Charles A. E. Goodhart
Emeritus Professor, Financial Markets Group, London School of Economics and Political Science

Abstract
Economics is not a hard science, subject to controlled experiments. In their absence, we economists construct narratives to explain events, often involving key sound bites. Among the sound bites of the financial crisis are the following:

- “Separate the casino from the utility bank”
- “Protect the taxpayer; bail-in the bank creditor”
- “Moral hazard is ever present”
- “Liquidate zombie banks and zombie borrowers”

This paper argues that each of these sound bites give a misleading picture of a far more complex reality.
Narratives of the Great Financial Crisis (GFC): Why I am out of step

The importance of narrative
In macroeconomics, we cannot undertake controlled experiments. It would be nice to do so. I would like to take a group, such as the readers of this paper, and reduce the money holdings of one half of the group by, say, £1,000, and increase the money holdings of the other part of the group by the same amount, and then see what happens in the next period. Perhaps fortunately for the people on whom such experiments might be unleashed, we economists are not allowed to do this. However, nature, politicians and surprise events unleash uncontrolled experiments on us from time to time; for example, the application of austerity to almost all economies in the Eurozone. But, unlike inanimate natural objects, such as atoms, we, individuals, learn and adapt in response to such uncontrolled experiments as are imposed upon us. It is, therefore, not necessarily true that the result of the first uncontrolled experiment will be like the outcome of the second such experiment as may be carried out, since we all will have learned in the meantime, and therefore our responses may differ, even when the initial shock is exactly the same in the two cases.

Hence, macroeconomics is not a hard science, in the sense that the natural sciences, such as physics and chemistry, are. Even so, we have to try to make sense of our macro context, with a view to trying to improve the welfare of everyone within the system. So, what we do, in the main, is to construct narratives of what has been happening, based more or less firmly on our existing theory and related to a set of empirical observations from the surrounding events. But both our theories and the available set of empirical observations are partial and limited. In this process of trying to make sense of what is happening, by some combination of theory and empirical observation, story-telling skills are valuable; especially so the ability to coin the memorable sound bite.

But I have become unhappy, out of step, with the dominant narratives and the sound bites that have developed and are commonplace about the causes of, and the cures for, the Great Financial Crisis (GFC).

“Separate the casino from the utility bank”
The idea put forward in this sound bite is that investment banking is largely gambling, undertaking proprietary deals for largely personal gain in exotic instruments (with generally lower ethics and worse culture), while retail banking provides services for the real economy. The suggestion is also made that it was the investment banking arms of (universal) banks that got us into trouble; and that only the retail part of banks is essential for the system to continue, and should therefore be the only segment of a bank to have taxpayer support in the case of extreme difficulties.

As an example, on the day when Fabrice Tourre, a trader at Goldman Sachs, was coming for trial on the grounds that he had fraudulently sold, or mis-sold, an exotic instrument that had been designed by a hedge fund trader taking the opposite position, in order to fail, one of the newspapers reported this with the following phrase: “This trial will show us whether the system was destroyed by the banks as a whole, or only by rogue traders within the banks.”

Most, if not all of this, is wrong. There is not space, or time, within this paper to talk at length about the real and valuable functions of investment banking. Suffice it to say, in the light of the tapes of the activities of the bankers at the Anglo-Irish bank, the idea that the ethics and culture of retail bankers is on a higher plane than those of investment bankers is at least questionable.

Rather more important, the GFC did not arise from dealings in exotic derivative instruments. It was, instead, driven by the expansion of credit to finance a property bubble, both in commercial real estate (CRE) and in housing mortgages. Indeed, the derivatives book of Lehman Brothers was actually profitable; what drove Lehman’s into bankruptcy was rather their stock (warehouse) of mortgage-based securities (MBS). Fuld, the CEO of Lehman Bros, had not expected housing prices and the valuation of MBS to fall as far as they did; he thought that purchasing further MBS at a time when these were showing price weakness would actually be profitable for his company.

What securitization, e.g., of such MBS, did was to spread the subprime crisis outside of the U.S. into Europe, via regulatory arbitrage. The capital requirement constraint that bit into the U.S. was the simple leverage ratio, since they had not adopted Basel II in detail. Whereas the capital requirement constraint that bit into Europe was the Basel II capital requirement (CAR), which was based on risk-weighted assets (RWA). Securitization allowed mortgages, especially the lower-rated mortgages, such as subprime, to be reconstructed into tranches with differing
exposures to default, and differing credit ratings. The higher quality tranches had a much higher credit rating, AAA, and could, therefore, be held in very large quantities by European banks without needing RWA capital involvement. Meanwhile, the American banks could hold the lower-quality tranches, i.e., the Mezzanine and equity (toxic) tranches, since the higher RWA did not have any effect on leverage. Thus, the European banks frequently had a leverage ratio of 40:1 or 50:1 in 2007, whereas the American banks had a much lower leverage ratio, perhaps 30:1 or lower, while the average risk weighting of the assets held by the American banks was considerably higher than that of the European banks.

It was the (botched) failure of Lehman Brothers, a pure investment bank, that caused the crisis. This was partly because the American authorities were careful to arrange the process of liquidation to allow the American parts of Lehman Brothers to be in a condition where they could continue to function. In contrast, the American liquidators paid virtually no attention to what might happen to Lehman International Europe bank, which was in London, or Lehman's in Japan. It was the chaotic failure of Lehman subsidiaries abroad that did a great deal of the damage. When the British liquidator, Tony Lomas, went into Lehman Bros in London on the Monday morning, there was no money, virtually no continuing utilities, such as electricity, and all the staff had been fired, including the IT staff, so he had no way of finding out where or what information might be stored.

Anyhow, investment banks are much more interconnected than retail banks, so the latter are much easier to liquidate, i.e., to separate into good/bad banks, and to sell the good bank together with the retail deposits to some other bank, or even to re-establish it and start it up as a public-sector, nationalized bank. The likely externalities and spillovers are certainly going to be much greater in the aftermath of the failure of an investment bank than that of a retail bank. So, the economic arguments would imply that if there are to be public sector, taxpayer bailouts, these should be of investment banks, not of retail banks. It is the political arguments that direct bailouts to the retail bank.

Meanwhile, the separation between investment banking and retail banking that has been proposed by the Independent Commission on Banking (the Vickers Commission), and has been widely applauded by most commentators, will lead to less diversification, and to an even greater focus of retail banks on risky property-related lending (CRE) and residential mortgages. Most of the banking crises, at least in developed countries, in the last 30 years have been related to excessive leverage in pursuit of property lending. The Glass/Steagall division will make the next property-lending boom/bust cycle even more likely, probably to arrive in about 15 years.

There is a common fallacy that banks take deposits from the personal sector in order to lend to the company sector. In fact, as Adair Turner (2010) has shown, whereas most deposits still come from the personal sector, the majority of lending is also to the personal sector, mainly for house purchase. Furthermore, such lending as there is to the company sector is usually also either property related, to construction or property developing companies, or collateralized on the basis of property. A large proportion of SME lending is collateralized in this way.

Hence, this separation is not likely to make the banking system safer; indeed it might actually exaggerate the risks within the system. For example, it will remove stable deposit funding from investment banks, and make them even more reliant on wholesale funds from informed counterparties — often other banks themselves that are more likely to run at the very first sign of fragility. Thus, the asset side of retail banks will become more concentrated and riskier, while the liability side of investment banks is also likely to become subject to greater risk and potentiality for runs.

Several economic commentators have recognized that the reforms, so far, have done relatively little to diminish the inherent weaknesses of the banking system. Some of them, such as Larry Kotlikoff and Michael Kumhof, have argued for further stronger restraints on the functions and structure of banks; by, for example, limiting the availability of transaction balances to narrow banks and requiring all risky assets to be held either on the basis of long-term fixed-interest liabilities or of equity. Although the logic of their position is understandable, their proposals are so radical and restrictive that they are unlikely to be acceptable, at least for the time being. There are, in my view, better ways to make the banking and financial systems safer, but this is not the place to go into such alternative proposals.
“Protect the taxpayer; bail-in the bank creditor”
By far the largest cost to the taxpayer, over the course of the GFC, arose not from the use of taxpayer funds to recapitalize a failing bank, but from the failure to do so, i.e., in the case of Lehman Brothers. The American authorities had had a forewarning, in the case of Bear Stearns, in the previous spring, of what might happen should a large investment bank get into difficulties. At that time, they had negotiated a merger of Bear Stearns with JPMorgan Chase. But they should have used the available intervening time to work out a plan B, should, in the case of another bank, such as Lehman’s, there might be no other bank prepared or willing to accept a merger with the failing bank. This failure to prepare an alternative plan B had calamitous results.

When taxpayer funds have been sensibly used to recapitalize banks, they have not only prevented disaster but have even often ultimately resulted in an accounting profit, as has been the case with TARP in the U.S. and in the earlier Scandinavian financial crisis in Sweden in 1991.

But such recapitalization needs to be done swiftly and decisively. This was not the case with the Royal Bank of Scotland (RBS). For political reasons, the then Labour Government did not want to nationalize RBS, although it then injected a vast majority of the ownership equity. As a result, it has never been entirely clear whether RBS should be run as a public-sector bank or as a private-sector bank. Its objectives were mixed and uncertain. It has, therefore, been much more difficult to manage and to be clear how it should be reorganized.

Moreover, without a proper funding backstop, stress-tests (for example, the Asset Quality Review, which the ECB is supposed to be carrying out next year) are castrated. Without such a backstop for recapitalization, any bank whose capital falls below the required limit is likely to face much higher funding costs, and even perhaps a run. So, there is pressure on those conducting the testing to show results that indicate that no bank at all has missed their requirements. But in that case, the audience will not believe the results of the stress tests, and will try to work out for themselves which banks they believe to be fragile.

The current intellectual fad is to propose that there should be a shift from taxpayer bailout of any failing banks to creditor bail-in. In my view, this would be a categorical error, though I realize that I am in the minority in this respect. Let me explain the reasons for this view.

1. The burden of loss from the failure is in no way going to be reduced. It will fall on a more concentrated group (of creditors), rather than on a more widely dispersed group (of taxpayers).
2. There will be exemptions of favored creditors, perhaps often extensive. In the case of Cyprus, the haircut on uninsured deposits was withdrawn, i.e., not imposed, in the case of charities and public-sector institutions, such as hospitals and municipalities. The meeting of the Eurogroup finance ministers proposed that there should be no such penalty on either persons or SMEs holding such large uninsured deposits. This means that the burden is likely to fall, almost entirely, on holdings of such assets by other financial intermediaries (banks will not be allowed to hold large amounts of such claims on other banks). Why it should be preferable to penalize investors in pension funds and in insurance companies, rather than taxpayers more generally, is far from clear.
3. Unfavored creditors can, and will, flee. Taxpayers cannot. The relative burden on taxpayers has been democratically reviewed; not so on creditors. There will always be hard cases, as indeed illustrated in the recent case of the Co-op Bank.
4. In order to ensure that there are sufficient creditors, so as to bail them in case of difficulties, banks will have to be forced to hold a sufficient proportion of “bail-inable” liabilities. This will raise funding costs throughout, but could raise the funding costs of weaker banks dramatically. The whole exercise of proposing creditor bail-in is likely to enhance the procyclicality of the system as a whole. Funding costs for weak banks, and banks in fragile countries, will rise much more steeply and much more rapidly during periods of pressure, under a system of bail-in, than when there was some expectation of taxpayer bailout. The exercise will exacerbate the depth and length of downturns and busts. In particular, the failure of the Banking Union, in the sense that there is now much less support likely either from the taxpayer or from banks in stronger countries, for those in weaker countries, is likely to mean that the weakness and increasing unemployment of the Southern Mediterranean states in Europe is likely to be prolonged and extended. The failure of the Banking Union means that the doom loop remains intact. Creditor bail-in is a way of continuing to focus on the weakness of the Eurozone peripheral countries within those same countries.
“Moral hazard is ever present”

But would not continuing reliance on taxpayer bailout encourage further risk-taking, i.e., lead to moral hazard? Yes, in principle, but that actually did not happen in 2008, as several articles [Fahlenbrach and Stulz (2009); Foote et al. (2012)] have shown. The bankers, again for example Dick Fuld, did not believe that they were taking risks by investing in mortgage-backed securities. Nor did the credit rating agencies, and nor did the markets, where CDS values had fallen to their lowest level, implying virtually no risk of default, in the early summer of 2007. For the previous 50 years in the U.S., if you held a regionally well-diversified portfolio of MBS, there was very little risk of significant falls in values, because housing prices, diversified over the U.S. as a whole, had never fallen significantly during that period. If you ran an econometric model, using those same 50 years, the expectation that housing prices might fall by more than 5%, on such a regionally diversified portfolio, was estimated to be a standard deviation of above 2, i.e., a most unlikely event.

Moreover, moral hazard should relate to the individual taking the decisions, i.e., to the banker, not to the institution, i.e., the bank. Liquidating the bank would have little effect on the decisions of the banker should the bankers’ expected pay-offs, i.e., his remuneration, remain the same whether the bank was liquidated or not. Moral hazard would be reduced by making the banker, not the bank, suffer from failure, for example by removing some part, or all, of the limited liability protection for top decision-making bankers.

But even so, if banks, particularly the large investment banks, were always to be supported by taxpayer bailout, would that not give larger banks an unfair advantage, and remove interest rate signals about their impending fragility? But why not rely on equity signals, because no one is proposing to remove the potential loss to equity holders, and then also require much more equity, as advocated by Admati and Hellwig (2013). Indeed, a banking system where banks were required to hold much more equity, relative to debt, would be much safer; and in other respects would have few disadvantages. The real issue, however, is not that the equilibrium state would be much preferable, with a much higher share of equity, but how to get from here to there. The problem is one of achieving dynamic transition from a low-equity to a high-equity equilibrium state, rather than residing in a comparison of the two states.

“Liquidate zombie banks and zombie borrowers”

Raising required equity ratios, without having a means of forcibly recapitalizing such banks (e.g., via taxpayers), at the same time as bankers continue to focus on the return on equity (RoE), will just lead to undercapitalized banks and deleveraging. Particularly at a time when banks’ market equity prices are low relative to book values, raising new equity will dilute existing shareholder value and transfer value from equity holders to bondholders. Hence, bank CEOs and other managers are only going to raise new equity when they have absolutely no other alternative.

Without having a means to force banks to raise new equity and not having the funds to recapitalize them themselves, regulators are having to connive, and to wink, at the continued existence of banks whose equity holdings are well below the higher ratios that are now generally desired. In effect, regulators themselves are practicing voodoo in order to bring zombie banks into (dismal) existence. There is increasing covert domestic political protection for each country’s banks. The banks in each country are being pressured to reduce leverage by cutting back on lending abroad or by selling assets situated abroad. The implicit agreement is that if domestic banks are prepared to go on increasing lending, either to the government or to the private sector, within each country, then the government and the monetary authorities there will do their best to protect them from failure or other outside interference.

But in any case, liquidation of such banks would just have made everything worse. In such a context, foreign banks are unlikely to come in to pick up the pieces, and in most cases, the other domestic banks are incapable of doing so. The forced sale of assets would have driven down asset values more generally. Why would any of this have helped other banks? Or the domestic economy?

If we assume, as surely we must, that the GFC is finite and temporary, then borrowers, like banks, will recover at some stage. Forbearance, at least up to a point, is good, not bad. In the U.K., foreclosures, for example on housing, and company failures have been comparatively low during the course of the depression. Why crystallize losses when you do not have to do so? By doing so, it drives down asset values, raises unemployment, and worsens the debt and extent of the depression. A problem of securitization in the U.S. was that the securitized mortgages were so comingleq that it was almost impossible to undertake forbearance on an.
individual (securitized) mortgage. Consequently, it was very difficult to find any alternative to foreclosure; though had as much initiative and thought gone into the reform of housing finance as went into the reform of banking finance, the outcome in the U.S. might have been better. Anyhow, this problem did not match, fortunately, in the U.K. In my view, the extent of forbearance shown by U.K. banks to U.K. borrowers in temporary difficulties has been one of the unsung, and unrecognized, success stories of this recent depression.

Why be any more brutal than you have to be, either in fiscal austerity or in driving borrowers into bankruptcy? What, to my mind is remarkable, is how many of those who, with considerable justification, believed that the degree of fiscal austerity in Europe was excessive, at the same time believe that it would have been better had banks fully recognized the current declines in values and foreclosed on those borrowers who were currently in difficulties. Of course, if a bank believes that the borrower can never recover, even when the economy does, then it will, and should, consider how best to recognize such failure in an orderly fashion. But with the recovery, both borrowers and banks are likely, one must hope, to do better, and eventually to manage to meet their debts. Premature foreclosure, like premature austerity, does not do any good to the economy more widely.

Conclusion
If investment bankers had been paid at roughly the same rate as academic professors or leading journalists, the discussion on the causes of the GFC would have been much more rational and balanced. An, entirely understandable, fury at such “obscene” remuneration has, however, led to distortions in such narratives, with the blame game zeroing in on that target. This has blinded too many people to the underlying cause of the GFC, which has been a massive and synchronized, but otherwise bog-standard, retail bank-led property bubble and bust. Those who cannot learn from history will be doomed to repeat it. Roll on the next property bubble and bust in the late 2020s.

References
Intelligent ERM: Evolving risk management

David Baxter
Head of ERM Enabling Programme, Swiss Re

Phil Vermeulen
Partner, EY LLP

Abstract
Enterprise Risk Management (ERM) is a familiar concept in management theory. However, sometimes its implementation can fall short of delivery on the potential benefits to the business. The need for effective ERM is becoming increasingly evident for companies that want to ensure that their practices are agile enough to respond intuitively to today’s environment. Implementing an enhanced, more intelligent risk management capability is exactly what global reinsurer, Swiss Re, has looked to do in a current and ongoing phase of work conducted in collaboration with EY. In this paper, we will present some of the initial findings from this program of work, some of the areas where enhancements to the existing risk management practice have added most value, as well as some of the more challenging aspects of implementing what is effectively a long-term strategy of cultural change across the business.

1 The opinions expressed in this article are the authors’ personal views and do not in any way represent the views of Swiss Re, EY LLP, their partners, or any associated organizations. Any remaining errors are the authors’ responsibility.
Introduction
This article presents a discussion on the lessons learned to date in an ongoing phase of work at Swiss Re, where EY has been working with the risk management function to embed an enhanced risk management strategy. This program of work was initiated in response to a business need to refresh and revisit the risk management plans and processes that had been developed over many years: the aim was not to fix something that was broken, but to identify improvements and enhancements that would increase efficiency and effectiveness, be sustainable in the long term and demonstrate a commitment to a leading risk management culture.

This paper is organized as follows. In the next section, we will present a high-level overview of Swiss Re’s vision. In the following section, we will look in more detail at the different areas of enhancement that were identified at the start of this ongoing piece of work – many of these are still evolving in the business. The following highlights some insights and lessons learned from the current experience of embedding change in the risk operating model of Swiss Re, looking at specific operational areas across organization, financial management, process, people, data management and IT systems.2 Finally, we provide an overview of how businesses that want to implement a similar process can take the initial steps to enhance their own risk management practice.

Swiss Re’s vision for “intelligent enterprise risk management” (IERM)
As Figure 1 shows, Swiss Re’s vision for a leading edge IERM framework to encourage a more adaptive and intuitive approach to risk is based on three levels:

- Intelligent protection: Swiss Re wanted to enable a more “intelligent” means of protecting the business: through understanding planned risks, supporting the taking of good risks and top-down management of the biggest threats, both known and emerging.
- Trust: the program aims to increase the level of trust in Swiss Re’s risk management capabilities, both externally and between different functions and team members within the organization.
- Competitive advantage: shifting the risk culture from a strategy of “blanket” protection or zero tolerance to one that encourages educated and transparent risk-taking combined with enhancements to risk management capabilities would ultimately enable better commercial decisions, without incurring unsustainable cost, thus improving competitive advantage.

To encourage trust, a large part of this initiative has been about driving transparency on governance and roles. To enhance intelligent protection, part of the strategy has been to standardize and industrialize processes where possible, and introducing new systems where appropriate to support risk activity effectively and efficiently.

In the next section, we will look at the different improvement areas that were identified to help embed and encourage this enhanced risk practice and culture. For each of these areas, we will discuss where value has already been added and where challenges have been faced. This program of enhancement is a work in progress and, as such, this paper presents a snapshot of an evolving process.

Focus areas for enhancement of risk management capabilities
Risk management is a core competency for Swiss Re and, therefore, the necessary “building blocks” for IERM were already in place in terms of risk appetite, limits, governance, models and reporting. When looking to choose focus areas at the start of this process, Swiss Re wanted to leverage their existing framework, identifying enhancements only where they would be most valuable.

After conducting interviews across the business and conducting an enterprise-wide assessment of their existing risk processes, the following areas of enhancement were identified, which we have grouped for ease of explanation into “seven E’s of Intelligent Enterprise Risk Management.” This is not to suggest that these enhancements were made in isolation from one another – each of these enhancements work together as part of a wider program of ongoing work:

1. Greater enlightenment: improving forward-looking analysis
2. Greater embeddedness: integrating risk management, capital management and business planning
3. **Greater empowerment:** further enhancing the first line of defense to empower those responsible for taking on risk
4. **Greater effectiveness:** integrating and aligning control activities across functional areas of expertise
5. **Greater entrepreneurialism:** positioning the Risk Management function to provide more value adding advice to the business, both on request and proactively as a valued business partner
6. **Greater efficiency:** improving processes, systems, and communication
7. **Greater engagement:** realigning the risk culture

In the following subsections, we look in more detail at each of these enhancements and how they have worked to date at Swiss Re.³

Greater enlightenment: improving forward-looking analysis
In most companies today, risk identification happens in a number of separate areas that rarely interconnect. For example, board members and senior executives have broad discussions with peers from other companies and industries to formulate views about long-term trends and risks, economic research teams analyze long-term macroeconomic development, while at the same time, traders, market risk managers, and actuaries identify medium-term risks from their analyses of market trends or claim patterns, and process managers and operational risk specialists look for potential control breakdowns or fraud. All of the activities that take place in relation to identifying individual risks have worked to date, but they may fail to detect the “next big thing,” which is often signaled by a number of smaller changes in different areas. Swiss Re wanted to deepen its understanding of complexity, and thereby improving its forward-looking analysis, by taking greater advantage of “networked intelligence” from inside and outside the organization.

To address this, Swiss Re looked to enhance the “top risk” identification process that had been in place for many years by

³ Each of these sections provide an overview of some of the key examples of how these enhancements have worked in practice and, it should be noted, that they do not encompass the entire program of work. Enhancements such as ‘Greater embeddedness’ in particular are too complex to cover in this article and, as such, we only touch upon a few specific examples to highlight the overall concept.
Intelligent ERM: Evolving risk management

bringing together the top-down and bottom-up processes to enable the sharing of risk identification insights from different areas of expertise: once fully embedded, this process will allow different risks to be prioritized, monitored and mitigated, with the information then being fed back for deeper modeling.

Swiss Re has also looked to enhance its existing emerging risk identification (SONAR) process, which brings together internal and external signals of existing and emerging risks. This process has been enhanced to include a deeper analysis of the interconnectivity between these signals in order to develop more complex threat scenarios and prioritize risks. The signals are analyzed, grouped and prioritized into key emerging risks. Each emerging risk is sponsored by a senior business owner, and is monitored through a formal review process with a range of stakeholders in the business. To ensure that these risks receive sufficient focus, they are also discussed at the most senior levels of the company. Swiss Re is now working on embedding this forward-looking thinking into decision-making. The approach varies between risks but includes actions, such as monitoring the "key risk indicators" and taking appropriate action when they show that emerging threats are crystallizing, ensuring that emerging risks are appropriately charged for, or clarifying risk appetite to ensure that certain risks are not taken onto the balance sheet.

Greater embeddedness: integrating risk management, capital management and business planning

Swiss Re is currently implementing improvements in its planning process to better integrate financial and risk management perspectives with the aim of enhancing their understanding of the range of potential outcomes and drivers over the planning horizon.

Greater empowerment: further enhancing the first line of defense to empower those responsible for taking on risk

Giving risk takers in the business greater responsibility can create a culture where compliance is not only about following rules, but is also about applying knowledge and experience contextually and interpreting risk accordingly. One of the goals of Swiss Re's current project has been to ensure that the first line of defense is obligated and fully empowered to manage risk, and that the role of the first and second line is clearly understood. Swiss Re has translated this into practice by implementing clear "delegated authority" thresholds within a defined framework for risk takers in the business before requiring approval from Risk Management, complemented with sample ex-post risk management peer review of any decisions taken. While this has been in place for some time, Swiss Re found that there was often a lack of clarity on ownership of the risks by the first line, and a reliance on the second line for analysis.

To address this, the project is developing a clear “target operating model” that maps all the key risk management activities (identification, assessment, monitoring, execution, and control) to relevant owners in the business and Risk Management, and syndicated this widely. Swiss Re is also investing in the development of directly accessible management information for first-line decision makers to enable them to perform a better risk analysis.

Greater effectiveness: integrating and aligning risk management activities

Even with a highly empowered first line of defense, there is a need for some degree of oversight from the second-line functions. However, in many companies, including Swiss Re, opportunities exist to better align the activities of the second-line functions in order to maximize their effectiveness and efficiency.
As the risk management functions of insurance companies have evolved organically over the years, separate control functions have typically been added as the need for them was identified, often with their own systems, processes and information flows; for example, for market, credit and liquidity management, operational risk management and compliance. These organizational units often interact independently with separate reporting processes, sources of data, and expertise. As such, they tend to be separate not only in terms of operations, but also in terms of skills. As an example, does the “liability” side of the risk function understand the “asset” side of the balance sheet to the appropriate level, and would they be able to identify risks outside of their area of expertise? In many organizations, the answer to this question would be a “no,” because an integrated view in the Risk Management function is not possible except at the most senior level. Furthermore, capturing risk exposure as a whole is hindered by the overlap and gaps between the stages of risk identification, measurement, monitoring and control, and inconsistent sourcing of information, which leads to staff wasting time reconciling information rather than adding value by truly managing risk.

Swiss Re is looking to implement a single risk register for all operational risks, alongside a top-down mapping of where these risks emerge in the process landscape. The aim of this enhancement is to establish a common universe to drive reporting so that all functions have a shared view on materiality and prioritization.

Greater entrepreneurialism: providing value adding advice to the business
Proper risk management is about enabling the right decisions, not controlling the business. The business has to be empowered and encouraged to take risk decisions with opportunity and scope for challenge to come at each level of the business. Just as risk takers should be empowered to be more risk-aware, the risk management framework has the potential to provide a forum for constructive peer review and advice on optimizing commercial decisions as opposed to “tick-box” controls.

This is primarily a cultural shift: to encourage this, Swiss Re wanted to ensure that the Risk Management function was involved in key decision-making processes to provide an independent perspective and suggest new ideas to enhance the risk and return profile. The intention was to tap into the unique experience of risk managers, who can provide forward-looking thinking on trends, which can then be used to directly influence products, investments and other business opportunities. As an example, risk managers consulted in the product development process could help business managers to incorporate safeguards in contract terms, which would potentially allow greater volumes to be written at a lower price for the same capital capacity, thereby increasing competitive advantage.

To successfully implement such a change in the relationship between the first and second lines of defense, Swiss Re’s risk managers would need to establish themselves as credible and skilled contributors to the debate, which in many cases requires enhancement of their current skill sets and approach. This is a clear example of an area where getting the most value out of an enhancement requires a cultural change across the business: in this case, changing both the skills of the Risk Management function and the business managers’ view of their potential role. This is not a simple or quick change to make, but is key to realizing the potential value of the other changes discussed in this paper.

Greater efficiency: improving processes and communication
A key objective of the program at Swiss Re has been to achieve material enhancements to the risk management capabilities of the organization while meeting increased regulatory requirements and without increasing the cost to the company. The core aim of empowering the first line to take more responsibility for risk is to allow risks to be managed more effectively at the point of origin. However, a secondary benefit is a more efficient and targeted decision-making process that thus frees up the core Risk Management function to focus less on controlling the day-to-day decisions and to look more strategically across the business. This delivers part of the desired productivity gains, but typically more work is needed to achieve these.

Through our review of capabilities, we have noted that a material proportion of risk managers’ time is spent on data cleaning and reconciliation. In addition, the complex network of systems that have grown over time in the Risk Management and Finance functions is leading to some processing inefficiencies. Swiss Re has chosen to address this by investing in a “minimal risk
Intelligent ERM: Evolving risk management

intelligent “infrastructure” program to materially simplify the IT landscape and align it to all systems. The aim is a fully integrated, “right-first-time” mentality with minimal applications that serve only the Risk Management function. This will be supported by straight through processing that uses data from the source to minimize data cleaning and reconciliation time. This is being designed to support the first line of defense and ensure they have the right information produced as efficiently as possible.

Greater engagement: realigning the risk culture
The key first step toward enabling cultural change is to set and clearly define the target culture. Fundamentally, Swiss Re wanted to encourage a greater engagement with risk and risk culture across the business by changing the way that the Risk Management function was perceived in the business. Swiss Re wanted risk management to not just be the responsibility of the Risk Management function, but to be something that should involve the entire organization. This message has been supported and clearly communicated through strong sponsorship of the change program from the Board and Executive Committee, which should eventually filter down to every level of the business.

Once this message had been clearly communicated, making the change and realigning the risk culture required putting the right incentives in place so that people could demonstrate a commitment to managing risk. To encourage this new behavior, it should be included in performance assessment and remuneration schemes and, eventually, form part of the recruitment and promotion processes.

Embedding change across the organization
In this section, we will provide a more detailed overview of where some of these changes are being embedded across the organization in four key areas: organization and governance; process; data and IT models; and people and culture. As per the previous section, these groupings have not been made to suggest that these changes are being made in isolation – there are overlaps and interactions that contribute toward the wider program of enhancement taking place across the business.

Clarity in organization and governance
A number of the focus areas mentioned above are about clarifying and changing people’s responsibilities around the organization. When designing a new framework, it is essential to address key questions early on: for example, what is the role of Group versus the Business Units and functional relationships, such as Risk, Finance, Underwriting, and so on? Swiss Re, like many organizations, has previously defined this at a high level in a number of policies, but this needs to be further defined to ensure that there is sufficient clarity and to facilitate change.

This exercise is being done at Swiss Re using two “lenses”:

- The first lens is the well-known “target operating model,” which shows graphically the target roles of different organizational units with respect to the key risk management activities, such as identification, measurement, monitoring and so on. This is being developed through substantial consultation and iteration within the business over a period of one year, driving understanding and ownership of the target state in the process.
- The second lens is the more innovative of the two: Swiss Re is creating a detailed mapping of all the economic balance sheet positions (or risk exposures) of the group and entities to the organizational decisions that cause those exposures to be added, removed or altered. This will then detail the ownership for the different cells in the matrix in terms of risk taking and controlling, and map these to relevant policies and standards. This will allow Swiss Re to test substantially whether certain exposures or decisions are subject to too much or too little oversight. The aim is to highlight areas where there is duplication of oversight between functions, and some areas where oversight is insufficient and should be strengthened.

This has been a challenging process. Swiss Re followed a highly consultative, bottom-up process to develop the target operating model, which had the benefit of achieving buy-in but required significant management time and resources. Executive, top-down mandates on some key organizational changes and design principles at the outset could accelerate the process; however, firms need to balance this with allowing team managers to develop their own solutions to meet the objectives of the program. These managers need to deliver the solutions day-to-day after all. An important requirement regardless of the approach chosen is to build sufficient challenge into the process from outside the Risk Management function. This will ensure that changes are addressing the needs of all stakeholders, and are made in full awareness of leading practice benchmarks.
Process standardization and transparency

The second key lever for implementing change is to make processes more transparent and standardize them where this makes sense. This is not to be mistaken for a SOx-type documentation exercise covering the whole process landscape. The key at Swiss Re was to be selective: the target operating model process is being used as a mechanism for testing with team members which processes suffer from lack of clarity on responsibilities and which could benefit from simplification, standardization and automation. These will then be further documented and standardized. Following standardization, centralization is a further action that is open to companies (especially those with offshore service centers, like Swiss Re) to further reduce costs where possible. The lesson learned here is to be smart about which processes are looked at, and define these in terms of the ultimate services or outputs that are delivered.

Data, IT and models

Companies need an IT architecture that offers cost-efficient support while also supporting efforts to standardize and industrialize business processes. This should enable the convergence of risk, finance and capital management, and facilitate the efficient development of risk and finance scenarios across the business.

At Swiss Re, the program has had a dedicated “risk infrastructure” work stream from the beginning, which has owned the detailed future state architecture vision. This team operates to some strict design principles to seek efficiency gains and simplifications in the “as is” landscape, but also to integrate new requirements identified by capability enhancement projects in the rest of the program.

Some of these design principles include:

- Ensuring that the first line of defense has systems in place to provide the necessary information of sufficient quality to allow risks to be measured, monitored and managed
- Implementing a minimal systems infrastructure within the Risk Management function that fully leverages the first line’s information without duplication
- Integrating the data and systems landscape between business management, finance, risk and actuarial, so that the finance systems are at the core and there is zero duplication of data or processing to support risk analysis; this requires IT systems with integrated data architecture, which enable more efficient and standardized analyses and reports that can speak to the variety of reporting needs across stakeholder groups
- Balancing flexibility and necessity by making systems bespoke and flexible only when this adds demonstrable value to the organization
- Ruthlessly decommissioning any systems that have had a material part of their functionality replaced by a new application
- Supporting the transparency of risk data, which should always be a resource for the wider business, not a resource for the Risk Management function

Although fully common data and systems between risk and finance would be ideal, this would be too challenging to implement in practice. As such, the IT landscape should be scoped and designed carefully from the outset: areas where full leverage is not possible will require automated reconciliation.

A significant success at Swiss Re was achieving cross-functional buy-in to a new, integrated landscape blueprint, which is now in the process of being implemented. The challenge will be maintaining momentum for the three-to-five-year time span to implementation.

Cultural change

Cultural change is probably the most important lever for achieving the goals set out above, but also the most difficult to implement. One of the key challenges of successfully implementing and maintaining this kind of change is in convincing people that they should change: not because they have been told to do so, but because it is the right thing for the business. This will be key in ensuring that the changes that are implemented are really embedded and sustainable and, therefore, last beyond the end of this phase of work. It is the people who will continue to shape the risk culture going forward and who should understand and be actively engaged with risk management.

One thing we have noted in the Swiss Re program is that there is no “silver bullet” for achieving cultural change: it is the cumulative effect of a number of “nudges” that gradually change behaviors. Some of these nudges are quite powerful and need to be implemented at the start:
1. It is important to have the right incentives in place to encourage the desired behavior: ideally, key staff should be measured based on intelligent risk-taking, with risk culture and behavior forming a key element of any incentives. At Swiss Re, all executives in key risk-taking positions have risk-based return or demonstration of appropriate risk culture as part of their goals, and the Risk Management function is also being offered incentives on its level of support and enablement to the business.

2. Setting the right tone from the top is essential: senior executives are the public face of risk management success, as well as business success more generally, and they need to ensure they are communicating this message properly and encouraging leadership at every level of the business to also communicate this message to their staff.

3. Education is essential to ensure people in the business know about the risk tools and techniques available: it is also necessary to provide training to those in the Risk Management function to ensure that they understand and are up-to-date with business issues. One way that this can be facilitated is, for example, by regular staff rotations between Risk Management, Finance and the wider business.

Alongside these nudges, cultural change is all about coordinated, sustained communication of the changes being made to the business and everyone’s role in these changes. At Swiss Re, this change program has the following elements:

- **Top down:** a cascade process to drive the ownership and sustainability of the process. This starts with a series of workshops with executive management to explain the vision and to facilitate thinking about how they will contribute to the change from within their divisions; they then facilitate a follow-on workshop with their team leaders with the same objective, and these team leaders then repeat the process with the next level down. Currently, this has only been within the Risk Management function, but the bigger challenge is to eventually expand this more widely in the business.

- **Bottom-up:** “Change Ambassadors” embedded at various levels of the business can be used to support volunteer projects within the overall change program: they will be responsible for communicating positive elements of change back to the wider business. Structured regular communications about how the projects within the program are changing the working lives of real people in the business can help bring the program to life for all employees.

It should be noted that the changes highlighted above are part of an ongoing program of work. As such, this paper presents a snapshot of some of the current areas where value has been added, and of some of the current challenges. The approach continues to evolve and lessons are continually being learned.

**What firms should do now**

In choosing to implement a vision for a more intelligent, embedded and efficient risk management practice, Swiss Re is continuing to refine and adapt the translation of its vision into a long-term strategic practice for their business. Although this phase of work is ongoing, there are some initial findings and observations that any business thinking about implementing a similar program of enhancement should consider:

- Businesses should start the process of thinking about their risk management capability and should consider whether it is likely to be fit for purpose for the future of their company.
- As a first step, they must be able to respond to regulatory requirements and articulate their vision for risk management and must be able to describe the outcomes they wish to achieve, at least in the medium term.
- They should talk to and get buy-in to the proposed changes from people across the entire organization.
- Firms should consider seeking an external perspective on best practice from outside the organization. As an example, Swiss Re sought input from a range of external sources, including CROs from within and outside of the financial services sector, regulators, rating agencies and academics.
- Most firms have a good base for enhancement and, as such, need only to define their focus areas for change based on current performance rather than needing to overhaul the whole business, or go into too much detail with this gap analysis. Firms should invest some time up front to identify areas where improvement will add the most value so that the business has a clear focus and change strategy, thereby preventing change fatigue. Consider developing a clear target operating model that maps all the key risk management activities (identification, assessment, monitoring, execution and control) to relevant owners in the business and risk management.
- Businesses undertaking such a program need to define concrete targets based on measures of success for each of the focus areas: these can be measurable targets (such as for productivity or efficiency), as well as more subjective ones.
feedback from the business or board members, for example).

Quick wins are possible to demonstrate value, but businesses should also approach the project with a long-term horizon and recognize that some of the changes that they would like to make (for example, data and IT overhaul or cultural change) will take several years to implement.

This intelligent culture of risk, if embedded efficiently and effectively, can potentially add value to the business, enhance existing processes and ultimately help businesses continue to manage and mitigate the existing and emerging risks that they continue to face in a changing world.
The shifting terrain of risk and uncertainty on the liability insurance field

Tom Baker
William Maul Measey Professor of Law, University of Pennsylvania Law School

Abstract
[A] high degree of scientific and technical uncertainty permeates the insurance industry – the very business that is charged with transforming uncertainty into risk. Scrutiny of how insurers face their own incapacities reveals that theirs is an uncertain business, and that we live not so much in a risk society as in an uncertain society.¹


* Thank you to Jay Feinman and Sean Fitzpatrick for helpful comments, and to David Pupkin for updating the data. This article was originally published in 2011 as part of the Clifford Symposium, Volume 60 of the DePaul Law Review. This version updates the figures to include insurance industry financial data through 2011.
Introduction
Concepts can impede understanding, sometimes by persuading us that we understand something that we really do not; other times, they impede understanding by diverting our attention from what we would readily recognize that we do not understand, if only we thought long enough about it. The conceptual link between insurance and risk — with risk understood here as the calculable subset within a larger set of uncertainties — may be one such impediment. If insurance and risk are tightly linked, and if risk is calculable, then we may think that we can understand the insurance business as a kind of applied econometric enterprise.

Work in law and political science has challenged the idea that insurance underwriting and pricing should be based on purely statistical considerations. Yet, even this work has tended to take as given the links between insurance and risk and between risk and calculability. Risk classification undoubtedly is “[t]oo [i]mportant to be [l]eft to the [a]ctuaries,” but the substitutes offered are other concepts — equal protection, privacy and solidarity — that would trump the (imagined) purely statistical analysis only in limited circumstances. Left uninvestigated and therefore not understood by these challenges is the process through which insurance institutions actually transform (or do not transform) uncertainty into risk.

Recent sociological and historical work — most prominently that of Richard Ericson and Aaron Doyle on the one hand, and Timothy Alborn on the other — has helped to remind us that insurance risks very often are not reliably calculable except in hindsight, at which point the risk has already been transformed into an all-too-measurable loss. Insurance is an “uncertain business,” characterized by competition for premiums that pushes insurers into the unknown. Even life insurance, the part of the business that manages the oldest and best-understood insurance risk — mortality — operates just beyond the limits of knowledge. For a recent example, consider the life settlements market through which hedge funds purchase and hold bundles of life insurance in a strategy to exploit and systematically frustrate life insurers’ actuarial assumptions.

This article takes some preliminary steps in the direction of extending the insights of this sociological and historical work to the liability insurance field. The next section begins with a simple quantitative comparison of U.S. property and liability insurance premiums over the last 63 years, setting the stage to make four points. First, liability insurance premiums have grown by about as much as property insurance premiums over this period, providing yet another piece of evidence supporting the view that the growth in U.S. liability costs represents an ordinary consequence of a growing economy rather than an unusual or pernicious feature of U.S. culture.

Second, there are systematic variations in the rate of growth in insurance premiums over time corresponding to what is known in the insurance industry as the underwriting cycle. The cycle is more pronounced in liability insurance than in property insurance, perhaps accounting for the greater popular attention to the uncertainties of liability risks as compared to property risks. Third, the greater variation in liability insurance premiums is almost entirely attributable to the smaller, non-statutorily required lines of insurance: most significantly the various lines of commercial liability insurance. Auto liability and workers compensation premiums, by contrast, largely vary in tandem with property insurance premiums. Finally, because media coverage of liability and liability insurance increases during the insurance “crisis” stage of the underwriting cycle — when premiums increase sharply — the widely held beliefs in myths about litigation may result, at least in part, from generalizations drawn from a systematically biased set of

References


See infra notes 15-22.
observations, analogous to the media bias toward big, unusual cases documented by William Haltom and Michael McCann.13

The following section moves in a more qualitative, speculative direction by considering liability insurance as a “business that is charged with transforming uncertainty into risk.”14 It discusses how one might explore more systematically in the liability and insurance context the significance of Ericson and Doyle’s remarkable conclusion that uncertainty-reducing innovations shift the limits of insurers’ knowledge but do not shift their need to operate just beyond those limits. The section explores some of the changing terrain of risk and uncertainty on the liability insurance field, closing with the admittedly speculative conclusion that the insurance underwriting cycle might be both a cause and a consequence of liability insurers’ efforts to push the liability and insurability frontier.

Comparing liability and property insurance premiums over time

In setting out to compile the aggregate data that would form the empirical backdrop for this article, I assumed two things that turned out to be wrong. First, I assumed that compiling the data would be easy, because the same statistical service (Best’s) has continuously tracked the property casualty insurance industry for more than 100 years.15 Second, and more interestingly, I assumed that liability insurance premiums would show much more significant growth than property premiums, particularly in the latter third of the 20th century.

Like other empirically minded torts teachers, I have followed the tort reform debates closely, and although I have disagreed with the defendants’ lobby about the interpretation of the data,16 I knew that aggregate liability insurance premiums in the U.S. had grown substantially in real terms over the 20th century. I had always assumed that this growth significantly outpaced the growth in property insurance premiums. Otherwise, why were insurance industry sources so outspoken about rising liability insurance premiums? In most cases, the same insurance groups write liability and property insurance, and I assumed (wrongly, it turns out) that their comparative silence about property insurance premiums meant that those premiums were not increasing as much.

Figure 1 shows that I was wrong about the relationship between property and liability insurance while also showing that, despite some difficulty, it is possible to compile the data.17 Figure 1 plots the aggregate liability and property insurance premiums from 1939 to 2011 in constant 2008 dollars for the entire U.S. insurance industry, as reported by Best’s based on data from the annual reports required to be filed with state insurance departments in the U.S. (and converted to 2008 dollars using the consumer price index table from the Bureau of Labor Statistics).18 The black line represents total liability insurance premiums; the gray line represents total property insurance premiums; and the yellow line represents the difference between the two.

13 See Haltom and McCann (2004), supra note 11, at 156-59.
14 Ericson and Doyle (2004), supra note 1, at 5. See also infra notes 25-51.
15 See Best’s, Aggregates and averages: property casualty edition and predecessor publications, AM Best.
17 The difficulty comes from the fact that the data categories have changed several times during the 1939 to 2011 period, reflecting changes in the categories in which insurance regulators have required insurance companies to report their premiums and in which Best’s has aggregated that data in its annual publication.
18 See Best’s, supra note 15, U.S. Department of Labor: Consumer Price Index, ftp://ftp.bls.gov/pub/special.requests/cpi/cpial.txt (last visited March 31,2011)(updated monthly). These aggregate premiums likely underestimate the growth in property and liability losses in the large commercial sector of the U.S. economy, as large corporations have increasingly retained the most predictable and most costly primary layer of exposure, but I am aware of no reason that this trend would have a differential effect on property and liability insurance premiums.
From 1958 – the year when total liability premiums first equaled total property insurance premiums – to 2011, property insurance premiums have grown at a real average annual rate of 3.21% and liability insurance premiums have grown at a real annual average rate of 3.67%. Over that period, property insurance premiums have grown a total of 368% and liability insurance premiums have grown a total of 436% in constant 2008 dollars. Over that same period, GDP grew 414% in real terms, indicating that the property insurance premium growth slightly lagged that of the economy generally and liability insurance slightly outpaced that growth, but the orders of magnitude are very similar.19

As the yellow line in Figure 1 reflects, liability insurance premiums have grown at a more variable rate than property insurance premiums. From 1958 to 2011, aggregate liability insurance premiums averaged 124% of aggregate property insurance premiums, but there was an 11-year period – from 1986 to 1997 – in which the liability insurance premiums were 145% of the property insurance premiums.20 The difference between the premiums in the two categories of insurance is greatest following the three most significant “hard market” phases of the underwriting cycle that occurred during this period: 1976–77, 1985–86, and 2002–03.21 Much of the difference is attributable to rapid growth in liability insurance premiums during just two years: 1985 and 1986 (a 56% increase in liability insurance premiums over those two years compared to 28% increase in property insurance premiums). The rest is attributable to differences in growth during 1975–76 (34% for liability and 26% for property) and 200–03 (27% for liability and 20% for property), and to quicker declines in growth in property insurance premiums following each of the three hard market periods.

This difference in the variability of premium growth rates helps explain the popular perception that liability insurance has grown more rapidly than property insurance. During hard market periods, liability insurance premiums in fact do grow more rapidly than property insurance premiums. These are also the periods in which articles about the problems in liability appear most widely in the media.22 Thus, although liability and property insurance premiums may be growing at roughly the same rate on average, liability insurance premiums grow more rapidly when people are most likely to pay attention, leading to the facially plausible but in fact incorrect belief that liability insurance premiums – and therefore the underlying liabilities – generally grow at a rate significantly exceeding the growth in GDP. In other words, the perception that liability and liability insurance premiums have grown disproportionately in relation to the underlying economy represents a generalization from a biased set of observations.

Figure 2, which presents a breakdown of the liability insurance premiums over time by category of insurance, reflects a reality that is often forgotten: at the aggregate level, the liability insurance market is dominated by the legally mandated forms of liability insurance – auto liability and workers compensation. Automobile liability insurance premiums account for more than half of the total in all but a few years during this period. Only in very recent years have all of the other kinds of liability insurance...
put together equaled the workers compensation share of liability insurance premiums. Moreover, the slight decline in the share of workers’ compensation insurance premiums reflects slower growth in workers compensation insurance premiums rather than particularly rapid growth in the other non-auto lines of liability insurance. Despite all the attention to medical malpractice and products liability in the media and the academic literature, medical malpractice insurance premiums represent a tiny fraction of the total; and the products liability insurance premium line is barely visible at the bottom of the chart.

Thinking about Figures 1 and 2 together raises a question about what portions of the liability insurance market are driving the difference in the variability of the liability and property premiums over time. Figure 1 shows that liability insurance premiums are more variable than property insurance premiums. Figure 2 suggests that the lines of insurance that are not the legally mandated forms of liability insurance have grown more significantly than those traditional lines (auto and workers compensation), but it does not allow for easy comparisons in the rates of growth. Figure 3 does just that: it shows the annual real growth in property insurance premiums (the gray line), the statutorily mandated liability insurance lines of auto and workers compensation (the yellow line), and all of the other lines of liability insurance (the black line). As Figure 3 shows, the growth in automobile and workers compensation lines has closely tracked that of property insurance since the 1950s. The big difference lies in the other lines of liability insurance, suggesting that there are greater year-to-year uncertainties in those other lines of liability insurance than in either automobile liability or workers compensation insurance.

Transforming uncertainty into risk?
Summarizing the discussion thus far, what we see is that aggregate U.S. property and liability insurance premiums have grown at roughly the same rate, on average, since the middle of the 20th century, but the growth in liability insurance premiums varies over time more significantly than the growth in property insurance premiums. The long-term total growth in liability insurance premiums is largely attributable to automobile liability insurance, with the premiums for all non-auto liability insurance exceeding the premiums for workers compensation insurance only in the last 10 years; this is apparently a result of decline in the growth of workers compensation insurance rather than a significant increase in the growth of these other kinds of insurance. Finally, developments in those other kinds of liability insurance account for an outsize portion of the variation in liability insurance premiums.

If Ericson and Doyle’s generalizations about risk and uncertainty hold true in the liability insurance market, then these aggregate results should be masking even more turmoil in that market. There are good reasons to believe that Ericson and Doyle’s generalizations do hold true in this context. Ericson and Doyle investigated three sectors of the insurance industry: life insurance, disability insurance and property insurance.23 Notwithstanding the significant differences among these sectors – different kinds of insured risks and different approaches to managing those risks – the researchers found a common pattern of competition pushing organizations just beyond the limits of knowledge.24 To believe that this pattern would not hold in the liability insurance context, we would need to have some reason to conclude that liability insurance risks are more calculable than life, disability or property insurance risks.

Yet there are good reasons to believe that liability insurance risks may be less calculable than these other risks. The forms of insurance that Ericson and Doyle investigated are loss based,

23 See generally Ericson and Doyle (2004), supra note 1.
24 See id. at 22–23.
meaning that they provide financial protection for the designated harm (death, disability or property damage) without regard to how the harm took place. Pricing these kinds of insurance requires estimating the rate and cost of the insured harms. By contrast, liability insurance provides financial protection only when the harm can be attributed to a liability-creating act, such as the breach of the standard of reasonable care or – in the workers compensation context – an injury arising out of work. This means that the uncertainties of liability insurance include not only developments in the rate and cost of insured harms but also developments in the standard of care and other aspects of liability. Uncertainties about these latter developments add an additional layer of uncertainty to the uncertainties about the underlying injuries and damage. The greater variation between property and liability insurance premiums at the aggregate level, as described above, provides some support for this conclusion.

Nevertheless, even if liability-related risks are only as uncertain as those addressed by disability, life and property insurance, that would leave more than enough room for the dynamic that Ericson and Doyle described: namely, that uncertainty-reducing innovations shift the limits of knowledge but not insurers’ need to operate just beyond those limits. In the remainder of this section, I will offer some admittedly impressionistic suggestions of some of the ways that this dynamic plays out in the liability insurance context.

Transforming uncertainty into risk
On the basis of the observation and study of the liability insurance market, my sense is that three key developments have improved liability insurers’ ability to predict liability risks. This sense represents informed intuition, without systematic empirical support. Thus, the following should be regarded as hypotheses, not proof: (1) advances in data gathering and analysis have led to a quantitative revolution in the tracking and valuation of routine claims, especially in the automobile liability context, resulting in more accurate forecasts of both the frequency and severity of insured liability losses; (2) because the liability insurance policy limits purchased by individuals have not kept pace with inflation, once again especially for automobile liability, an increasing percentage of claims are capped at the policy limit, in effect truncating the right tail of the severity distribution (i.e., the high-severity claims that happen less frequently); and (3) insurers have shifted an increasing proportion of their professional and commercial liability risks into a kind of insurance contract that poses less of the difficult-to-evaluate “long tail” liability risk than the contracts of the past. The paragraphs that follow describe each of these hypotheses in turn.

The quant revolution
Although still very much a work in progress, the quant revolution has significantly changed the valuation of routine accident claims, especially in the automobile liability context. The automobile insurance adjusters from the 1960s, as profiled in H. Lawrence Ross’s *Settled out of court*, used their intuition and experience to craft settlement offers. Today, automobile insurance adjusters use claims-valuation software that takes the injury and demographic information and produces a range of permissible settlement offers. To my knowledge, there are no publications reporting the results of horse races between the quantitative models and experienced adjusters, but my informed impression is that the automobile insurance industry has completely shifted to data-driven adjusting for routine claims and relies heavily on claims-valuation software even when adjusting serious, non-routine claims. This shift strongly supports the view – expressed forcefully in Ian Ayres’ *Super crunchers* – that “intuition and experience are evolving to interact with data-based decision making.”

The insurance industry has long had massive amounts of data. Since the 1990s, the industry has been working on making that data accessible for (proprietary) analysis, for example, by putting underwriting and claims data on the same platform, with the result that the “truly massive” datasets that Ian Ayres describes are now available at all of the major automobile insurance companies. Other lines of insurance may have lagged

26 See generally Ross, H. L. 1980, *Settled out of court*; the social process of insurance claims adjustment, Walter De Gruyter.
The clear trend in the liability insurance industry is to extend the quant revolution into other lines and into more severe claims. This trend points toward the realization of the “actuarial” approach to liability insurance that many lawyers and law professors probably thought was already more fully embodied in insurance practice than was actually the case.

The increasingly binding nature of liability limits

Absent the poor claims-handling that leads to liability for “bad faith,” an insurance company’s liability for any given claim is capped at the dollar limit of the applicable insurance policy. No matter how serious the injury or how wrongful the defendant’s conduct, the liability insurance policy limit is the most that a liability insurance company is contractually obligated to pay as long as the claim is handled properly. Because plaintiffs almost never collect “real money” from “real people,” even a claim that involves damages that significantly exceed the liability insurance policy limit will ordinarily be settled for no more than the available insurance amount. Indeed, the more the value of the case exceeds the policy limit, the more quickly the case will be settled for that policy limit because of the well-established insurance law “duty to settle.” This dynamic is well established and well appreciated by informed observers of the liability and insurance universe.

Less well appreciated, however, is the aggregate consequence when liability insurance premiums fail to keep up with inflation. As Kenneth Abraham explained in The liability century, the mandatory minimum automobile insurance policy limits have significantly lagged behind inflation, with the result that tort damages as determined by the law on the books are increasingly likely to exceed the available automobile insurance. Because the insurance companies’ obligation to pay is capped at the policy limit, and because “blood money” is so rarely paid, this trend means that the damages being increasingly awarded by tort law in action are simply the liability limit of the defendants’ automobile liability insurance policy, sometimes supplemented by the plaintiffs’ underinsured motorist coverage. Research using the Texas Department of Insurance medical malpractice claims database suggests that the same phenomenon may be occurring with respect to medical malpractice claims, though with less impact on liability insurance overall because of the relatively small size of the medical liability insurance market.

The changing nature of liability insurance contracts

The two developments just discussed are most significant for automobile liability insurance. This third development affects commercial liability. Beginning with medical malpractice insurance in the 1970s, the liability insurance industry has gradually been shifting its “long tail” business (i.e., liability risks in which there is a long lag between the wrongful conduct and the eventual lawsuit) into liability insurance contracts that provide coverage on a claims-made basis.

Traditionally, liability insurance provided coverage on an accident basis, meaning that the policy covered accidents that took place during the policy period. This is still the case for automobile liability insurance. In the 1960s, the insurance industry shifted toward providing coverage on an occurrence basis, which extended the original concept of accident to incorporate events that took place gradually, potentially over a long period of time. This approach led to massive exposures for asbestos and environmental injuries in the commercial context, and for birth and childhood-related conditions in the medical malpractice context.

In response, the insurance industry shifted in the 1970s and 1980s to providing coverage on a claims-made basis in professional liability insurance and somewhat later – more

---

30 CF Simon, supra note 3, at 786 (describing the critique of actuarial practices in which the individual is “the sum of the many roles he plays as a result of being a member of many status groups”).
33 Syverud, supra note 31, at 1116.
36 See Zeiler et al. (2007), supra note 34, at 59.
38 See id.
The shifting terrain of risk and uncertainty on the liability insurance field

Gradually, and less completely — in the product and environmental liability context. A claims-made policy provides coverage for claims that are first made during the policy period. The claims-made form of coverage provides greater predictability to insurers for two reasons. First, the policy is more likely to be sold close in time to when the claim is paid, thereby allowing the insurer to incorporate more recent information about the frequency and severity of claims into the price for the coverage. Second, because a claim can be first made only once, there is less likelihood that policies from multiple years will be triggered by a claim or a set of related claims (such as a mass tort), thereby capping the insurer’s exposure at the limit of a single policy (in marked contrast to the situation under an occurrence policy).

Transforming risk into uncertainty
At the same time that these developments transformed uncertainty into risk, other developments pushed in the opposite direction. In the personal lines liability insurance context, which mostly means automobile liability insurance, the two most important developments were the “arms race” in risk segmentation and the politics of automobile insurance pricing. The commercial lines context is more complicated because commercial liability risks evolve with the underlying economy. One important and easy-to-explain commercial lines development concerns corporations’ increased willingness to retain risk, with the result that commercial liability insurance increasingly involves the right tail risk, precisely the opposite of personal automobile liability insurance. The sections that follow explain each of these.

The arms race in automobile risk segmentation and distribution
Of all of the forms of insurance investigated by Ericson and Doyle, personal automobile liability insurance is most like life insurance, at least in terms of the dynamics of uncertainty and risk. Like mortality, automobile liability risk is well understood in the aggregate. And, because of the quantum revolution in claims evaluation and the increasingly binding nature of automobile liability insurance limits, the loss associated with any given automobile liability claim is almost as determinable as the loss associated with any given death claim. So, as with life insurance, the competitive action lies in innovations in the slicing and dicing of the insured population and in product distribution.

The politics of automobile insurance pricing
The politics of automobile insurance pricing inserts an additional uncertainty into the automobile liability market that does not exist in the life insurance market. With the transition of the Massachusetts automobile insurance market to a kind of managed competition, there are no longer any state insurance departments that set rates. But all state insurance departments retain some authority over changes in rates. Given inflation and the fact that insurance rates are set in nominal dollar terms, all insurance companies eventually need to raise their automobile insurance rates. In most cases, insurance regulators do not object. But sometimes they do object. That is a source of uncertainty.

40 The recent “accident forgiveness” programs represent a partial exception to this rule. Strictly speaking, accident forgiveness programs represent a change in pricing, not in product design, but the programs are marketed as a feature that is analogous to a change in design.
Commercial liability insurance: a window on the underlying economy

It is difficult to make useful generalizations about a business that is as diverse as commercial liability insurance, a field that in its broadest definition includes professional liability insurance, many specialized forms of liability insurance and the more commonly understood general liability insurance. Liability insurance tends to extend over time to match liability itself (with lags and with exceptions), and thus developments in liability law, both on the books and in action, represent a source of uncertainty for liability insurers. Almost all changes in the economy produce at least some change in the mix of liabilities assumed by liability insurers, and thus the creative destruction of a capitalist economy represents another source of uncertainty for liability insurers. The shift from manufacturing to services, the globalization of the supply chain and the expansion of the high tech and health care sectors of the economy, among many other developments, all have consequences for liability insurers if only by reducing the predictive value of historical information about the frequency and severity of claims.

Corporations retain more risk

There is one major cross-cutting development in the commercial lines marketplace that is worth singling out: businesses of all kinds are retaining greater levels of risk, as represented by rising deductibles and self-insured retentions, and by the use of captive insurance organizations (wholly owned “insurance” subsidiaries of commercial business organizations). Businesses, unlike individuals, have learned what may be the most important lesson of risk management theory and research: because of the inevitable loading costs of insurance, it makes no sense to buy insurance for losses that are small in relation to easily available assets. Since at least the 1980s, businesses have preferred to buy liability insurance policies with higher deductibles and higher limits. The aggregate result is to remove the most predictable liabilities from the liability insurance pool, leaving insurers dependent on providing protection against relatively extreme liabilities. Replacing the premium lost when a corporation raises its workers compensation or liability insurance deductible by U.S.$1m requires selling a lot of excess insurance, because excess coverage is less expensive than primary insurance on a per-dollar-of-coverage basis. As a result, commercial liability insurance has become increasingly focused on the right tail of the liability claim distribution, which is characterized by larger, less frequent losses, especially for large organizations. Liability insurers believe that, in the jargon of the business press, this is a “fat tail,” meaning that commercial liability losses do not follow a normal distribution. As the quant revolution enables large organizations to transform their liability uncertainties into risk, the organizations will retain more of that risk, and they will endeavor to shift the remaining uncertainties to their liability insurers.

The insurance underwriting cycle

As reflected most clearly in Figures 1 and 3, the property casualty insurance industry goes through a business cycle that is characterized by long periods of relatively stable insurance prices, often slowly declining in real terms, and short periods of rapid price increases. There is an extensive literature on this cycle — including a contribution by me — and perhaps the one thing upon which there is universal agreement is that no one fully understands what drives the cycle. Not surprisingly, the cycle is a major preoccupation in the industry. Industry trade publications regularly report on what industry leaders and analysts think about where the industry is in the cycle, and predictions of this sort can have very significant strategic implications for companies.

During a “hard market period,” prices significantly exceed costs and insurers can implement new restrictions on coverage and underwriting. During a “soft market period,” prices gradually decline in real terms and insurers gradually abandon coverage and underwriting restrictions. By the end of the soft market

43 Baker, supra note 25, at 133.
44 Schumpeter, J. A., 1942, Capitalism, socialism, and democracy, Harper Perennial Modern Classics
46 Id. at 527 (reporting that “medium to large business insurance policies often include relatively large deductibles or self insured retentions”).
The shifting terrain of risk and uncertainty on the liability insurance field

period, almost all insurers are almost certainly selling insurance below cost, retaining market share in the hope of making money once the market turns hard. Warren Buffett, whose most significant holdings are property casualty insurance companies, has famously advocated paying underwriters not to write insurance during severe soft market periods, because the company will almost certainly lose money on the book of insurance sold at that time. The problem is, of course, that no one really knows when the cycle reaches the point at which liability insurance costs exceed prices — because the costs will not be known for sure until the future — and no one really knows the value of keeping market share during the soft market period. The underwriting cycle contributes to the uncertainty in liability insurance in a variety of ways. The periodic relaxation and restriction of underwriting and coverage practices reduce the predictive value of each insurance company's claims history, because the mix of risks insured does not remain consistent over time. Keeping market share during a soft market period involves heading off into the unknown. By definition, the relaxation of underwriting and coverage restrictions means selling insurance against risks that the insurer had not been willing to insure in the prior period. Among reinsurers, a continuation of a soft market period means that the reinsurer needs to be easier for insurers to deal with and that means relaxing — and at some point abandoning — detailed reporting requirements at the individual insured and claim level. The result is that reinsurers do not have continuous individual and claim level data of the sort that would allow them to control for the changing mix of business among the insurance companies that they reinsure. When the hard market period begins, insurance companies raise prices and jettison some of their worst risks, almost certainly writing profitable business. But just how profitable remains uncertain, as does the amount of time before enough new capital enters the insurance market for the underwriting cycle to turn once again. In that sense, we can regard the liability insurance underwriting cycle as both a cause and a consequence of liability insurers' efforts to push the insurability frontier.

Conclusion

These competitive consequences of the liability insurance underwriting cycle resonate well with the dynamic of Ericson and Doyle (2004). Even as liability insurers develop technologies that would better predict the losses of a stable risk pool, competition reshapes companies' risk pools, so that they operate just beyond the edge of their knowledge. This dynamic helps explain how the liability insurance field can feel out of control even to expert participants. “Fear is the key,” in the words of insurance industry leader Sean Fitzpatrick. Those feelings put expert passion behind the popular beliefs that have fueled the restrictive tort reform movement. Taking the long view, liability and liability insurance premiums have grown at about the same rate as the U.S. economy. But, few people live their lives consistently taking this long view, which may not even be adaptive in the competitive arena.

51 See Dolan, M., 2003, “Repeating the sins of market cycles,” Insights, October (on file with author) (article by the president of a liability insurance company reporting “our industry realizes the ‘right priced’ environment only momentarily during the shift from the hard to the soft market” and arguing “the cyclicity of this business will be perpetuated, and capacity will engage in wild and irrational movements over the price/quality line, unless carriers” adopt a variety of measures that are unlikely to be adopted given competitive pressures).
52 Author’s confidential, personal communication with reinsurance underwriters as part of the field research reported in Baker, T., and S., Griffith, 2010, Ensuring corporate misconduct: how liability insurance undermines shareholder litigation, University Of Chicago Press
53 See generally Baker, supra note 22 at 393; Dolan, supra note 51, at 1; Fitzpatrick, supra note 12, at 255.
54 See Fitzpatrick, supra note 12, at 255–56 (quoting title of Alistair MacLean's novel, Fear is the key).
Principles and policies for in-house asset management

Gordon L. Clark
Halford Mackinder Professor of Geography, and Director of the Smith School of Enterprise and the Environment, Oxford University
Ashby H. B. Monk
Executive Director, Stanford University’s Global Projects Center, Stanford University

Abstract
Spurred on by the recent financial crises, a growing number of institutional investors are working to bypass traditional financial intermediaries, agents and centers through the development of in-house teams of investment professionals. As such, the institutional investment community, which is often characterized by broadly diversified and outsourced organizations, is becoming much more involved in the day-to-day asset management of their portfolios. Research shows that this new path offers a variety of important benefits, including higher net-of-fee returns. And yet, there remain significant pitfalls as well. In this paper, we outline the challenges facing would-be “in-sourcers” and offer a series of principles and policies for effective in-house asset management. Drawing on 20 case studies, we conclude that successful in-house asset management is a function of the people, processes, systems and overall resources at the disposal of management.

1 The authors are pleased to acknowledge the support of Stanford University, and members of the research consortium on in-sourcing and outsourcing coordinated by the Global Projects Center in the Faculty of Civil and Environmental Engineering. None of the above should be held responsible for the views and opinions expressed herein.
Principles and policies for in-house asset management

Introduction
Spurred on by the recent financial crises, a growing number of institutional investors are working to bypass traditional financial intermediaries, agents and even centers through the development of in-house teams of investment professionals. As such, the institutional investment community, which is often characterized by broadly diversified and outsourced organizations, is fast becoming more active in the day-to-day asset management of their portfolios. And, significantly, this appears to be having a positive effect on financial performance, as research now demonstrates that in-house asset management can lower costs and generate higher net-of-fee returns [MacIntosh and Scheibelhut (2012); Fang et al. (2012)].

Notwithstanding the benefits, such as higher returns and more sustainably constructed portfolios, there remain significant challenges and pitfalls associated with in-sourcing. For many institutional investors, such as public pensions and sovereign funds, moving assets in-house requires a complete reorganization of the investment process. The operational DNA of a traditional, externally managed institutional investor looks completely different – and is set up with different objectives – than the operational DNA of an internally oriented institutional investor. “Going direct” thus requires an overhaul of many of the existing institutions and processes; which means in-house asset management is much more difficult to implement than just hiring a few dealmakers. It requires mobilizing significant internal resources and committing them to a specific course of action [Leiblein et al. (2002)].

At issue, then, is the degree to which the broader community of long-term, institutional investors will be able to change their organizations enough to be successful as in-sourcers. Many institutional investors have proven incapable of managing change within their organizations, allowing inertia to be the main factor influencing decision-making. Encouragingly, however, there is a small community of institutional investors that have already gone down this path and can provide useful lessons to other investors considering the same [see MacIntosh and Scheibelhut (2012)]. In this paper, then, we detail some of the experiences and lessons from in-sourcing institutional investors. Specifically, we draw on the insights of 20 in-depth and in-person case studies of public pension and sovereign funds around the world with the objective of developing a series of principles and policies for effective in-house asset management.

The 20 funds that make up our cases are drawn from four continents and comprise many of the biggest institutional investors in the world. We selected our cases on the basis that each had in-sourced at least a portion of their asset management and could offer insights into how it should (or should not) be done. While there are many names you would recognize among our case studies and respondents, we respect the social science guidelines concerning the confidentiality and anonymity of our respondents [Clark (1998)]. As such, no funds are named in this paper, which is why we focus on broader principles and policies rather than dwelling on the specific practices of any one fund. In this respect, our methodological and conceptual framework is similar to that used by Clark and Urwin (2008).

The paper proceeds as follows: we begin with a discussion of the principal-agents problems embedded in the financial services industry. We then highlight the nature of the institutional investment business. Subsequently, we discuss the challenges associated with innovation in this domain, as institutional investment organizations tend to be bureaucratic and sclerotic. Next, we develop a series of principles and policies for effective in-house asset management. In the end, we conclude that the key touchstone to a successful in-house investment policy is governance, which refers to the medium through which the time, skills and commitment of the board is managed to make decisions and set strategy. In our view, it is ultimately the board that holds the strategic levers of success for institutional investment organizations.

“Broken agency” and finance
Institutional investors, such as pension funds, endowments and sovereign funds, exist because their sponsors decided to manage a set of future liabilities and/or uncertainties by setting aside financial assets and investing those assets in financial markets. The rational for doing this can be simplified down to two key reasons. On the one hand, prefunding was a way to ensure that sponsors were making credible and legitimate promises; this was a way to ensure that they could actually meet the obligations they were promising today in the distant future. On the other hand, it was also hoped that the prefunded financial assets would grow at a rate faster than that of the sponsor and the overall economy, which meant that future liabilities could be met at a relative discount today. There was an assumption embedded in both factors that the financial markets offered a reliable mechanism
to manage a variety of long-term policies. Moreover, it was also assumed that institutional investors could successfully achieve their corporate and financial objectives over the long run.

Most of the sponsoring entities of these institutions (e.g., governments or unions) did not, however, build up the capabilities in these organizations to invest their assets in financial markets on their own. Rather, they took the position that they would be better served by using their institutional investment organizations as simple conduits for the financial services industry. In this respect, the institutional investor’s job was to contract for investment management services with external asset managers in the private sector. As such, it was the private asset managers (and not the institutional investors) that made the actual investments in stocks or bonds or other financial securities (which were themselves derivative of some underlying asset or business). In addition, institutional investors used custodians to manage the actual securities and had an array of consultants, accountants, auditors and actuaries to provide oversight.

As this suggests, the traditional institutional investor was almost entirely outsourced, rarely possessing the expertise and competencies to execute even the most basic financial transactions without the help of some external advisor. Said differently, sponsors of institutional investing organizations were seemingly comfortable with the idea that their funds’ success was a function of the effective oversight and management of a long chain of principal-agent relationships, bringing together the sponsors (e.g., governments) with institutional investors (e.g., Wall Street), and advisors (e.g., consultants). And, as it happened, this model did work for a while, thanks in particular to a series of bull markets. But, over time, the extended chain of principal-agent relationships became problematic. In particular, the injection of new incentives and motivations at each link of the chain served to distort the original motives of the asset owners. Too often, the ultimate investment decisions made by asset managers maximized the utility of the asset managers (and not the asset owners); this is a phenomenon known as a “broken agency” problem [Sheffer and Levitt (2010)].

Broken agency stems from misalignment of interests due to an inappropriate distribution of risks and expected returns between the principals and the agents. It is most obvious in large-scale construction and infrastructure projects where the agents that

Box 1: Why in-source?

There are five key factors pushing institutional investors to move assets in-house:

1. **Access:** There are instances where the third-party vehicles are not attractive, and access to a given asset or market can be more effectively achieved on a direct basis.
2. **Alignment:** Principal-agent problems are pervasive in the asset management industry, and some institutional investors view in-sourcing as a useful mechanism to minimize agency costs.
3. **Capabilities:** By developing an investor’s internal resources, all aspects of the organization’s capabilities are improved, as internal teams can identify “unknown unknowns” about the business.
4. **Performance:** Perhaps the most cited reason for in-sourcing by institutional investors was the desire to maximize net-of-fee investment returns.
5. **Sustainability:** Managing assets in-house offers an investor the ability to think critically about how to tailor a portfolio to meet its needs (as opposed to trying to cobble together a series of external mandates).

bear the short-term risk and receive the short-term rewards are different from those that bear long-term risks and rewards. In large-scale infrastructure projects, broken agency results in a number of contradictions between the short-term and long-term parties that can lead to suboptimal outcomes. In the case of institutional investment organizations, however, broken agency distorts portfolio construction; the portfolio that a long-term investor would like to hold and the portfolios that long-term investors actually hold are quite different due to an over-reliance on short-term oriented asset managers. This is the broken agency of institutional investing. The only way to realign interests in this case would be to identify asset managers that have the same long-term objectives as the asset owner. The idea here would be to take a “life cycle perspective” that explicitly involves long-term interests in the design and framing of short-term investment policies and strategies.

This is sound in theory, but, in practice, it is difficult to achieve. It is rare to find private sector asset managers willing to make
decisions that will benefit a portfolio 20 or even 10 years later. In addition, it is rare to see the asset owners (i.e., institutional investors) able to manage the end-to-end investment process required to take a life cycle perspective. And yet, in the wake of the financial crisis, a growing number of plan sponsors have begun to wake up to the misalignment of interests embedded in their principal-agent based investment model. In fact, the onslaught of recurring financial crises over the past few years has jolted some sponsors to push their own organizations toward a more internally oriented investment strategy. The idea is to disintermediate the short-term asset managers through the development of internal teams to refocus on long-term priorities and avoid short-term distortions and broken agency.

**The new business of institutional investing**

Institutional investors organize their production process to realize a given objective, which is typically communicated to management in the form of a return target. In order to achieve this target, investors are obliged to make a series of strategic decisions regarding asset mix, market access points and execution of investment strategies [Campbell and Viceira (2002)]. In making these decisions, institutional investors will tend to have three key resources internal to the organization: human capital, systems and processes. It is important to note that the level of sophistication the fund has in these three areas will tend to drive the decisions made by the organization about assets, access and execution. As readers might imagine, the fact that most institutional investors are lacking in human resources, systems and processes, the asset mix, access and execution has been, for the most part, conservative and outsourced.

Financial intermediaries, it should be noted, were also very adept at packaging the risk and return characteristics of underlying assets into products that institutional investors could easily digest. These products came with specific risk-return characteristics that could be combined with other products to construct a credible portfolio to meet institutional investors’ planned rate of return. This intermediation and derivation was deemed necessary and indeed useful and was the basis for talented asset managers to work with lay asset owners.

As noted above, however, the business of institutional investment has begun transitioning away from intermediation and derivation for lay investors toward direct access to assets by skilled institutional investors. Indeed, in the wake of the financial crisis, a growing community of institutional investors is transitioning toward a new business model based on direct investing through in-house teams. This move toward in-sourcing represents a profound shift from the traditional business of institutional investment. For example, a fund that outsources asset management to external parties will be focused on external searches for managers, manager due diligence, performance assessments and benchmarks, and ongoing monitoring of external parties. A fund focused on in-house asset management will need to hire people to execute the strategies and people and systems to support them. Indeed, it is important that in-sourcers do more than just hire “investors”; they must also focus on building an entire organization that includes a back, middle, and front office. It is important that the investment teams do not get too far ahead of the fund's organizational abilities; otherwise, the fund will disappoint partners and co-investors. (And it will not be given many second chances.)

To sum up, realizing the advantages of being a long-term investor will inevitably require some level of in-sourcing due to broken agency problems. But this will be challenging to implement since for much of the 20th century institutional investors have been restricted from developing internal capabilities. In our view, overcoming this deficiency will require good governance. Following Clark and Urwin (2008), we find that good governance is crucial to be an effective in-house investor, as it is the primary mechanism to mobilize the resources of the institution. And so, given the dramatic organizational and management changes required to move assets from external managers to in-house teams, it should come as no surprise that a strong and capable governance framework is an absolute prerequisite.

**Innovation and governance**

The key ingredient driving the long-term success of an institutional investor is its governance. As noted above, governance is the mechanism that can mobilize the resources of the institution to realize its objectives. It is also the mechanism that can create new resources (or further develop existing resources). The independence of the organization, its resources and systems, and the ability to identify areas of opportunity, as well as challenges, are all crucial elements of governance that can dramatically impact the success of any in-sourcing policy. So, before moving assets in-house, institutional investors should first...
assess their governance capabilities to determine whether a given investment strategy is commensurate with its organizational capabilities.

The idea is to root the fund’s investment style and strategy in a clear assessment of the organization’s capabilities and the resources at the disposal of staff. Similar to using a “risk budget” to guide portfolio construction and investment decision-making, institutional investors should use a “governance budget” to guide the process of investing, the development of the organization and the guidance of management [Clark and Urwin (2010)].

A good board will recognize the unique nature of the financial services business and compensate internal staff accordingly. The board will defend the independence of the organization, which also helps when trying to compensate staff at rates far exceeding traditional government pay scales. In-sourcers have to streamline investment decision-making and be ready to move at the pace of transactions. For example, this may require public pension fund boards to break from calendar time meeting schedules and be prepared to meet in real time, which requires a high governance budget.

Developing a governance budget first requires cataloging the resources and assets that drive sustainable returns. In general, these include the talent and skills of portfolio managers, the processes and protocols of decision-making, and the information processing tools that support judgments. In other words, it is the people, processes, protocols and systems that seem to matter most in this business [Clark and Monk (2012)]. And where do the people, process, protocols and systems ultimately come from? It is the board that makes the resourcing decisions; it is the board that controls the strategic levers of success. A fund’s governance budget, which ultimately refers to the resources available to the fund to build the investment operations, is thus a crucial determinant of an investor’s capacity to innovate and has clear relevance to in-sourcing.

Governance budgets comprise three ingredients: 1) the amount of time that a fund’s board can apply to a given investment problem, 2) the level of expertise that can be called upon at the level of the board, and 3) the commitment of the board, which refers to its effectiveness at getting things done; this is the dynamic capabilities of the board (e.g., real-time meetings versus calendar time meetings). We can look at each of these as being scarce resources that can be drawn down as a fund engages in more innovative or risky behavior. Thinking about governance budgets, then, should become just as important as thinking about asset allocation or manager selection, especially for funds interested in bringing assets in-house.

In summary, any fund considering a policy to move assets in-house should first ensure that their governance budget aligns with their new risk budget. If the levels of governance required exist, then the fund can begin to develop the organization in accordance with the principles and policies below. In order to achieve the level of governance required, we think it prudent to first examine the nomination procedures of boards. It is reasonable to argue that the single most important factor driving the success or failure of an institutional investor over the long run is the procedures used to nominate board members. Ideally, these procedures should prioritize commercial, financial and entrepreneurial expertise over political or stakeholder affiliations.

### Principles and policies for success

Moving assets in-house resolves a variety of principal-agent problems. However, it also raises an entirely different set of operational challenges for institutional investors and, in particular, the funds’ governing boards. In this section, we distill the lessons learned from our 20 case studies and set out the principles and policies that boards should focus on when considering (or managing the process of) moving assets in-house. Readers will notice that we develop these principles in a deliberate manner such that the foundational elements are presented first. Subsequently, the more advanced elements are raised. Said differently, we think effective in-sourcing requires a foundational approach that starts, at the base, with good governance and then adds additional principles and policies in a cumulative manner. Indeed, among our cases each additional element seemed to build on the previous element, which gives the impression that a sophisticated in-sourcer is developed like a pyramid (Figure 1). In short, boards should focus on the following characteristics and focus on them in the following order.

---

2 Here it is important to recognize that governance is a finite and measurable resource, and the investment style and strategy of the organization should match the available levels of governance. Said slightly differently, a fund’s governance budget should implicitly be reflected in its planned and expected performance and explicitly be reflected in the fund’s risk budget.
Fundamental attributes: the following three elements are deemed to be foundational among in-sourcers:

1. **People:** recruitment and retention of talented individuals is the single most important factor driving the success of any investment organization [Bertram and Zvan (2009); Ambachtsheer (2011)]. And so, when it comes to institutional investors managing assets internally, the attraction and retention of people is a prerequisite. It should come as no surprise, then, that our cases all had sophisticated human resources policies and teams. Specifically, the funds tried to combine a competitive salary (near the mean for the specific industry or asset class) with some compelling extra-financial characteristics (e.g., mission, location or responsibilities). In addition, the funds often targeted their recruiting efforts on those individuals for whom the investor had an inherent competitive advantage in hiring, such as those with ties to the region [Bachher and Monk (2012)].

2. **Organization:** internal asset management requires building out much more than an investment office; it requires the entire organization be improved. For example, successful internally managed pension funds require nearly two people in the back office for every one person in the front office [MacIntosh and Scheibelhut (2012)]. In-house investors also have to be able to analyze (and be confident in) their data, which requires a data warehouse with the ability to aggregate and normalize data with a view to presenting it to key decision-makers in a manner that can be interrogated (ideally in real time). This will require expensive and customized data infrastructure.

3. **Risk management:** by taking responsibility for end-to-end investment of assets, the fund will have to develop sophisticated risk management capabilities. These can be broken down into financial and operational risk management:
   
   - **Financial:** investors are going to have to develop the systems and protocols for managing financial risks. Among our cases, this often was done with a combination of traditional financial models as well as informal models and metrics to get a “real” sense of exposure. This may require a dedicated team led by a Chief Risk Officer, as well as sophisticated data visualization systems that can allow the senior management to slice portfolios according to risk exposures.
   
   - **Operational:** all of our cases focused a great deal on compliance, control and “fail” prevention. Some funds attempted to build redundancy into almost everything they do. This was deemed necessary to prevent “unwanted events” that could delegitimize the organization in the eyes of sponsors. And yet, a major downside with this level of (attempted) control and compliance is the increased hierarchy and bureaucracy, which results in a prolonged decision-making process for strategic changes to the fund.

   In order to minimize bureaucracy, we encourage funds to map operational risks and set “tolerance levels” for losses to help guide behavior of risk managers.

Intermediate attributes: Building on the above factors, culture, asset selection and mandates offer boards important touchstones in developing an in-house team:

4. **Culture:** as a fund moves from external to internal asset management, a cultural transformation is inevitable (and necessary). In-sourcers need to promote a culture of responsibility, accountability and high performance. It is...
also important that people in the organization speak the same language. This will ensure mutual understanding and help to assess competing opportunities. The board and the management team have to be willing to take some personal career risk in this process too. Remember that outsourcing to asset managers provides pension and sovereign fund staff a sort of political and career risk insurance policy, so setting up an in-house team will mean giving up this comfort level and getting involved in the challenging and sometimes painful business of investing. Note that building a cohesive culture may require keeping all staff local for the first few years, which may be deemed contrary to the collection of local information in foreign markets.

5. Asset selection: in-sourcing strategies should be conceived in relation to an institution’s specific characteristics. In other words, the investor should first think about what is actually feasible within a given institutional environment and then tailor the strategy to that environment. Investors should obviously exploit their endowments when they bring assets in-house, as they are, in effect, signaling that they can outperform external providers. Here we think location, liability profile and resourcing are paramount considerations. For example, some funds see their size and time horizon as being competitive advantages that allow for strategies and investments that extend far beyond what the private sector is capable of doing. In these cases, internal teams might focus on illiquid assets, such as infrastructure. In other cases, some funds look to start with local, small and low-risk assets. These funds try to build legitimacy with their boards to be able to bring a growing proportion of the assets in-house over time. Investors must consider where they are physically located before launching an in-sourcing policy that requires special skills or access. As Spence (2002) notes, firms should outsource those functions that require high levels of specialization, while in-sourcing makes sense when information and transaction costs are high.

6. Mandates: many funds with successful internal programs think of internal mandates in much the same way as they do external mandates. In large part, this helps to ensure the in-house mandates are adding value. So, for example, each in-house mandate would go through a similar documentation and oversight process as an external mandate. In addition, by treating in-house and external teams similarly, the fund can be sure that developing an internal team is worth the effort (and headache). However, the in-sourcer should also recognize that this can go too far, as one of the benefits of running the mandates in-house stems from flexibility and the ability to deviate from time to time.

Advanced attributes: the following aspects of in-sourcing will help boards sustain in-house operations:

7. Delegation and segregation: for the internal teams to be able to do their jobs effectively, they will require formal delegation of responsibilities from the board (and management). In principle, this delegation of authority should include a variety of different delegated bands of authority, depending on the size of investment and asset class. Some key duties and functions should also be segregated among employees to ensure rules are being respected and indeed followed.

8. Communication: institutional investors that in-source need to be proactive about communications with stakeholders. It is important that everyone — the board, the sponsors, and even employees and plan members — truly understand what the fund is doing and why it is valuable and required. This will help build legitimacy for the policies and ensure their proper resourcing. Moreover, once the fund takes responsibility for deploying the capital, stakeholders will undoubtedly challenge losses generated by internal teams vociferously. This can lead to conservatism due to perceived career risks. Instead, management and the board need to get out ahead of the markets by articulating the strategies and their risk return profiles.

9. Networks: investors that bring assets in-house are, in effect, giving up many of the agglomeration economies enjoyed by asset managers. These investors thus have to find ways to put themselves at the center of deal flow and investment opportunities, which can be a challenge for funds situated far from financial centers. One mechanism for sourcing deals and tapping into local knowledge is to develop relationships with like-minded investors around the world. A direct investor should thus look to “network economies” within its peer group as a (potential) replacement for the “agglomeration economies” they are giving up on Wall Street or London.
Principles and policies for in-house asset management

In sum, the fundamental, intermediate and advanced attributes described above represent the nine key factors deemed to be at the core of the success of any in-house investment project. And, moreover, in order to develop these nine characteristics the fund has to have a tenth: good governance.

Implications and conclusions

A growing number of institutional investors are working to bypass traditional financial intermediaries and agents through the development of in-house teams of investment professionals. Research shows that this path to financial markets, if executed well, offers a boost to long-term performance. In large part, this is about realizing the advantages of being a long-term investor and overcoming broken agency problems. Notwithstanding the benefits, there are significant challenges and pitfalls. In fact, for many investors, it requires a complete reorganization of the business of institutional investment. At issue, then, is the degree to which the broader community of long-term institutional investors will be able to change their organizations to be successful as in-sourcers and long-term investors.

Drawing on 20 in-depth case studies of in-sourcing institutional investors from around the world, we develop a series of principles and policies to guide boards in their in-sourcing journey. At the highest level, we conclude that governance is the key touchstone to a successful in-house investment capability. In our view, the board holds the strategic levers of success for institutional investment organizations. As such, it is the board that will help develop the following nine elements of in-house asset management:

- **People:** the board has to understand the importance of people and be willing to offer compensation packages sufficient to get the talent required, keeping in mind that even generous compensation paid to internal teams will almost always be dwarfed by the compensation routinely paid to external managers.
- **Organization:** a board must have final authority over the allocation of resources and work to develop all aspects of the organization.
- **Risk:** given that investors are little more than risk managers, operational and financial risk management is crucial. Note that the risk management function should not prevent the dynamic capabilities that can lead to success.
- **Culture:** developing in-house asset management capabilities requires creating a culture of risk-taking that is not present in most externally oriented institutional investors.
- **Assets:** no institutional investor can or should do everything in-house. The costs of complexity are all too apparent in some of the world’s largest funds. This means that the boards of institutional investors need to be careful in selecting those activities for in-house provision that are both cost-effective to produce and complementary in their relationships to one another. Otherwise, outsourcing is the best option.
- **Mandates:** internal mandates should be conceptualized, sold and launched with the same rigor demanded of external managers.
- **Delegation:** successful in-sourcing requires delegation to experts and segregation of duties to ensure that these experts are being held accountable for their actions.
- **Communication:** a long-term investor with in-house operations has to be proactive with its stakeholder outreach to ensure the ongoing legitimacy of the fund, even when markets are not cooperating.
- **Networks:** successfully executing an in-house strategy almost certainly requires developing a network of like-minded peers to share opportunities and experience.

As this article highlights, the path to move asset in-house is a long and difficult one. As such, it is important to be extremely selective in the areas where the fund looks to add value through in-sourcing and be humble about what is possible. Institutional investors considering internal mandates should always look to the market first to see if the service can be purchased at a lower price (with comparable returns) than the internal teams. If a fund can do no better than anybody else, it should outsource it to those people and focus efforts on something else.
References
Why foreign life insurers did not achieve their ambitions in China: Structural and operational obstacles

Maoqi Wang
School of Economics and Management, Tsinghua University

Bingzheng Chen
Professor of Finance, School of Economics and Management, Tsinghua University

Joseph Qiu
Senior Portfolio Manager, Global Property, AIG

Abstract
China’s economy has experienced rapid growth in the last 30 years. A highlight is in the growth of its insurance industry, whose premiums increased at an annual rate of about 30%. Naturally, people would think that the foreign insurance companies who entered the Chinese market during this period should have achieved impressive performances, especially given that they are competing with the domestic insurance companies that do not have the advantage of experience and technology. However, the reality is that foreign insurers, especially foreign life insurers, often found it difficult to expand their market shares in this emerging market. Many foreign life insurers experienced frequent management turnover and some even withdrew from the Chinese market. This paper aims to analyze this interesting market phenomenon from two different perspectives: (1) structural obstacles, including the regulatory environment and the influence of the recent financial crisis and (2) operational issues faced by foreign life insurers, including operational efficiency, shareholding structure and distribution channels. Based on these analyses, we offer several suggestions to the market participants and regulators.
Why foreign life insurers did not achieve their ambitions in China: Structural and operational obstacles

Introduction
Many economists forecast that China is on track to turn itself into the world's largest economy in a few decades. Such development needs a solid insurance industry to support the growth and provide stability. To accommodate this need, China's insurance industry has gone through dramatic changes in the last 30 years, essentially developing itself from a single-player market to the current situation where 83 domestic insurers and 47 foreign insurers were operating and competing in this "promising land" in 2012. Many foreign insurers consider the Chinese market a target of their strategic expansions with the ambitions to establish another growth engine. Their ambitions are not unreasonable, given that the Chinese insurance market achieved about 30% annual premium growth in the last 30 years.

Most people, including economists, anticipate that the entry of foreign insurance companies should be beneficial. These foreign insurers generally have more experience, better technology, and well-established management teams. A study by Leverty et al. (2009), using dates between 1999 and 2004, found that productivity improvements among insurers in China was related to the entry of foreign insurers. Other studies, including Shen (2000), Sun (2003), D’Arcy and Xia (2003), Guan (2003), Xu and Li (2003), and Wang et al. (2006), also argued that foreign insurers have made a positive impact on the Chinese insurance market.

Although the overall Chinese insurance market has been improving, many foreign insurers, especially life insurers, did not achieve their initial goals in this country. This is reflected in their inability to increase market share, frequent management turnover and the fact that some foreign insurers even withdrew from the Chinese market. Although it is a rather important and interesting market phenomenon that is unique to China, we are not aware of any research that thoroughly analyzes the economic reasons behind this trend. This paper investigates the structural obstacles, including regulatory constraints and the recent financial crisis, and the operational challenges, including operational efficiency, shareholding structure and distribution channel, faced by foreign life insurers. We trust that these can at least partially explain the foreign life insurers’ situation in China and will benefit both researchers and practitioners to further understand China's insurance market.

Historical development and current challenges
Opening up of China's insurance industry
China's insurance market has achieved impressive growth in the last three decades. Total premiums grew from RMB460m in 1980 to RMB1,434b in 2011, which implies a nominal annual growth rate of 29.63%. In 2011, life insurance accounted for 67.8% of all premiums. Although China has become the sixth-largest insurance market in the world (measured by total premiums), it is still considered to be undeveloped. Its insurance density and penetration rate are among the lowest in the world.

The rapidly growing market and the great future potential have attracted the attention of many global insurance companies. However, it did take many years for China to open up its insurance market to foreign insurers. In 1980, the American International Group (AIG) created a representative office in Shanghai. However, according to the Provisional Regulations Governing the Administration of Foreign Insurance Companies, foreign insurers were faced with significant challenges when entering the market. In 1994, foreign insurers were allowed to set up wholly-owned subsidiaries in China. However, due to various regulatory constraints, many foreign insurers encountered difficulties when trying to establish a presence in the market. This paper aims to shed light on the structural obstacles and operational challenges faced by foreign insurers in China.

1 According to CIRC, there have been 42 domestic and 26 foreign life insurers, as well as 41 domestic and 21 foreign non-life insurers in China through the end of 2012. See http://www.circ.gov.cn/web/site0/tab61/.

2 New accounting standards were implemented from December 2010, in which the measurement of premium income on life and health insurance was no longer the same as before. So, the growth rate of premiums was not comparable since then.


4 The data are from “The World Insurance” by SwissRe Sigma, 2012.03.
Insurance Enterprises, foreign insurers’ equity participation in the Chinese insurance industry was still completely banned. In 1992, Shanghai was selected by the Chinese Government to test the entry of foreign insurers. Soon after that, American International Assurance (AIA) received the first insurance license ever issued to any foreign insurer since 1949, but the geographic scope of its business was limited to Shanghai. In 1995, American International Underwriters (AIU) set up the first foreign P&C branch in Shanghai. Before 1998, 113 applications for insurance licenses filed by foreign insurers were declined (Leverty et al. (2009)). Only 10 joint ventures and branches of foreign insurers obtained licenses to write insurance coverage in China. Even these 10 were strictly limited in what insurance they could write, where it could be written, what rate they could charge, and what kind of shareholding structure they could have. As a result, foreign insurers only represented less than 1% of the Chinese insurance market in 1998.

However, with the growth of China’s economy, insurance demand increased as well. After China signed the WTO agreement in 2001, the Chinese Government launched an official schedule to open its insurance market to foreign players. Restrictions on locations, products, reinsurance and ownership were expected to be removed gradually. A three-year plan was created according to the WTO agreement (see Table 1).

The rest of the paper will focus solely on life insurance because: (1) life insurance and non-life insurance have fundamentally different business models and therefore cannot be mixed together, (2) the institutional restrictions on foreign non-life insurers are different and more severe and (3) foreign non-life insurers have not been able to achieve a meaningful market share in China (1.06% in 2010).

Development of foreign life insurers in China
Before we study the development of foreign life insurers in China, we need to clarify the definition of “foreign.” According to Li (2007), foreign investors mainly use two approaches to enter the Chinese insurance market, namely, participation and establishment. In the case of participation, the foreign company invests in the equity of an existing Chinese insurance company as a minority shareholder, less than 25%. When foreign insurers decide to go the establishment route, which means holding an ownership stake of in excess of 25%, they set up independent business operations through a joint venture, a wholly owned operation or a branch.

Using statistics provided by the China Insurance Regulatory Commission (CIRC, China’s insurance regulation authority), Table 2 presents information about the number of foreign life insurers,
Why foreign life insurers did not achieve their ambitions in China: Structural and operational obstacles

their market shares and ownership categories from 1992 to 2010. Although the number of foreign life insurers increased rapidly, those who entered the Chinese market post 2001 were required to adopt a 50/50 shareholding structure with a Chinese partner.

When foreign life insurers initially entered the Chinese market, they were confident and ambitious. A survey conducted in 2008 by a major accounting firm indicated that foreign life insurers’ executives were expecting a 10% market share by 2011. However, the situation has changed in more recent years. As shown in Table 2, total market share of foreign life insurers stopped climbing in 2008 and has been hovering around 5% since then. The ownership category also indicates that several foreign life insurers changed their status from “foreign majority” or “50/50” to “Chinese majority,” meaning that they gave up their controlling interests. As a result, the confidence level was reduced as well. A similar survey conducted in 2010 found that executives of foreign life insurers did not anticipate any further increases in their Chinese market share, representing a dramatic change in their ambitions from 2008.

Challenges for foreign life insurers in China
Underperformance in market share
As shown in Table 2, the market share of foreign life insurers stopped its upward trend in 2008 and has remained around 5% since then. It makes sense to compare this data with other major insurance markets in the world. Figure 2 shows foreign insurers’ market share in China and OECD countries. It is rather clear that foreign insurers in China (both life and non-life) have failed to gain as much market share as they did in other countries. For example, foreign insurers have a 30% life market share and a 50% non-life market share in the U.K. In the U.S., foreign insurers have a 22% life market share and a 12% non-life market share. Even in other Asian countries, such as Japan and South Korea, foreign insurers could obtain 20% in the life insurance market.

Furthermore, leading insurance companies’ market share performance in China is well below their global performance, as is shown in Table 3.

5 In fact, the OECD database defines foreign insurers as: insurers with foreign capital as the major shareholder, or branches of foreign insurers. So, the market share data that we used for the other OECD countries were actually underestimated compared to the CIRC standard, which suggests a larger gap between China and other OECD countries.
The marker share issue is reflected not only by foreign life insurers’ historical performances in China, but also by the ongoing trend of their slow expansion in China’s regional business. Table 4 contains a comparison between a number of recently established domestic life insurers and foreign life insurers in terms of how many new provincial branches were created in a year. One can easily see that foreign life insurers expanded their regional business at a much slower rate than their Chinese counterparts.

Frequent management turnover
In addition to underperformance in terms of market share, another interesting market phenomenon is foreign life insurers’ frequent management turnover in China. This is further highlighted when they are compared with their domestic competitors. For example, Mingzhe Ma has led the China Pingan Insurance Group since 1994 and Dongsheng Chen has led the Taikang Insurance Company since 1996. There is no question that these fast-growing domestic life insurers benefit from their stable management teams.

On the other hand, change of senior management of the life insurers in China is not unusual. For example, a major joint venture between a foreign insurer and a Chinese corporation has changed its chief executives three times since 2007. Some observers argue that this was due to the unsatisfactory performance of the company in China, while others think that it was probably due to the differences between the domestic and foreign partners over strategies. Although we cannot prove that foreign life insurers should not change senior management that frequently, it is not difficult to imagine that frequent management turnover does present challenges for the companies in carrying out strategies and achieving long-term success.

Ownership decline
The life insurance equity transactions observed in China from 2009 to 2011 have attracted attention. These transactions are listed in Table 5. We can see that some foreign life insurers reduced their shareholding in joint ventures; some even withdrew from the Chinese market completely. These transactions provide additional evidence that foreign life insurers have been encountering difficulties in the Chinese market.

Structural obstacles
Regulatory environment
Entry barriers and regulatory restrictions have been widely blamed for causing difficulties for foreign insurers in China. However, few credible analyses with solid data exist to support this argument. In this study, we employ the International Association of Insurance Supervisors (IAIS) database to investigate the impact of insurance regulation.

Using IAIS’ Insurance Laws Database, we compiled in Table 6 a comparison of licensing, scope of business, shareholding restriction, and the average market share of foreign life insurers in a number of countries.

Figure 2: Foreign insurers’ market share comparisons: China and OECD countries (2008)
Source: OECD database

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AIG</td>
<td>3.60%</td>
<td>1.49%</td>
<td>0.78%</td>
</tr>
<tr>
<td>Metropolitan Life</td>
<td>3.20%</td>
<td>0.01%</td>
<td>0.23%</td>
</tr>
<tr>
<td>AXA</td>
<td>3.30%</td>
<td>0.05%</td>
<td>0.10%</td>
</tr>
<tr>
<td>Allianz</td>
<td>3.00%</td>
<td>0.06%</td>
<td>0.15%</td>
</tr>
<tr>
<td>ING</td>
<td>2.50%</td>
<td>0.04%</td>
<td>0.12%</td>
</tr>
<tr>
<td>Generali</td>
<td>2.50%</td>
<td>0.10%</td>
<td>0.58%</td>
</tr>
<tr>
<td>Aegon</td>
<td>2.30%</td>
<td>0.02%</td>
<td>0.16%</td>
</tr>
<tr>
<td>Aviva</td>
<td>2.00%</td>
<td>0.03%</td>
<td>0.47%</td>
</tr>
<tr>
<td>Manulife</td>
<td>1.70%</td>
<td>0.19%</td>
<td>0.18%</td>
</tr>
</tbody>
</table>

Table 3: Major catastrophe reinsurance pricing approaches
Source: CIRC website and Sigma 2006, Swiss Re

6 See news reports, such as http://insurance.zgjrw.com/News/2011414/Insurance/358619383400.shtml.
Why foreign life insurers did not achieve their ambitions in China: Structural and operational obstacles

Table 4: Provincial branches establishment of some domestic and 50-50 foreign life insurers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Union Life (2005)**</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jiahe Life (2005)</td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PICC Life (2005)</td>
<td>1</td>
<td>4</td>
<td>16</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nissay-Greatwall Life (2005)</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreign life insurers***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cathay Life (2004)</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Aegon-CNOOC (2003)</td>
<td></td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Heng An Standard Life (2003)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skandia Life (2004)</td>
<td></td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Chongsheng Life (2003)</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIGNA &amp; CMC Life (2003)</td>
<td></td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sino-French Life (2005)</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Samsung Air China Life (2005)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sino-US MetLife (2004)</td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Great Eastern Life (2006)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Some equity transactions of foreign insurers in China over 2009–11

<table>
<thead>
<tr>
<th>Current company name</th>
<th>Year of foundation</th>
<th>Equity transaction descriptions</th>
<th>Time of transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun Life Everbright Life Insurance Co., Ltd.</td>
<td>2002</td>
<td>The company was founded in 2002, with two major shareholders, the China Everbright Group and Canada Sun Life Financial, each holding 50% of the total shares. The China Everbright Group injected more capital in 2009, reducing the share of Sun Life Financial Services to 20%, and turned the company into a domestic one.</td>
<td>2009.07</td>
</tr>
<tr>
<td>Tian An Life Insurance Co., Ltd.</td>
<td>2001</td>
<td>The company was founded in 2001, with two major shareholders, Manulife Group and the China Tianan Group, each holding 50% of the total shares. Manulife Financial withdrew all its capital in 2009, leaving the Tianan Group holding all the shares, and the company turned into a domestic one.</td>
<td>2009.09</td>
</tr>
<tr>
<td>ICBC AXA Life Insurance Co., Ltd.</td>
<td>1999</td>
<td>The company was founded in 1999, and the AXA Group held 51% of the shares initially. In 2010, the Industrial and Commercial Bank of China made a deal with the AXA Group and the Metlife Group by purchasing 60% of the company's shares, leaving 27.5% for the AXA Group and 12.5% for the Metlife Group.</td>
<td>2010.10</td>
</tr>
<tr>
<td>BoComLife Insurance Co., Ltd.</td>
<td>2000</td>
<td>The company was founded in 2000, with two major shareholders, China Life and the Commonwealth Bank of Australia, holding 51% and 49% of the shares respectively. In 2010, the China Bank of Communications purchased all the shares from China Life and injected more capital to hold 62.5% of the company's shares, leaving the Commonwealth Bank of Australia with 37.5%.</td>
<td>2010.01</td>
</tr>
<tr>
<td>ING BOB Life Insurance Co., Ltd.</td>
<td>2002</td>
<td>The company was founded in 2002, with two major shareholders, the Beijing Capital Co. and the ING Group, each holding 50% of the shares. In 2010, the Bank of Beijing purchased all the shares from the Beijing Capital Co., and changed the name of the company from ING Capital Life Insurance to the ING BOB Life Insurance Co.</td>
<td>2010.04</td>
</tr>
<tr>
<td>Founder Meiji Yasuda Life Insurance Co., Ltd.</td>
<td>2002</td>
<td>The company was founded in 2002, with two major shareholders, the Haier Group and New York Life Insurance, each holding 50% of the total shares. In 2010, New York Life Insurance withdrew from the company, leaving the Haier Group as the major shareholder. After another equity transaction in 2012, the company is now called Founder Meiji Yasuda Life Insurance.</td>
<td>2010.12</td>
</tr>
<tr>
<td>CCB Life Insurance Co., Ltd.</td>
<td>1998</td>
<td>The company was founded in 1998, with two major shareholders, the China Pacific Insurance Group and the ING Group, each holding 50% of the shares. In 2010, the China Construction Bank (CCB) acquired the 50% shares from ING group, turning the company into a domestic insurer. In 2011, the CCB, as well as other strategic partners (such as China Life Taiwan and China Jianyin Investment Co.), took over the other 50% shares from the China Pacific Insurance Group, while CCB became the majority shareholder.</td>
<td>2011.03</td>
</tr>
</tbody>
</table>

Table 4: Provincial branches establishment of some domestic and 50-50 foreign life insurers

*Data of domestic life insurers are from the China Insurance Yearbook 2006–11.

**Number in parentheses is the foundation year of the company.

***Data for foreign life insurers are from the CIRC website, published by the International Department for Authorization Announcement.

---

8 All the companies were founded around 2005, so the establishment of new branches in the same year is comparable.
of countries. We can see that the major difference in the regulatory environment of China, as compared to OECD countries, is that it has a shareholding cap of 50%. This resulted in the “50/50” shareholding structure that has been adopted by many foreign life insurers in China, which has further led to other operational issues, such as high management turnover and low efficiency.

In addition to the direct impact on the shareholding structure, the regulatory environment in China also has an indirect impact on the foreign insurers. Using reinsurance as an example, the Chinese market experienced a “statutory reinsurance period” from 1985 to 2006, a “priority reinsurance period” (from 2006 to 2009) and finally entered the “free competition period” (since 2009). However, even in the free competition period, China Re, the traditional market dominant owned by the Chinese Government, has an obvious advantage in client relationship, customer resources, and operational data. Although we admit that it is not possible to quantify these indirect impacts, they do exist and deserve market participants’ attention.

Financial crisis
The recent financial crisis significantly shrunk the global equity market, and therefore created additional constraints for insurers to develop in the emerging markets. Many foreign markets faced capital constraints and were even bailed out by their governments. As a result, some were forced to revise their global strategies and move resources from emerging markets back to their core markets.

Table 6: Comparisons of the supervision environment for foreign life insurers

<table>
<thead>
<tr>
<th>License of doing business</th>
<th>Maximum shareholding for foreign investors</th>
<th>Limits before notification of intended participation in an insurance company</th>
<th>Average market share of foreign life insurers (2006-08)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreign companies through a branch</td>
<td>Foreign companies by way of cross border business</td>
<td>Are foreign insurers entitled to underwrite all risks?</td>
<td>For a single investor</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Mexico</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Poland</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Turkey</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Italy</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Australia</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Netherlands</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>U.S.</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Sweden</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Japan</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Austria</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Germany</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Korea</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Canada</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Denmark</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Switzerland</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Spain</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>China</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
</tbody>
</table>

Table 6: Comparisons of the supervision environment for foreign life insurers

Source: IAS Insurance Law Database, tabulated by the authors

9 In this period, 20% of the non-life insurance premiums had to be ceded to reinsurers, which mainly referred to China Re.
10 In this period, 50% of the reinsurance business needed to be ceded with priority given to at least two domestic reinsurers. China Re, of course, received most of the business.
As Figure 3 indicates, the equity capital of life insurance companies declined by more than 25% in 2008. Apart from asset loss, two aspects of the life insurance business were also hit by the financial crisis – variable annuities and securities lending. Variable annuities businesses with equity market guarantees suffered from the falling equity markets when regulators asked these insurers to post reserves against the rising market value of their guarantees.

Securities lending businesses caused illiquidity problems for the insurers since they invested the collateral in high-risk assets before the crisis. Since these assets could not be sold easily during the financial crisis, some life insurance companies had difficulty meeting the liquidity requirement.

Given the pressure from the financial crisis, it was not unreasonable for foreign life insurers to reduce their shareholding in China. After all, these companies failed to deliver satisfactory results in the Chinese market. When they desperately needed cash to boost liquidity, selling the shares in China became a natural option.

Operational challenges
The structural obstacles, such as the regulatory environment and financial crisis, are external factors that impacted the development of foreign life insurers in China. In addition, there are internal issues that can also explain the current challenges faced by foreign life insurers. These internal issues include operational efficiency, shareholding limitation, and distribution channel.

Operational efficiency: the DEA model
The nature of the life insurance business determines that it is difficult to generate immediate returns from initial investments. In other words, it is common to expect no profits at the early stages of development. However, management decision is driven by shareholders’ interest and investors’ need to see how their investment is performing. Consequently, efficiency analysis that essentially measures the output-input ratio becomes an important topic. Logically, if the business operation is not efficient, then investors would want to withdraw their investment.

Chen et al. (2007) investigated the efficiency performance of domestic and foreign life insurers in China. They found that, in contrast to the common view, domestic insurers possess advantages in terms of both pure technical efficiency and scale efficiency. This suggests that maintaining growth is an important strategy for foreign life insurers that wish to compete in the Chinese market.

This paper analyzes efficiency from a different perspective. Our primary interest lies in companies’ decisions to leave the Chinese market. We, therefore, think that premium income and profits are appropriate output indicators. Compared to realized benefits, additions to reserves and invested assets used by Chen et al. (2007), the current approach fits the context of business decision-making.

Methodology and data
We use the Data Envelopment Analysis (DEA) model developed by Banker et al. (1984) to test the operational efficiency of life insurers in China. The model is based upon the goal programming below:11

\[
\begin{align*}
\text{Min} \{ & \theta \cdot e(S' + eS') \} \\
\text{s.t.} & \sum_{i=1}^{n} \lambda_i x + S = \theta x_c \\
& \sum_{i=1}^{n} \lambda_i y = y_c \\
& \lambda_i \geq 0, j = 1, \ldots, n \\
& S' \geq 0, \text{and} S \geq 0
\end{align*}
\]

11 For more references and details about DEA and the model, please refer to Chen, Powers and Qiu (2009).
In our study, number of employees (X1), equity (X2), and expenses (X3) are used as input indicators; premium income (Y1) and profit (Y2) are used as output indicators. Inputs are in general agreement with insurance efficiency research and the selection of output indicators is based on the discussion above. Our dataset includes company-level data from 2001 to 2010, obtained from the China Insurance Yearbook. Summary statistics of input and output indicators are provided in Table 7.

### Analysis results

The power of the DEA model is reflected in its ability to distinguish pure technical efficiency from scale efficiency. Mathematically,

\[
\text{Technical efficiency} = \text{pure technical efficiency} \times \text{scale efficiency}
\]

Technical efficiency measures a firm’s overall efficiency performance, pure technical efficiency captures the efficiency of resource allocation and internal management, while scale efficiency reflects the degree to which the firm is operating at an optimal level. Average efficiency scores from domestic and foreign life insurers in China are presented in Table 8.

We observe that foreign life insurers in China exhibit lower efficiency than their domestic competitors. This is especially true for scale efficiency. Even for pure technical efficiency where foreign players should have the competitive advantage, domestic life insurers’ performance is slightly better.

### Table 7: Summary statistics of input and output variables

<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>
</tr>
</thead>
<tbody>
<tr>
<td>Number of companies</td>
<td>51</td>
<td>49</td>
<td>45</td>
<td>45</td>
<td>38</td>
<td>34</td>
<td>27</td>
<td>25</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>Number of domestic companies</td>
<td>27</td>
<td>24</td>
<td>21</td>
<td>21</td>
<td>15</td>
<td>12</td>
<td>19</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Number of foreign companies</td>
<td>24</td>
<td>25</td>
<td>24</td>
<td>23</td>
<td>23</td>
<td>22</td>
<td>8</td>
<td>17</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Number of employees (X1)</td>
<td>Mean 6,944</td>
<td>6,025</td>
<td>7,623</td>
<td>6,568</td>
<td>5,961</td>
<td>6,104</td>
<td>6,256</td>
<td>5,684</td>
<td>8,078</td>
<td>6,648</td>
</tr>
<tr>
<td></td>
<td>Median 1,056</td>
<td>1,151</td>
<td>750</td>
<td>621</td>
<td>438</td>
<td>290</td>
<td>236</td>
<td>954</td>
<td>890</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum 103,868</td>
<td>105,111</td>
<td>102,777</td>
<td>96,786</td>
<td>77,318</td>
<td>75,728</td>
<td>75,437</td>
<td>72,900</td>
<td>65,232</td>
<td>51,880</td>
</tr>
<tr>
<td></td>
<td>Minimum 11</td>
<td>62</td>
<td>33</td>
<td>47</td>
<td>39</td>
<td>75</td>
<td>42</td>
<td>63</td>
<td>54</td>
<td>56</td>
</tr>
<tr>
<td>Equity (X2), in million RMB</td>
<td>Mean 2,517</td>
<td>6,297</td>
<td>3,791</td>
<td>1,251</td>
<td>833</td>
<td>2,579</td>
<td>3,203</td>
<td>3,174</td>
<td>1,173</td>
<td>1,318</td>
</tr>
<tr>
<td></td>
<td>Median 1,056</td>
<td>1,151</td>
<td>750</td>
<td>621</td>
<td>438</td>
<td>290</td>
<td>236</td>
<td>954</td>
<td>890</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum 33,654</td>
<td>212,776</td>
<td>135,881</td>
<td>17,189</td>
<td>9,957</td>
<td>62,951</td>
<td>62,436</td>
<td>13,269</td>
<td>7,941</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum 5</td>
<td>17</td>
<td>12</td>
<td>2</td>
<td>6</td>
<td>56</td>
<td>12</td>
<td>23</td>
<td>122</td>
<td>16</td>
</tr>
<tr>
<td>Expenses (X3), in million RMB</td>
<td>Mean 2,893</td>
<td>2,480</td>
<td>794</td>
<td>1,166</td>
<td>961</td>
<td>581</td>
<td>1,780</td>
<td>1,579</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Median 2,893</td>
<td>2,480</td>
<td>794</td>
<td>1,166</td>
<td>961</td>
<td>581</td>
<td>1,780</td>
<td>1,579</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum 177</td>
<td>333</td>
<td>257</td>
<td>212</td>
<td>149</td>
<td>97</td>
<td>70</td>
<td>192</td>
<td>166</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum 4</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>Premium Income (Y1), in million RMB</td>
<td>Mean 10,026</td>
<td>6,672</td>
<td>12,166</td>
<td>5,537</td>
<td>9,682</td>
<td>3,831</td>
<td>5,694</td>
<td>13,169</td>
<td>10,465</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Median 666</td>
<td>615</td>
<td>1,495</td>
<td>363</td>
<td>505</td>
<td>236</td>
<td>188</td>
<td>146</td>
<td>240</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Maximum 95,587</td>
<td>71,876</td>
<td>294,939</td>
<td>78,237</td>
<td>183,843</td>
<td>58,849</td>
<td>149,983</td>
<td>58,959</td>
<td>128,781</td>
<td>81,313</td>
</tr>
<tr>
<td></td>
<td>Minimum 0</td>
<td>1</td>
<td>14</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Profit (Y2), in million RMB</td>
<td>Mean 1,002</td>
<td>1,050</td>
<td>71</td>
<td>940</td>
<td>338</td>
<td>150</td>
<td>103</td>
<td>251</td>
<td>-1,606</td>
<td>-52</td>
</tr>
<tr>
<td></td>
<td>Median -16</td>
<td>-2</td>
<td>-173</td>
<td>-34</td>
<td>-56</td>
<td>-38</td>
<td>-51</td>
<td>-14</td>
<td>-32</td>
<td>-16</td>
</tr>
<tr>
<td></td>
<td>Maximum 33,811</td>
<td>33,036</td>
<td>10,205</td>
<td>28,297</td>
<td>9,601</td>
<td>5,456</td>
<td>2,919</td>
<td>5,857</td>
<td>1,825</td>
<td>1,761</td>
</tr>
</tbody>
</table>

Table 8: Average of efficiency results (domestic and foreign)

<table>
<thead>
<tr>
<th></th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic efficiency</td>
<td>0.66</td>
<td>0.82</td>
<td>0.76</td>
<td>0.82</td>
<td>0.71</td>
<td>0.52</td>
<td>0.84</td>
<td>0.89</td>
<td>0.84</td>
<td>0.33</td>
</tr>
<tr>
<td>Pure technical efficiency</td>
<td>0.27</td>
<td>0.78</td>
<td>0.87</td>
<td>0.78</td>
<td>0.78</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>Scale efficiency</td>
<td>0.47</td>
<td>0.69</td>
<td>0.69</td>
<td>0.70</td>
<td>0.60</td>
<td>0.23</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>Foreign efficiency</td>
<td>0.58</td>
<td>0.79</td>
<td>0.79</td>
<td>0.74</td>
<td>0.56</td>
<td>0.70</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>Foreign efficiency</td>
<td>0.58</td>
<td>0.79</td>
<td>0.79</td>
<td>0.74</td>
<td>0.56</td>
<td>0.70</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
</tr>
</tbody>
</table>

The Journal of Financial Perspectives 57
Interestingly, foreign life insurers appeared to perform better in 2007, with technical efficiency higher than domestic companies. However, this momentum failed to continue in 2008. In the following discussion, we found that underdevelopment of the distribution channel was the major cause of this issue.

Low efficiency performance provides an incentive for foreign life insurers to reduce their shareholding in China and even leave the Chinese market. Those foreign life insurers with such equity transactions spent more labor, equity capital and operating expenses but failed to generate sufficient premium and profit for their shareholders. After the equity transactions, we see that efficiency performance of these companies improved (Figure 4).

**Shareholding limitation**
Due to the regulatory restrictions on foreign shareholding, many foreign life insurers adopted a 50/50 shareholding structure when they entered the Chinese market, meaning that the foreign and domestic shareholders each hold 50% of the total shares. In this section, we utilize the efficiency analysis results to investigate whether the 50/50 shareholding structure has a negative influence on foreign life insurers. From Figure 5, we can see that there are significant efficiency differences between those foreign life insurers with a 50/50 shareholding structure and other foreign life insurers, especially in scale efficiency. This concurs with market observers’ criticism that the 50/50 shareholding structure held back the scale development of the companies. According to Fang (2011), under such a structure, the chairman of the board, financial managers and human resources managers are usually appointed by the domestic partner. The foreign partner usually names the chief executive and other management team. This illustrates how the 50/50 shareholding creates potential conflict among the management team and possibly results in low efficiency.

**Distribution channel**
Bancassurance is a very important distribution channel in the Chinese insurance market. According to CIRC, bancassurance accounts for 30% of the premium income (all product lines). Driven by the synergy effect, banks naturally want to serve as strategic partners and shareholders of insurance companies. However, they were not able to do so until the regulation was changed in 2008 allowing banks to hold insurance shares.
After that, banks played an increasingly important role in the insurance equity transactions and mainly served as acquirers.

Table 9 lists the proportion of premium income attributable to bank and postal insurance for each of the foreign life insurers from 2008 to 2010. Insurers that experienced equity transactions are highlighted. Prior to the equity transactions, they generally had poor performance in China and needed bancassurance to boost their business volumes. At the same time, China's domestic banks were looking for insurance partners. This at least partially explains why the equity transactions occurred and the fact that bancassurance of these companies grew rapidly, post transactions.

Conclusions
The importance of the Chinese insurance market is determined by China's economic growth. This market deserves the attention of any ambitious insurance company. In fact, many foreign insurers entered the Chinese market with the goal of establishing their strategic positions in this emerging market and possibly secure another growth engine. However, although it has been widely acknowledged that those foreign insurers have had a positive impact on the Chinese market, the foreign insurers themselves, especially the life insurers, are having difficulties expanding their market share in China. Their competitive advantages in experience, technology and management were not fully realized. Some foreign life insurers have experienced frequent management turnover and some even withdrew from the Chinese market recently. It is important to thoroughly understand this unique market phenomenon and the reasons behind it. Market participants and regulators should adjust their strategies accordingly.

This paper utilizes both quantitative and qualitative analysis to investigate the challenges faced by foreign life insurers. We identify the regulatory environment and the financial crisis as the structural obstacles that impact foreign life insurers' performance externally. Regarding operational challenges, we employ DEA to compare the efficiencies of domestic and foreign life insurers and find that latter are not as efficient as their domestic competitors, especially in terms of scale efficiency. Insurers that reduced their equity shares in recent years exhibit particularly poor efficiency performance. In addition to efficiency, we further study the 50/50 shareholding limitation and distribution channel issues that also negatively impact the performance of foreign life insurers in China.

On the basis of these analyses, it is not surprising to see that foreign life insurers in the Chinese market have become more conservative and lowered their expectations about their future growth potentials. Some prefer to serve as minority shareholders instead of operating the daily business in China. Some have shifted their focus and resources from China to other countries.

In order to improve the current situation, Chinese regulators need to look at the long-term benefit that free market competition can bring to Chinese customers. A more relaxed regulatory environment will promote competition, and therefore provide

<table>
<thead>
<tr>
<th>Name of the company</th>
<th>Proportion of premium income from bank and postal channels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2008</td>
</tr>
<tr>
<td>ING-BOB Life</td>
<td>-</td>
</tr>
<tr>
<td>Chongsheng Life</td>
<td>-</td>
</tr>
<tr>
<td>Shin Kong &amp; HNA Life</td>
<td>-</td>
</tr>
<tr>
<td>Tian An Life</td>
<td>-</td>
</tr>
<tr>
<td>HSBC Life</td>
<td>-</td>
</tr>
<tr>
<td>King Dragon Life</td>
<td>-</td>
</tr>
<tr>
<td>BoComm Life</td>
<td>-</td>
</tr>
<tr>
<td>CCB Life</td>
<td>0.96%</td>
</tr>
<tr>
<td>Manulife-Sinochem Life</td>
<td>0.00%</td>
</tr>
<tr>
<td>Heng An Standard Life</td>
<td>3.97%</td>
</tr>
<tr>
<td>Founder Life</td>
<td>6.61%</td>
</tr>
<tr>
<td>AIA Life</td>
<td>13.53%</td>
</tr>
<tr>
<td>ICBC-AXA Life</td>
<td>31.81%</td>
</tr>
<tr>
<td>Prudential Life</td>
<td>41.55%</td>
</tr>
<tr>
<td>Skandia Life</td>
<td>55.96%</td>
</tr>
<tr>
<td>Sun Life Everbright Life</td>
<td>57.54%</td>
</tr>
<tr>
<td>Aviva-COFCO Life</td>
<td>57.99%</td>
</tr>
<tr>
<td>Skino-US MetLife</td>
<td>58.20%</td>
</tr>
<tr>
<td>General China Life</td>
<td>63.27%</td>
</tr>
<tr>
<td>Cathay Life</td>
<td>66.53%</td>
</tr>
<tr>
<td>CIGNA &amp; CMC Life</td>
<td>70.07%</td>
</tr>
<tr>
<td>Samsung Air China Life</td>
<td>70.49%</td>
</tr>
<tr>
<td>Great Eastern Life</td>
<td>71.18%</td>
</tr>
<tr>
<td>Aegon-CNOOC</td>
<td>76.21%</td>
</tr>
<tr>
<td>Allianz Life</td>
<td>76.22%</td>
</tr>
<tr>
<td>Huatai Life</td>
<td>87.63%</td>
</tr>
<tr>
<td>Sino-French Life</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Table 9: Bank and postal insurance distribution percentages for foreign life insurers in China
Source: China Insurance Yearbook, 2009–11
Why foreign life insurers did not achieve their ambitions in China: Structural and operational obstacles

Incentive for both domestic and foreign insurers to improve their businesses. For example, gradually relaxing the shareholding limitation, streamlining the license approval process and encouraging product innovation are good things to consider. Regulators should also play a role in removing the current obstacles and ensure fair market conditions.

Foreign life insurers operating in China need to employ a customized approach. The unique regulatory environment and market characteristics, such as the importance of bancassurance, must be considered when forming business strategies for the Chinese market. These can be achieved by accumulating more local knowledge, employing more experienced Chinese managers and showing the willingness to deviate from the standard business model. Additionally, the fact that foreign insurers have more experience and better technology does not mean that they have nothing to learn from their domestic competitors. Realistically speaking, foreign insurers need to improve their business volume first in order to realize competitive advantage. Finally, foreign insurers should be patient because regulatory environment and market condition will not change in one day. They should remember that it took AIG many years to establish its current position in China.

References
CIRC, China Insurance Yearbook, 2000-2010
Financial policymaking in the context of the known, the unknown and the unknowable

Richard J. Herring
Jacob Safra Professor of International Banking, The Wharton School, University of Pennsylvania

Abstract
This article examines the challenge of financial policymaking from the perspectives of what is known, unknown and unknowable about the financial system. It focuses on policy challenges during the Great Recession and changes in regulations in response to lessons drawn from the global financial crisis. The article concludes that since many of the risks to the financial system are unknown and some are unknowable, regulators and supervisors need to place much greater emphasis on the resilience of the system. Rather than imposing increasingly prescriptive regulations, a wiser course may be to devise simpler rules that can be implemented and monitored more easily.

1 This essay is based on a book that I co-edited with Frank Diebold and Neil Doherty, The known, the unknown and the unknowable in financial risk management, Princeton University Press, 2010. I am grateful to my co-editors, the authors of chapters in this book, and Ralph Gomory for stimulating many of the ideas in this essay and to the editor of this journal, P. Jackson and A. Clifford for comments on an earlier draft. Some of these ideas were presented earlier in the Yale Journal of Regulation (Herring (2009)).
The financial crisis of 2007–08 has caused a massive reassessment of what we thought we understood about financial markets and financial policymaking. Many of these issues can be usefully examined within a conceptual framework developed in a classic essay by Ralph Gomory (1995). This framework can be applied to financial policymaking by distinguishing the known from the unknown and the unknowable. The “known” (K) refers to a situation where we know the probability of a future event with sufficient certainty that we can specify all the parameters of the probability distribution with a high degree of confidence. This may occur because we have extremely powerful a priori knowledge — for example, a theory of what determines the event about which nearly all experts agree — or because we have data collected over a sufficient range of conditions under which the underlying structure has remained stable so that we can estimate the distribution with a high degree of confidence.

The “unknown” (U) refers to a situation in which we can specify an event, but there is considerable uncertainty about when or whether it may occur because we have a variety of alternative theories without consensus among experts about which is correct, or because we lack adequate data or sufficiently powerful statistical techniques to estimate a distribution with even negligible levels of confidence. The “unknowable” (U) refers to shocks that we have not even identified because they have not happened or because we have no widely accepted theory that leads us to expect they might happen. To some extent, U may simply be the result of a failure of imagination. Much of the history of financial policymaking can be viewed as converting U to U to K, although it would be a mistake to infer that we have enjoyed continuous progress to K. The abrupt transition from the Great Moderation to the Great Recession serves as a recent reminder. The boundaries are not fixed and behavior is dynamic. We may find that K has become u and that an entirely new U has erupted to become an important U. Indeed, if shocks are subject to fractile distributions instead of well-behaved Gaussian distributions, K may contain much more u than is commonly acknowledged [Mandelbrot and Taleb (2010)]. The past is never a perfect predictor of the future. New factors may become important and relationships estimated in times of normal market functioning tend to break down at times of market stress. What was thought to be mild randomness often proves to be wild randomness — or at least more often than it should if it were governed by a well-behaved Gaussian distribution.

In Will Roger’s phrase, one of the key risks may be what we think we know “that just ain’t so.” Undue dependence on ratings by a wide variety of financial market participants preceding the financial crisis illustrates this problem clearly, as did the reliance of regulators on the Basel II capital adequacy regulations to protect the safety and soundness of the banking system.

**Information challenges for financial policymakers**

Financial policy becomes most relevant when a shock that was unknown or unknowable shifts the financial system from the domain of K into the domain of u. Financial policymakers are charged with limiting the vulnerability of the financial system to such shocks and mitigating the consequences of such shocks once they occur. Financial policymakers aim to promote monetary and financial stability, but virtually every aspect of financial policymaking is subject to substantial uncertainty. For example, how precisely should these objectives be defined? With regard to monetary policy, what amount of inflation is consistent with achieving stable, sustainable growth? For that matter, what measure of inflation is appropriate? After the monetary authorities have lowered real interest rates to negative levels, can they continue to have an impact on economic activity by flooding markets with liquidity? Is it feasible, both technically and politically, for the monetary authorities to prevent asset bubbles during periods of low and stable inflation? Moreover, the monetary authorities must operate with incomplete knowledge about the current state of the economy and how their actions (or inaction) may affect economic activity. Monetary policy operates with long and variable lags and it is difficult to anticipate market responses to shocks. Yet, the monetary authorities must immediately determine whether the financial system has adequate liquidity and whether monetary policy needs to be adjusted to counter the effects on the economy of a crisis-induced tightening of credit [Kohn (2010)].

---

2 The U.S. had not implemented Basel II for banks in 2008, but the five large investment banks, which were heavily involved in the crisis, had adopted a form of the Advanced Internal Ratings System as “Consolidated Regulated Entities” overseen by the Securities Exchange Commission (SEC). Outside the U.S., many of the large banks had adopted Basel II by 2008.

3 Jacob Frenkel, former Governor of the Bank of Israel, has expressed doubt about whether the monetary authorities know enough to deflate bubbles before they become dangerous (Michaels and Thornhill 2009). He asserts that the real choice is “Which system do you want: one in which the [monetary authority] pricks three bubbles out of five or five out of three bubbles? Because we know for sure that it will not be able to solve four out of four.”
Prudential supervision focuses on safety and soundness issues. Micro-prudential supervision focuses on the safety and soundness of individual institutions. In contrast, macro-prudential supervision focuses on the safety and soundness of the financial system. Most supervisory resources have traditionally been allocated to micro-prudential supervision, but that does not ensure macro-stability goals will be achieved. Indeed, sometimes an action taken to make an individual institution safer can undermine macro-financial stability. Demanding that individual institutions hold more capital and reserves can appear to make those institutions safer, but banks in turn may restrict credit and refuse to roll over loans to other institutions, thus reducing aggregate demand at a time when the economy is already weakened.

With regard to prudential policy, the primary goal of financial stability must be to protect the functioning of the financial system to ensure the provision of payments services and facilitate the efficient allocation of resources over time and across space. This may be threatened by a loss of confidence in key financial markets or institutions, as we have seen in the recent financial crisis and the still more recent sovereign debt crises in the Eurozone. But how safe should financial institutions be? Should all failures be prevented? Would the restrictions on risk-taking necessary to achieve that goal reduce the efficiency of financial intermediation and reduce investment? Would this deprive the economy of the dynamic benefits of creative destruction? Or would it shift risk-taking and innovation to unregulated entities in the shadow banking system where it would be less well monitored and more difficult to control? What tools should be used to achieve these objectives? And what governance structure is most likely to motivate policymakers to act in the public interest?

Public-sector compensation contracts are much more highly constrained than compensation contracts for senior executives in financial services firms. Yet even private institutions, which have had complete freedom of contracting, have failed to devise incentives to manage risks effectively. During the Great Recession, we learned that several firms had adopted incentive systems that led executives to take on much more risk than their shareholder principals or taxpayers, who were called upon to provide a bailout, would have wanted them to do. More fundamentally, when objectives are not crisply defined, it is difficult to establish and enforce accountability. Blame avoidance is, by default, the primary objective of most bureaucrats [Kane (1980)].

Although the prudential supervisory authorities have enormous leverage over fragile institutions that require liquidity assistance, they have relatively little influence over well-capitalized, apparently profitable institutions that they may nonetheless believe to be taking on excessive exposures to uncertain shocks.

In order to guard against the arbitrary use of regulatory and supervisory power, most countries subject disciplinary decisions by officials to some sort of judicial or administrative review. The result is that in order to discipline a bank, a supervisor must not only know that a bank is taking excessive risk, but it must also be able to prove it to the satisfaction of the reviewing body – perhaps beyond a reasonable doubt. This leads to a natural tendency to delay disciplinary measures until much of the damage from excessive risk-taking has already been done. The U.S. has tried to limit the scope for forbearance through Prompt Corrective Action measures contained in the Federal Deposit Insurance Corporation Improvement Act of 1991.4 This attempted to remove a substantial degree of supervisory discretion by triggering mandatory regulatory interventions when an institution’s capital ratio falls below prescribed levels. The implementation of Prompt Corrective Action measures, however, has often been less than prompt. For example, Indy Mac lost nearly U.S.$9 billion [see Adler (2009)] before it was closed. A basic problem is that the triggers for Prompt Corrective Action are defined in terms of accounting ratios and experience has shown that accounting values systematically lag real economic values when markets turn down [Herring (2012)]. Moreover, during times of crisis even explicit rules may be insufficient to limit forbearance.

A more fundamental problem is that both bankers and supervisors have strong motives to withhold unfavorable information as long as possible. For example, frontline risk-takers will sometimes try to conceal losses from their supervisors in the hope that they can recoup their losses before they are discovered. Higher-level executives, in turn, may try to conceal losses from security analysts and regulators. Executives may believe that they can protect their compensation (or jobs) by withholding adverse information. Equally importantly, they may wish to protect their scope for dealing with the problem before it must be disclosed.

4 See http://thomas.loc.gov/cgi-bin/query/C?c102:./temp/~c102x064ch for details of the FDICIA legislation.
Supervisors may have similar incentives for withholding unfavorable information. They may hope to remedy the problem before it needs to be publicly disclosed and, especially, they may want to protect their scope for making decisions if adverse information was shared with other regulators who might take preemptive action. In addition, if the negative information leaked to the public and precipitated a run, the regulators might be forced to take actions they would prefer to avoid.

Review of supervisory actions also leads officials to react mainly to what has already happened (and is, therefore, objectively verifiable) rather than to act on the basis of expectations about what may happen (which are inherently disputable). Moreover, supervisors are seldom, if ever, held accountable for the misallocation of resources that occurs when an insolvent institution is permitted to continue operations because there is no official scrutiny of opportunity costs. Unfortunately, supervisors are criticized mainly for taking action even when it is justified, rather than for not taking action, even though earlier action might have prevented financial disorder.

Charles Goodhart (2010) has suggested that K should be partitioned into actual past data and expected values. Supervisors generally react to past actual losses rather than mean expected losses, much less unexpected losses, even though they believe they know the probability distribution that governs outcomes. Greenspan (2008) has expressed doubt about whether regulators know enough to act preemptively.

Information issues present a fundamental challenge to supervisory authorities who must evaluate the solvency of regulated financial institutions. Neither actual past data nor expected values can be relied upon in times of crisis when markets become illiquid and difficult supervisory decisions must be made. Bank accounting has traditionally been a mix of historical cost accounting, accrual accounting and mark-to-market accounting that can obscure the true financial condition of a bank. This has sometimes undermined incentives for hedging risks by requiring that a risky position and the offsetting hedge be valued differently. Thus, the volatility of reported earnings may increase, even though risk has been reduced.

Many external observers question whether this mix of valuation approaches conveys a true and fair account of the current position of a financial institution. New financial accounting standards introduced more or less simultaneously by the International Accounting Standards Board and the Financial Accounting Standards Board require that assets held at fair value be classified in three different categories: (1) assets that can be marked to market based on prices in active markets for identical instruments; (2) assets that are marked to matrix, based on observable market data for similar assets and (3) assets that are marked to model, based on judgment regarding how the market would price such assets if they were traded in active markets. (This is sometimes referred to as marking to model, or more cynically, marking to myth.)

This third category presents significant difficulties for regulators, who face a severe asymmetric information problem vis-à-vis the regulated institution. How can the regulatory authorities comfortably rely on the estimated values of category three assets? Yet the question of the correct price for such assets is critical for implementing policies, such as the original version of the Troubled Asset Relief Program (TARP) in the U.S., asset guarantee programs deployed in the U.S. and other countries, and numerous bad bank resolutions implemented in many countries. If the price is set too low, banks will not participate willingly. Yet if the price is set too high, it can be the source of enormous concealed taxpayer subsidies. Elizabeth Warren, then the independent monitor appointed by Congress to scrutinize the bank recapitalization program, concluded that the Treasury overpaid U.S.$78b of the U.S.$254b spent in TARP transactions [Guha (2009)].

Part of the problem is that financial theory offers only two kinds of tools for valuing assets that are not traded in active markets: (1) the present value of discounted cash flows, which works well in a world of K, where cash flows can be predicted and risks estimated; and (2) real option theory, which works well only

---

5 An encouraging post-crisis trend in prudential supervision has been a greater emphasis on supervised simulations to evaluate capital adequacy under a variety of scenarios.

6 For additional discussion of fair value standards see Herring (2012). For example, S&P studied one bond, formerly rated AAA, backed by 9,000 second mortgages. Nearly one-quarter were delinquent and losses on those that had defaulted were 40%. One financial institution carried the bond on its books at 97 cents on the dollar. S&P ran a favorable scenario and concluded the bond was worth 87 cents on the dollar, but under a less favorable scenario it was worth only 53 cents on the dollar. At the same time, the bond traded in the secondary market at 38 cents on the dollar [Bajaj and Labaton (2009)].
if you can write a decision tree that captures most of the key uncertainties and decision points in the future and assign them plausible probabilities. When shocks occur, these fundamental prices move from the realm of K to u. Thus, during a crisis, fundamental values rest on relatively infirm foundations.

Even category 1 assets may become a problem in a crisis. Setting aside the issue of asset price bubbles, market values can be relied upon so long as assets are traded in broad, deep resilient markets. In such markets, however, assets tend to be priced on the basis of comparisons to their own past prices or to the prices of comparable assets. When a shock undermines confidence in these relative values and causes losses, traders tend to withdraw from markets until they regain confidence in their valuation models. Concerns may arise about counterparties who may have had excessive exposures to the shock and markets become thin. A flight to quality may occur and liquidity will be restored only when confidence in valuation models and counterparties is restored.

This dynamic has been very much in evidence during the recent crisis and led to the destruction of much of the endogenous liquidity in the system. Two channels of contagion have been identified [Brunnermeier et al. (2009)]. First is the tendency of counterparties to demand more collateral or larger “haircuts.” This reduces the leverage that their counterparties can obtain by borrowing against their assets and they must either raise more capital (which is expensive and difficult to do during a crisis) or deleverage by selling assets. If several institutions are selling assets at the same time, prices will drop and the market will become less liquid. This is likely to lead counterparties to demand still more collateral and still larger haircuts until the market simply evaporates. The second channel is driven by the losses that occur in institutions that hold long positions in the assets that decline in value. They too will be forced to try to raise additional capital or to attempt to deleverage by selling once liquid assets into an increasingly illiquid market.

Efforts by central banks to add liquidity to the system to revive these markets have had little effect so long as concerns persist regarding the solvency of key counterparties. Indeed, during the crisis and the agonizingly slow recovery, many U.S. banks have chosen to hold excess reserves at the central bank rather than participate in interbank markets. On occasion, the flight to quality has been so extreme that Treasury bill yields have turned negative.

**Crisis prevention**

Most policymakers would agree that it is better to prevent such crises than to try to manage and mitigate them once they have occurred. Crisis prevention is an enormous burden, however, which falls mainly on the shoulders of the prudential authorities. Prudential regulation attempts to establish rules for the sound operation of financial institutions and critical elements of the financial infrastructure, such as clearing and settlement arrangements. Ideally, prudential policymakers should be looking beyond the known to anticipate emerging sources of systemic vulnerability in order to calibrate appropriate prudential policies. In the dynamic world of modern finance, this requires trying to understand how changing institutions, products, markets, and trading strategies create vulnerabilities to new kinds of shocks and new channels of contagion, such as the margin/ haircut spiral and the loss spiral just described. But K cannot be neglected. Institutions still fail in familiar ways by taking excessive concentrations of credit risk, such as Wachovia, or by imprudently borrowing short and lending long as did Northern Rock and dozens of structured investment vehicles.

Prudential supervisory authorities confront a number of trade-offs that must be made in uncertain terms. How safe should banks be? Scott (2010) has argued that a central feature of corporate governance is aligning the risk neutral preferences of well-diversified shareholders with risk-averse managers. This calculus is unlikely to take account of the systemic costs of an institution’s failure and so the prudential authorities will presumably prefer a higher degree of safety, but how much higher? Goodhart (2010) has observed that it is relatively easy to establish a set of penalties that would make the banking system perfectly safe, but largely irrelevant in intermediating between savers and investors.

This is closely related to the degree of competition in the financial system. Competition is generally viewed as a positive feature. It stimulates innovation and lowers the cost of financial services. But, it also reduces the charter values of incumbent institutions and may lead to increased risk taking [Herring and Vankudre (1987)]. Should financial innovation be encouraged? Securitization has facilitated diversification of risk, reduced costs and liberated borrowers from dependence on particular lenders, but the subprime crisis showed that the fragmentation of responsibility in the securitization process...
can also undermine credit standards and enable banks to achieve higher leverage by evading capital requirements. Derivatives had enabled financial institutions to partition and manage risks much more efficiently, but they can also be used to assume enormous, highly leveraged risks. The growing sophistication of risk management techniques has enabled institutions to push out the boundaries of the known, but the very complexity of these techniques presents a challenge in the event of a crisis because it is very difficult for the authorities and potential investors and counterparties to comprehend the full range of risks and how they should be managed.

The supervisory authorities have a number of tools. These include licensing requirements, restrictions on certain kinds of activity deemed to be excessively risky, liquidity requirements, capital requirements and disclosure requirements. The authorities may also try to identify and encourage the widespread adoption of best practices in risk management, in effect urging the private sector to convert u into K.

Before the crisis, by far the most ambitious effort at prudential regulation was the development and implementation of the Basel II standards for capital adequacy and their successor Basel III, which try to emulate “economic capital”,8 that was used by many sophisticated institutions to measure and aggregate risks. But the concept is firmly rooted in the world of K. It depends on the bank being able to assess the expected loss and the equations provided by the Basel Committee being prudent in converting this into unexpected loss at a high level of confidence. This then provides at least one basis for the firm’s assessment of how much capital it needs to retain a given credit rating.

The original Accord on capital adequacy made a very crude attempt to reflect credit risk in the banking book by applying risk weights to on-balance-sheet assets and off-balance-sheet positions set capital requirements that formed the denominator of the required ratios. The concept of economic capital made clear that the role of capital should be to absorb unexpected losses, while reserves should be established to absorb expected losses. Banking supervisors from the 13 countries that then comprised the Basel Committee on Banking Supervision began to take note of the evolving concept of economic capital, and so when they expanded the original Basel Accord to include market risk, they permitted banks to employ their own models of market risk (usually some variation of Value at Risk (VaR)) under certain conditions designed to assure the regulators that the models were reasonably accurate and maintained with integrity. The 1996 Market Risk Amendment [Basel Committee on Banking Supervision (1996)] provided an entirely new approach to setting capital requirements that relied on the way that leading banks were measuring and managing this risk, although banks were also provided with the alternative of applying mechanical asset price haircuts to compute their regulatory capital requirement for market risk.

The market risk internal models approach was expected to deliver several benefits. First, it would reduce or eliminate incentives for regulatory capital arbitrage because the capital charge would reflect the bank’s own estimate of risk. Second, it would reward diversification to the extent that a bank’s internal models captured correlations across risk positions. Third, it would deal more flexibly with financial innovations, incorporating them in the regulatory framework as soon as they were incorporated in the bank’s own risk management models. Fourth, it would provide banks with an incentive to improve their risk management processes and procedures in order to qualify for the internal models approach. And fifth, compliance costs would be reduced to the extent that the business was regulated in the same way that it was managed. By and large, the internal models approach for market risk proved to be highly successful, even when it was severely tested by the extreme market disruption in 1997, 1998 and 2001 [Herring (2005)]. But the VaR internal models failed spectacularly during the recent crisis when several banks experienced 25 standard deviation movements in prices several days in a row.9

Nonetheless, this early success, in combination with the progress made in modeling credit risk, led to calls from industry to revise the original Basel Accord to incorporate an internal models approach to capital regulation of credit risk. Basel II attempts to extend this new approach to setting capital requirements to credit risk and operational risk [Basel Committee on Banking Supervision (2006)].

8 Economic capital is the amount of capital that a firm needs to ensure that it remains solvent over a specified period of time with a specified degree of confidence.

9 Haldane (2009) observed that a 25s event would be expected to occur once every 6x10124 lives of the universe.
The Basel II structure rests on three pillars: Pillar 1 specifies capital charges for exposure to credit, market and operational risks; Pillar 2, the supervisory review process; and Pillar 3, market discipline. Although the supervisory authorities were convinced that credit scoring models had significantly expanded the amount of credit risk that could be regarded as falling in the domain of the known, they were skeptical that internal models of credit risk were as reliable and verifiable as models of market risk. While some kinds of credit risk, such as retail lending, have rich and granular datasets comparable to market risk, other kinds of credit risk are less amenable to empirical analysis because data is sparse relative to past credit cycles and distinctly non-granular.

In the end, the regulators rejected the unconstrained use of internal models, but permitted qualifying banks to use their estimates of inputs in their own internal models. For the Advanced Internal Models approach, qualifying banks may use their own estimates of probabilities of default, losses given default, exposures at default and durations of exposure — as inputs in the regulatory model determining capital requirements. Thus, Pillar 1 capital requirements recognized (to a limited extent) the analytical and empirical advances banks had made in expanding the domain of K to include credit risk. Unfortunately, the losses experienced by the institutions most heavily engaged in packaging and selling subprime-related debt have cast considerable doubt on the reliability of such models.10 And the Standardized version of Basel II, which relies on ratings issued by Nationally Recognized Statistical Ratings Organizations did not fare much better. In many cases, more than half the tranches of subprime-related securities were subjected to triple notch downgrades, which are extraordinarily rare in corporate issues.

Because Basel II is an agreement negotiated among the members of the Basel Committee on Banking Supervision, it reflects a number of political compromises that undermine its aspirations for technical precision. This is most evident in the definition of regulatory capital, which is based on accounting values and includes a number of items that do not reflect an institution’s capacity to bear unexpected loss. This undercuts the link to economic capital and the logic of the approach.

Pillar 1 capital charges are intended to deal with known risks. Pillar 2, the supervisory review process, is intended to deal with unknown risks that can be identified, but are not sufficiently well quantified to establish Pillar 1 capital charges. Presumably, as theoretical and empirical advances succeed in moving some of these risks into the domain of K, Pillar 1 capital charges will be established for them as well.

The principal tools of supervisory analysis in the domain of the unknown are stress testing and scenario analysis. Stress testing requires economic judgment to formulate and calibrate scenarios that expose potential vulnerabilities. It requires a careful consideration of which relationships will continue to hold and which relationships will break down in time of stress. Mandelbrot and Taleb (2010) have cautioned that traditional stress testing, which relies on selecting a number of worst-case scenarios from past data, may be seriously misleading because it implicitly assumes that fluctuations of this historical magnitude would be the worst that should be expected. They note that crashes happen without antecedents. Before the crash of 1987, for example, stress testing would not have included a 22% drop in share prices within a single day or that 63% of the returns on the stock market over the past 50 years occurred in just 10 trading days. If shocks are subject to wild randomness, risk managers and prudential supervisors must evaluate the robustness of a portfolio over a broad spectrum of extreme risks.

Goodhart (2010) has emphasized a different concern regarding stress testing and scenario analysis. What may matter most in a crisis are interactive effects that occur when many institutions attempt to adjust their portfolios in the same way at the same time. These are critical to understanding an institution’s vulnerability in a crisis and the impact on the financial system, but are omitted from most scenarios.

Stress testing and the simulation of crises may be of value even if such crises never occur. The data necessary to simulate a crisis may prove useful in monitoring vulnerability and a careful consideration of the consequences of such a crisis may lead to changes in strategy and/or risk management. Crises seldom unfold according to the anticipated scenario, but strategies for responding to one kind of shock may prove useful when a different kind of shock occurs. For example, evacuation procedures that Morgan Stanley established after the bombing

10 U.S. banks remained on the Basel I standard, but the five major investment banks, which were heavily involved in structuring subprime debt, had adopted a variant of the Basel II advanced internal models approach as “Consolidated Supervised Entities.” Most securitized assets were held in trading books rather than banking books.
of the World Trade Center in 1993 enabled the firm to safeguard all of their employees in the much more severe terrorist attack on 11 September 2001. Learning from close calls or near misses may be as important as learning from actual losses.

The key element of regulatory discipline under Pillar 2, however, is the ability of the prudential supervisor to impose an additional capital charge on an institution if they are uncomfortable with the results of its stress tests. This places supervisors in the role of imposing discipline on an institution thought to be vulnerable to a shock of unknown probability. This will inevitably prove challenging for supervisors who are usually much less well paid and less well informed than bank managers. The history of bank supervision does not provide much basis for optimism that they will succeed. Indeed, Northern Rock provides a recent example to the contrary. In June 2007, just before the near collapse of the bank, the British Financial Services Authority [FSA (2008)] authorized Northern Rock to apply the Pillar 1 internal-ratings-based risk weights, which reduced its required regulatory capital by nearly 30%. Northern Rock, in turn, planned to increase its dividends to shareholders by 30% [FSA (2008)]. The FSA made no attempt to offset the reduction in Pillar 1 capital charges with an increase in capital charges under Pillar 2, nor did it require Northern Rock to conduct a stress scenario that would have shown that it was fatally exposed to a liquidity shock because of its heavy reliance on wholesale markets to fund large portfolios of mortgages while they were being seasoned for securitization. The collapse of the market for securitized debt left Northern Rock with an inventory of mortgages that it could not sustainably finance. But in some cases, the authorities have found it impossible to even imagine the shock that may occur. 11

How should prudential supervisors deal with U? As Scott (2010) has noted, firms can limit their leverage and maintain enough capital and liquidity to absorb unknowable losses if they should occur. But how much financial slack is sufficient? By assumption that is unknowable, but almost all of the things that banks could do to cope with the unknowable are very costly, and competitive pressures may make it very difficult to sustain such precautions. Should regulators, therefore, require that banks hold capital substantially in excess of the regulatory minimum as a safeguard against unknown and unknowable shocks? Andrew Crockett (Herring (2009)) has observed that policymakers find it inherently difficult to strike the proper balance between the efficiency losses associated with excessively onerous preventative policies and the cost-effectiveness of responding ex-post to adverse events. For regulators as well as firms, the appropriate amount of financial slack is an unknown.

Pillar 3 of the Basel II approach was intended to enhance market discipline by improving disclosure. The authorities may collect and publish data that helps market participants understand the current state of the economy and financial markets, and the condition of regulated financial institutions. But growing reliance on dynamic trading strategies to manage risk has made it increasingly difficult to provide a meaningful picture of risk exposures. Positions may change so rapidly that information is out-of-date before it can be published. Moreover, the chief motive for market discipline – the fear of loss – is often undermined by the reluctance of the authorities to permit the creditors and counterparties to suffer loss. Sometimes, even shareholders of systemically important financial institutions have been shielded from the full cost of an institution’s losses. The extraordinary bailouts of a number of major global financial institutions are recent cases in point. The exception, the bankruptcy of Lehman Brothers, proved so disruptive that the heads of state who comprise the Group of 20 vowed that it should never happen again.

The ambitious Basel II approach attempted to incorporate in capital regulation what is known about risk management, but it generated several unintended consequences that made the system more vulnerable to a sudden shift into the domain of the unknown. The attempt to force all major firms to adopt one version of “best practice” and especially the imposition of a regulatory model of credit risk increased the likelihood of herding so that banks were more likely to attempt to move in the same direction at the same time, which undermines the liquidity of markets and increases price volatility.

To the extent that Basel II succeeded in making capital requirements more risk sensitive, it also made bank lending more procyclical. This tendency was reinforced by the fact that banks increasingly hold marketable assets that must be fair valued. In a boom, measures of risk are likely to decline and prices of assets are likely to rise and so, in combination with higher retained

11 See FSA (2008) for the FSA’s assessment of its failings with regard to the supervision of Northern Rock.
earnings, measured capital will rise just as required capital falls, thus facilitating additional lending and accentuating the boom. In a recession, internal ratings will migrate downward, thus increasing required capital. In addition, asset prices will fall, increasing the erosion of the bank’s capital position caused by credit losses. This will exacerbate the constriction of the supply of loans. This interaction of more risk-sensitive capital requirements with a higher proportion of fair-valued assets tends to accentuate booms and busts. This is a clear instance in which micro-prudential and macro-prudential objectives conflict. Basel II aimed to make individual banks safer, but may have inadvertently weakened the banking system. More fundamentally, Basel II failed to deal with systemic risk.

Crisis management

Because it is so difficult for prudential supervisors to fulfill their responsibilities ex-ante, policymakers must often shift into crisis management mode to mitigate, ex-post, the consequences of a shock. In a financial crisis, the ratio of $u$ and $U$ will be especially large relative to $K$ (Kohn (2010)). Policymakers must deal with unknowns, such as the size of the disruption. How large will it be? How many firms will be involved? How long will it last? How likely is it to have serious spillover consequences for real economic activity?

Part of the problem is the challenge of anticipating the channels of contagion. Which firms have direct exposure to the shock? Which firms have indirect exposure because they are counterparties or creditors of the firms that sustain the direct impact or because they have similar exposures and could lose access to external financing? Which other firms might be placed in jeopardy because of the forced liquidation of assets in illiquid markets as risk-averse lenders and counterparties demand larger haircuts and/or more collateral? Risk preferences and perceptions of risk are dynamic and so a flight to quality often occurs. Market participants may sell assets whose prices are already declining to close-out positions. Moreover, they will avoid any counterparty that might be impaired. During the recent crisis and its aftermath, policymakers have learned that their ability to substitute central bank liquidity for endogenous liquidity created by the financial system is less effective than was once believed.

In a crisis, policymakers must try to convert $u$ into $K$ as quickly as possible. This requires close cooperation across regulatory authorities within a country and, increasingly, across borders. Inevitably, the authorities must rely on major market participants for much of their information about current conditions. But conflicts of interest may corrupt flows of information. Information may be selectively communicated to serve the self-interests of market participants who might be the beneficiaries of crisis management policies.

Does this argue for a direct role of the crisis manager in supervising systemically important institutions? The Fed insists that it does, but central banks lack such authority in many other countries (Herring and Carmassi (2008)). After the crisis, the prudential supervision function has been moved back into the Bank of England and the Financial Services Authority has been abolished. More recently, the E.U. has agreed to establish a separate banking supervision function inside the European Central Bank. Nonetheless, how best to organize prudential supervision and crisis management remains a significant unknown.

In addition to gathering information during a crisis, policymakers must also convey information. They may urge firms to do what the policymakers believe they should do in their own self-interest, as happened in the LTCM crisis in 1998 and was attempted again in the two days preceding the Lehman bankruptcy. But when is it appropriate to be reassuring? And when might reassurance prove counterproductive by leading market participants to expect outcomes that cannot be assured?

Ironically, crisis management that is successful in the short run may inadvertently lead to larger future crises. If risk-takers are protected from the full negative consequences of their decisions, they may be likely to take greater risks in the future. This presents a difficult dilemma for crisis management. The costs of inaction are immediate and obvious. It is easy to imagine damaging outcomes and self-interested market participants will press for official support and can easily muster political support. Inaction in a crisis is likely to be subject to blame even when it is not appropriate. This may contribute to an inherent tendency to oversupply public support. Once it has been provided, entrenched interests will lobby to keep it and new additional activity may depend on it. Moreover, moral hazard manifests itself slowly and may be difficult to relate to any one particular policy choice. The history of crises teaches us that it is very difficult for the
Financial policymaking in the context of the known, the unknown and the unknowable

authorities to exit from guarantees issued in a crisis once normal conditions are restored, although the FDIC’s successful removal of extraordinary guarantees for large transactions accounts provides an encouraging example to the contrary.

The evolving policy response to the crisis: attempting to extend the domain of K

Even before the crisis was fully resolved, policymakers attempted to patch some of the more obvious gaps in the supervisory network and to apply lessons learned during the crisis. Basel II had proved to be a comprehensive failure. The denominator did not fully reflect the banks’ exposures to risk, the numerator did not accurately portray an institution’s ability to absorb loss and remain a going concern, and the required minimum ratio was much too low. Indeed, banks that failed tended to have higher Basel II ratios than banks that encountered less serious troubles. For example, Citibank reported a Tier-1 capital ratio of 11.8% when its stock market capitalization fell to about 1% of its accounting assets [Citicorp (2008)]. Surprisingly, despite the emphasis on making capital requirements more risk sensitive, reported regulatory capital ratios scarcely moved, even though banks were experiencing the worst crisis since the Great Depression.

An independent observer might have been tempted to jettison the Basel II approach in favor of a much simpler system. Indeed, simple leverage ratios tended to be much better predictors of banking strength during the crisis than the much more complex Basel II ratios. But the Basel Committee, instead, attempted to salvage and extend the Basel II approach. They began by trying to adjust some of the risk weights that were more out of kilter. The Basel Committee tried to discourage resecuritizations by raising the capital charge on BB-rated tranches of resecuritizations from 350% to 650% and raising the capital charge on AAA-rated tranches from 20% to 40%. The market response to this change in regulations illustrates the difficulty that supervisors have in predicting the impact of what would seem to be a very straightforward increase in the stringency of regulations. Within months, financial engineers had found a way to work around the increased risk weights. They developed a new financial structure called a Re-Remic that enabled banks to reduce their regulatory capital charges on existing downgraded securities by swapping portfolios of newly tranched securities [IMF (2009)].

Ultimately, the Basel Committee produced a sweeping revision of Basel II known as Basel III, which continued the pattern of increasing complexity in the framework for the regulation and supervision of capital standards for internationally active banks. The Basel I framework was set out in a mere 28 pages. Basel II required 347 pages and Basel III, 616 pages. This rough measure grossly understates the increasing complexity of the international regulatory framework. This has imposed burdens on both banks and regulators. Basel I required six to eight calculations for a bank to determine its regulatory capital requirement. Haldane (2011) has estimated that a large complex bank using the Advanced Internal Ratings Based models for Basel II would have over 200,000 risk buckets and would require more than 200 million calculations to determine its regulatory capital ratios. The costs of implementing Basel III will be even larger and are estimated to require that banks hire an additional 70,000 full-time employees to comply.

The challenges in dealing with the crisis have caused fundamental change in banking regulation in the U.S., the U.K., and the E.U., but focus will be on the changes in international institutions and regulations. World leaders were keenly aware of the massive resources required to support the international banking system. Haldane (2009) has estimated that the authorities in the Eurozone, the U.S., and the U.K., committed over U.S.$14t (approximately one-quarter of world GDP) to support their banking systems. Moreover, the loss of GDP rivaled that lost in major wars. Since this crisis had worldwide ramifications, the infrastructure for making economic policy decisions was expanded. Instead of the G-7, G-8 or G-10, the new policy commitments were made by the G-20 that included representation from the emerging powerful economies of Africa, Asia and Latin America. In addition, the G-20 transformed the Financial Stability Forum into the Financial Stability Board and gave it the mandate to make sure that the policy decisions by the G-20 were implemented.

The G-20 recognized several major gaps in pre-crisis oversight

---

12 Citicorp continued to report Basel I capital ratios because the U.S. had not yet implemented Basel II.

13 Membership in the Basel Committee has been similarly augmented to better reflect the dispersion of financial power since the Committee was originally formed.
of the financial system. First, policymakers realized that Basel II had failed to promote safety and soundness. Second, they wanted to mitigate the procyclical dynamics of Basel II. Third, they concluded the incentives built into Basel II to encourage the largest banks to adopt more sophisticated risk management techniques in exchange for lower required capital ratios was perverse. Fourth, they wanted to limit leverage more effectively. Fifth, they believed that the regulators had failed to oversee liquidity risk adequately. Sixth, they wanted to make the derivatives more transparent by shifting as much activity as possible from over-the-counter markets to exchanges or central clearing parties where derivatives activity could be more easily monitored. Finally, they recognized that some institutions had become not only too big (or too complex or too interrelated) to fail, but also that some institutions had become too big (relative to home country resources) to save.

Revisions in Basel III were intended to address the first six issues. In addition to tinkering with the risk weights and making some of the assumptions in the regulatory model more conservative, they increased capital charges on credit risk in the trading book, took account of wrong-way counterparty risk and penalized holdings of derivatives that continued to be traded over-the-counter relative to derivatives that were traded on exchange or cleared through a central clearing party.

The most important change, however, involved the numerator in the capital adequacy ratio. Indeed, the main thrust of Basel III was more and higher quality capital. Basel I had begun with a fairly clear-cut distinction between Tier-1 capital (shareholder’s equity and instruments, such as non-cumulative perpetual preferred that were virtually indistinguishable from equity in times of stress) and Tier-2 capital, which contained a wide variety of instruments and allowances that included a concession to virtually every major country that had been reluctant to sign the Basel Capital Accord. Over time, Tier-1 capital was diluted by the introduction of a number of hybrid instruments that were not only enough like debt to satisfy the tax authorities that banks should be permitted to deduct interest payments on these instruments, but also enough like equity to satisfy the regulators that they would be the functional equivalent of equity in a downturn. Hybrid instruments eventually grew to become half of the Tier-1 requirement.

The crisis, however, clarified thinking about what should count as capital. The Basel Committee realized that many of the hybrid instruments were not useful for sustaining confidence in a crisis. They made an analytical distinction between going-concern capital – capital that would enable the firm to continue operating even though it was suffering losses – and gone-concern capital – capital that would buffer other creditors from loss only after the bank had gone through some sort of bankruptcy or resolution process.

On the basis of this distinction, the Basel Committee redefined Tier-1 capital. Unfortunately, they chose to complicate this straightforward distinction by distinguishing two types of Tier-1 capital – “common equity” Tier-1, which is what most analysts focused on in the crisis, and “additional” Tier-1 that permitted the inclusion of some of the most rigorous forms of hybrid capital. The following discussion focuses on the Tier-1 common equity requirement because it is likely to be the binding constraint.

Basel III requires a new higher minimum of 4.5% Tier-1 common equity ratio. In order to reduce the extent to which capital requirements exacerbate business cycles, the Basel Committee has introduced an additional “capital conservation buffer” of 2.5% Tier-1 common equity. The Basel Committee seems to be a bit conflicted about how this should work. The principal purpose of this buffer is to encourage the bank to reduce its capital in hard times rather than to reduce its lending. But at the same time, the Basel Committee has provided disincentives for banks to use this flexibility by imposing constraints on a bank’s ability to make discretionary distributions if it does draw down the buffer.

The Basel Committee has also attempted to provide a tool for regulators to attempt to prevent bubbles. Under Pillar 2, the supervisor may negotiate an additional 0-2.5% of Tier-1 capital to dissuade banks from excessive lending that might fuel an asset price bubble. In order to make the regulation effective, however, the Basel Committee has ruled that the same add-on must apply to all foreign banks in the country. Thus, the countercyclical buffer must be announced at least 12 months in advance.

---

14 Additional equity will include only hybrid capital that is issued and paid in, subordinated to depositors, general creditors, and subordinated debt, neither secured nor covered by a guarantee of the issuer, a perpetual maturity, with no step-ups or other incentives to redeem and non-cumulative. While these constraints are reasonable, creating yet another ratio: a Tier-1 common equity ratio and a Tier-1 ratio. This has introduced yet another ratio in an already very complicated system.
Financial policymaking in the context of the known, the unknown and the unknowable

to provide time for foreign banks to comply. This seems to take a remarkably optimistic view of the ability of the regulatory authorities to forecast bubbles and to time their interventions with such precision that they prevent the bubble from forming rather than exacerbating financial conditions after the bubble has burst.

Reversing the favoritism extended to large banks under Basel II, Basel III will impose an add-on to capital requirements for “systemically important institutions.” This is in recognition of the fact that large banks posed a more serious risk during the recent crisis than smaller banks. The Financial Stability Board posts a list of such institutions each November. The add-on may vary from 1% to 3.5% of Tier-1 common equity. To date, the largest add-on has been 2.5%, but the 3.5% bucket is held out as a threat to discourage banks from becoming still bigger and more complex.

While many of these changes appear to be improvements in the international regulatory framework, several defects were not addressed. The denominator in the risk-weighted ratios under the standardized option relies on fixed risk weights that are often inaccurate. Moreover, these risk weights do not change over the cycle even though the riskiness of bank assets surely does. Perhaps even more troubling is the internal models-based approaches, which continue to rely on a bank’s own internal models that can be easily manipulated and are difficult to monitor. More fundamentally, the complexity of the risk-weight ratios makes it virtually impossible to compare capital adequacy across banks.

The most striking innovation in the Basel III capital requirements, however, is the introduction of a leverage ratio. This puts a floor on the extent to which banks can manipulate risk weights to lower their capital requirements. Moreover, it is easy to monitor for regulators and facilitates comparisons across banks by the market. It is, moreover, what market participants relied on during the crisis, when it became apparent that risk-weighted capital ratios were misleading at best. The new approach is being introduced under Pillar 2, but it is intended to migrate to Pillar 1 eventually.

Concerns about liquidity are addressed in two new liquidity ratios. The first is intended to ensure that banks have enough high-quality liquid assets to cover total net outflow over a 30-day period of stress. Unfortunately, this concept has been so weakened in the negotiation process that the numerator contains numerous assets that will be liquid only under the most optimistic of assumptions and the denominator assumes a scenario that is much less stressful than most banks experienced during the recent crisis. The long-term net stable funding ratio is so controversial and the feared consequences so serious, that it seems unlikely to be implemented.

Thus, Basel III has added to the complexity of Basel II while attempting to correct some of the more obvious shortcomings. Stefan Micossi (2013) has observed, “Basel III has made the system even more complicated, opaque and open to manipulation.” This complexity is itself a problem. In addition to the deadweight costs in complying with and monitoring compliance with literally thousands of pages of regulations, the new regulatory initiatives introduce new sources of uncertainty. Banks are uncertain of how the rules will ultimately be implemented and enforced, and the authorities are uncertain about how banks will react to the barrage of new regulations. If the past is any guide, however, the authorities would be wise to expect new and ingenious forms of regulatory arbitrage.

The attempt to deal with the too-big-to-fail issue represents an even more striking innovation in international regulatory and supervisory policy. First, the authorities have realized that their own actions — subsidizing mergers between large, weak institutions with large, strong institutions — have been a major contributor to the growth of such institutions, which have more than doubled their share of banking markets since the mid 1990s. In addition, numerous studies have shown that the perception that such institutions are too-big-to-fail gives them an advantage in the cost of funds that is wholly unrelated to the quality of their services or capital strength. Ultimately, efficient resolution policy may be the best safeguard against this kind of moral hazard, but the crisis revealed that policymakers in most countries lack the appropriate tools to resolve a large, complex financial institution without jeopardizing the rest of the financial system [Herring (2004)]. And large, complex financial institutions have adopted corporate structures that defy efficient resolution.

15 More precisely, they are considered too large to be permitted to cause loss to creditors.
The 28 large, complex financial institutions identified by the Financial Stability Board as systemically important have 2.5 times as many majority-owned subsidiaries as the 28 largest non-financial corporations (Herring and Carmassi (2009)). Moreover, most of these financial institutions are managed in an integrated fashion, along lines of business with minimal regard for national boundaries or the entities that must be taken through some sort of bankruptcy process in the event of failure.

The regulatory authorities have recognized that to counter the presumption of automatic bailouts created by the response to the crisis, they must devise ways of resolving institutions that will not create intolerable spillovers, but will punish managers, shareholders and, if necessary, some creditors of the faltering institution. The financial authorities could accomplish more by doing less, if they can credibly restore a role for market discipline in the system by devising a credible resolution plan for every systemically important institution.

The Financial Stability Board has been charged with leading this effort. All of the members of the now extended Basel Committee have agreed that each systemically important financial institution must file a recovery and resolution plan. Several countries have already required that their largest banks submit recovery plans — more popularly known as living wills or, more sardonically, as funeral plans. If these resolution plans indicate that the institution cannot be resolved in a reasonable amount of time without causing financial disruptions, the authorities reserve the right to require that an institution restructure or even sell off units to creditors. The aim has been to ensure that taxpayers will never again be vulnerable to such potentially massive costs to support their financial systems. This has also led to an agreement that creditors should be bailed-in before deposit insurance funds or taxpayers suffer a loss. This is a dramatic change from practice during the crisis. The first implementation of the bail-in doctrine in Cyprus was not encouraging. Cypriot banks did not have sufficient Tier-2 capital to absorb losses and so it was necessary to bail-in general creditors. Since uninsured depositors have the same standing as general creditors in much of Europe, this cause ripples across the Eurozone.

**Concluding comment**

Improving the effectiveness of prudential policy should be an important priority. But this requires an honest assessment of what we actually know about the risks assumed by systemically important institutions and, indeed, what the managers themselves know. Since many of these risks are unknown and some are unknowable, supervisors need to place much greater emphasis on increasing the resilience of the system. Rather than imposing increasingly prescriptive regulations, a wiser course may be to devise simpler rules that can be defined and monitored more easily. As Ralph Gomory (1995) noted, “[A]s the artifacts of science and engineering grow ever larger and more complex, they may themselves become unpredictable.” This may very well be the situation we now face. Rather than heaping still more amendments on the Basel III structure, the Basel Committee might be more effective if it focused on simplifying the system to make it more transparent, flexible, and easier to monitor.

**References**

Adler, J., 2009, “FDIC to sell IndyMac; but cost estimate grows,” American Banker, January 5


Basel Committee on Banking Supervision, 1996, “Amendment to the Capital Accord to incorporate market risks,” January


Gomory, R., 1995, “The known, the unknown and the unknowable,” Scientific American, June


Greenspan, A., 2008, “The Fed is blameless on the property bubble,” Financial Times, April 7, 9

Financial policymaking in the context of the known, the unknown and the unknowable


Kane, E., 1980, “Politics and Fed policymaking: the more things change, the more they remain the same,” Journal of Monetary Economics, 6, 199-211.


Part 2: Tactical

Japanese financial institutions expanding abroad: Opportunities and risks

Safe to fail

Market risk of real estate: Using indirect data to understand direct risks

Insurance risk transfer and categorization of reinsurance contracts

Superior information and compensation fees of active mutual funds
Japanese financial institutions expanding abroad: Opportunities and risks

W. Raphael Lam
International Monetary Fund

Abstract
Overseas activities of Japanese financial institutions have risen, mainly in Asia, since the global financial crisis. Stagnant growth and low interest margins in Japan have added to incentives to seek opportunities abroad. This paper explores the determinants of Japanese banks’ overseas expansion, and assesses whether these cross-border activities will continue under the new macroeconomic policies often referred as “Abenomics.” The analysis finds that Japanese banks are well positioned to scale up foreign exposures, thanks to their relative resilient balance sheets and the robust growth in the region. Stronger domestic growth in Japan could mitigate the pace, but is unlikely to reverse a long-standing trend because empirical estimates suggest that global and regional factors play a more prominent role in the growth of cross-border claims. The increasing cross-border activity would pose funding risks and supervisory challenges that require continued close monitoring. An incomplete set of domestic policies that fails to raise growth and exit deflation could, however, undermine prospects of Japanese banks in expanding abroad.

1 The views expressed herein are those of the author and should not be attributed to the IMF, its Executive Board, or its management.
Japanese financial institutions expanding abroad: Opportunities and risks

Introduction
Cross-border activities of Japanese financial institutions have risen over the past few years, particularly in the Asian region. Overseas loans by major banks are growing by more than 20% year-on-year. Major Japanese banks have attained an important global and regional presence, particularly in areas of syndicated lending and project finance. Foreign claims on Asia have recouped the decline at the height of the global financial crisis and are now at levels comparable to the 2005–08 period. Moreover, major brokerage firms and life insurers have sought acquisitions or strategic partnership overseas.

The current trend is often compared to previous episodes of overseas expansion by Japanese financial institutions over last decades. Those episodes can be broadly classified in three waves (Figure 1): (i) the rapid expansion in the 1980s – up until the burst of the asset bubbles in 1990; (ii) the expansion during the mid-1990s; and (iii) the expansion abroad beginning from 2006 but temporarily slowed during the global financial crisis. A question to explore would be how the current trend of overseas expansion has similarities to these episodes.

In addition, this paper assesses whether this trend will continue under the government’s new policy framework often referred to as Abenomics. The Bank of Japan’s (BoJ) new quantitative and qualitative monetary easing (QQME) framework – part of the three-pronged strategies to revive growth and exit deflation – intends to encourage financial institutions to shift away from government bonds and take on greater exposures of risky assets (such as loans and investment securities).2 An improved domestic outlook could increase financial institutions’ inward focus to satisfy rising credit demand. On the other hand, uncertainty over the Japanese Government bond (JGB) market and yen movements may stimulate diversifying needs outside Japan. Are there new risks that may emerge in light of increasing cross-border activity and the implications for financial institutions and supervisors?

To answer these questions, the paper analyzes what factors contributed to the Japanese financial institutions’ decisions to expand abroad recently. We undertake an empirical study on several banking systems in advanced countries and examine their foreign claims. We then apply those empirical results to the Japanese banking system and assess the role of each contributing factor. The paper builds on the literature on cross-border banking [Berger et al. (2000), De Haas and Lelyveld (2010), De Young et al. (2009)] and international capital flows [Fratzscher (2012), Jotikasthira et al. (2012)].

Our empirical results show that several regional and domestic factors have contributed to overseas expansions. Stagnant growth and limited domestic credit demand have added incentives for Japanese financial institutions to seek opportunities abroad. Modest global uncertainty, large growth differentials and the resilience of domestic banking systems are key drivers for cross-border claims. Outside Japan, robust growth in Asia and deleveraging of European banks in the region contributed to a rise in cross-border lending. The exchange rate appreciation in the past years might have added incentives for expanding abroad.

This paper argues that the trend of expanding overseas is likely to continue, but will depend on a supportive domestic economy under successful Abenomics and careful risk management and supervision. Robust growth in Asia and sufficient liquidity at home would imply that the trend of expanding abroad is likely to continue. Although stronger domestic growth might slow the expansion pace, it is not expected to reverse the trend unless an incomplete Abenomics poses risks to domestic financial stability. Increasing cross-border activity also adds to funding risks and supervisory challenges that require continued close monitoring.

Past experience of overseas expansion by Japanese financial institutions
The various attempts of expanding abroad by Japanese financial institutions over the last few decades can be broadly classified into three waves. These earlier waves of expansion have somewhat resulted in losses for the financial institutions without helping them secure major global footing, except for a few megabanks and securities firms. The first wave of overseas expansion by Japanese financial institutions occurred during the mid-to-late 1980s, about the same time as when many real-estate and construction companies markedly increased their outward foreign direct investments (FDIs) (Figure 1). In addition to financing those FDIs by real-estate and construction

---

2 The three-pronged Abenomics strategies include flexible fiscal policy, aggressive monetary easing, and structural reforms to exit deflation and raise growth.
Figure 1: Japanese financial institutions: global ranking and expansion overseas
Japanese financial institutions expanding abroad: Opportunities and risks

Japanese banks increased foreign lending to Asia again in the mid-1990s, in part to take advantage of the rapid growth in the region. Those loans were often denominated in foreign currency related to investment projects in the region. Following the Asian financial crisis, Japanese banks, however, incurred sizeable valuation losses and their nonperforming loans (NPL) rose sharply, forcing them to recede on overseas lending. External bank assets fell by about 40% in two years (Figure 1). Overseas losses, on top of domestic problems, had contributed to the subsequent banking crises in Japan that lasted until early 2000s. The banking crises have resulted in significant restructuring and consolidations that gave rise to a concentrated market, with a few megabanks in the form of financial groups, securities firms and life insurers.3

Overseas activities by Japanese financial institutions has risen again since 2005 despite the temporary decline during the peak of the global financial crisis. Japanese banks have also broadened their financing to non-Japanese entities and local demand. Recent developments and the factors contributing to the rise will be discussed in the next section.

Despite the rise and fall of overseas activities, net external assets for Japanese banks have been on an increasing trend over the last decades. This possibly suggests a change of funding source on overseas activity. In the years leading up to the Asian Financial crisis, Japanese banks have relied on foreign-currency financing that created a net liability position. Over time, Japanese banks have accumulated net external foreign assets while the short-term liabilities have remained stable, implying that banks have increasingly financed long-term overseas loans with domestic yen-denominated funds.

<table>
<thead>
<tr>
<th></th>
<th>Year-on-year growth</th>
<th>Cumulative growth from 2005-08</th>
<th>Cumulative growth since end-2008</th>
<th>Share of total consolidated claims</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All countries</strong></td>
<td>0.1</td>
<td>72.6</td>
<td>37.0</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Europe, of which:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peripheral European</td>
<td>10.8</td>
<td>83.3</td>
<td>2.5</td>
<td>25.4</td>
</tr>
<tr>
<td>countries</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core European countries</td>
<td>-1.2</td>
<td>73.8</td>
<td>-22.5</td>
<td>2.8</td>
</tr>
<tr>
<td><strong>Europe</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>15.1</td>
<td>88.6</td>
<td>7.0</td>
<td>17.0</td>
</tr>
<tr>
<td><strong>Asia Pacific, of which:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>8.1</td>
<td>67.0</td>
<td>42.1</td>
<td>43.6</td>
</tr>
<tr>
<td>Emerging</td>
<td>13.4</td>
<td>124.1</td>
<td>102.9</td>
<td>15.9</td>
</tr>
<tr>
<td>Offshore centers</td>
<td>17.2</td>
<td>126.9</td>
<td>105.6</td>
<td>9.9</td>
</tr>
<tr>
<td></td>
<td>7.6</td>
<td>119.5</td>
<td>98.7</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>19.2</td>
<td>62.4</td>
<td>12.4</td>
</tr>
</tbody>
</table>

Table 1: Cross-border consolidated foreign claims of Japanese banks
Source: BIS.
1. As of September 2012.
2. Peripheral European countries include Italy, Ireland, Greece, Portugal and Spain.
3. Core European countries include France, Germany, Switzerland and the U.K.

Recent developments in overseas activities
Japanese financial institutions have increased their overseas activities, mainly through takeovers and lending, subsequent to the global financial crisis, mostly in the Asian region. Stagnant domestic growth, relative resilience of Japanese banks through the global financial crisis and strong growth in Asia have contributed to the expansion abroad.

Banking sector
Japanese banks have increased their cross-border activities, mostly to the Asian region. Cross-border consolidated claims of Japanese banks abroad have increased since 2005 and reached almost U.S.$3t (about 15% of total banking and trust assets), according to the Bank of International Settlement (BIS). Claims on Asia have more than doubled since the global financial crisis (now accounting for about 16% of total foreign consolidated claims), and is now at levels comparable to the 2005-08 period. Exposures to Europe, however, have significantly slowed after the global financial crisis (Table 1 and Figure 1). A large share of the rising foreign claims is attributed to growing overseas loans by major banks. Overseas loans account for about 15%-20% of the total outstanding loan balance as of September 2012, with the lending activity to Asia being the strongest.

3 Their asset sizes are among the largest global financial institutions.
Figure 2: Project finance and syndicated loans in Asia and the role of Japanese banks
Japanese banks expanded their overseas network through various forms of ownership. Besides setting up local branches and subsidiaries, banks have sought the expansion of customer base and business functions through business alliances and investments in overseas financial institutions, and exploiting different forms of ownership structures tailored to local markets (e.g., financial holding company in the United States, bank subsidiary in China). The expansion abroad has placed Japanese banks among the key players in regional and global syndicated loans and project finance (Figure 2). Megabanks have stepped up project finance and syndicated loans business, particularly in Asia, because of their strong balance sheets and long-term approach in lending. Besides interest income on lending, banks also earn fees from arranging and underwriting deals. In Asia, syndicated loans are often raised in local currency and by multiple banks across a number of countries, thereby requiring a sound financial base for the lead banks. Project finance is largely related to financing infrastructure projects such as utilities, transportation and communications. The three megabanks in Japan have been increasing their participations in these markets, particularly following the exit of European banks (Tables A1 and A2 in the Appendix).

The performance of overseas lending among major Japanese banks has been stronger in several ways compared to their domestic lending. Overseas gross profits now account for about 30% of total gross profits (about half of which arise from net interest income). Net interest margins for overseas loans have improved after the global financial crisis and exceeded those for domestic loans. As megabanks have been cautious in lending abroad to firms with established credit history, credit risks on overseas loans are moderate. The average risk-monitored loans ratio for overseas lending was about 0.7% as of September 2012, much lower than that on domestic lending (about 2%). Syndicated loans underwritten by Japanese major banks to established foreign firms usually have high investment ratings and, therefore, relatively low credit risks. Those loans to firms in emerging markets are relatively small as a proportion of the overall overseas lending (less than the global average in proportion), and about one-fifth of the syndicated loans have covenants that limit credit risks. Although project finance could be more risky because of the longer duration, it is usually backed by underlying infrastructure assets. At the margin, overseas loans therefore appear more profitable in general but are associated with less risk.

Nonetheless, overseas expansion also brought new risks. Foreign currency and maturity mismatches are likely to rise going forward as the long-term funding base in Japanese banks has fallen short of total external loans. Japanese banks extend most overseas lending in U.S. dollar and in long maturity for which they do not have a natural funding base. These create dollar funding risks and add to the maturity mismatch in foreign lending. Major banks have increased their local deposit base (e.g., corporate deposits) – accounting for about half of the funding base – but it still falls short of the total external loans. Banks, therefore, rely on short-term financing sources, such as yen-dollar basis and currency swaps, that are subject to volatility, and by issuing foreign exchange-denominated bonds. The loan-to-deposit ratio for overseas loans continues to exceed 100% (compared to the loan-to-deposit ratio for domestic loans of about 70%), potentially contributing to funding risks. Funding cost and availability depend on credit ratings, which also affect prospects of securing certain lines of business. In the event of credit downgrades, funding costs could rise substantially and the loss of certain lines of business precipitate initial difficulties.

Nonbank sector: life insurers and securities firms
The trend of expanding abroad is not limited to Japanese banks only. Major life insurers have begun to strengthen their overseas business, especially in Asia, by acquiring or affiliating with local insurers for long-term profitability. They usually expand via incremental capital and building alliances, typically involving minority stakes rather than aggressive acquisitions. To gain competitiveness in local markets, they broaden the range of products and services (e.g., medical insurance) and increasingly rely on more efficient distribution channels (e.g., selling through banks “bancassurance”). To date, as the majority of overseas investments are minority interests, the risk and return from overseas business for major life insurers tends to be modest.

---

4 According to the Bank of Japan, default rates on selected overseas loans ranged from 0.4–1.3 percent, much lower than the respective loan margins (Financial System Report Chart III 3-11). Banks are relatively cautious in choosing overseas loan extension and setting loan conditions.

5 As life and nonlife insurers tend to lower their weight of their stockholdings, they are likely to reinvest these funds in overseas mergers and acquisitions.

6 “Bancassurance” accounts for about 35-70 percent of new business premium in Asia.
Despite the recent setback in global operations, leading Japanese securities firms sought to counter diminishing prospects by expanding overseas. Outward FDI on financial services (banks and insurers) surged in 2008 on the account of Nomura Holdings, a securities firm, acquiring the European arms of Lehman Brothers. The market share of the leading Japanese securities firms outside Japan is relatively limited in most areas such as financial advisory roles in mergers and acquisitions, capital market issuance and underwriting (Thomson Reuters (2012)). This is because of limited expertise in executing services outside Japan, and relatively higher funding costs on foreign-currency instruments than other leading global peers, possibly due to lower credit ratings.

Factors contributing to increasing cross-border activity
Several domestic and regional factors have contributed to the increasing trend of overseas activity among Japanese financial institutions. First, limited domestic opportunities have forced major Japanese banks to look for opportunities and expand overseas. Credit demand was sluggish in the past few years due to stagnant growth, though it has picked up recently. Large corporations have limited funding needs as they accumulated sizeable surpluses (rising to about 6% of GDP). Structural factors – such as high leverage among SMEs, aging population and sluggish growth in Japan’s regions – have limited domestic opportunities. At the same time, low interest rates have persisted for a decade but lingering deflation has limited the decline of real interest rate to sufficiently stimulate credit demand. Shrinking net interest margin on loans (about 0.6%-1.2% now relative to about 1.2%-2.1% in early 2000s) tends to limit banks’ core profitability as interest income accounts for more than two-thirds of banks’ total income.

In addition, major banks have remained resilient during the global financial crisis and have the capacity to take on more foreign exposures. They have abundant yen liquidity supported by a stable deposit base, and have further strengthened their capital adequacy (Tier-1 ratio at 12%) after the global financial crisis, in part to meet the Basel III requirements. The resilience of the balance sheets of Japanese banks has placed them in a better position to further expand overseas, despite lingering global uncertainty. The exchange rate has appreciated until recently, which may offer an additional incentive for expanding abroad.

Regarding regional and global factors, robust growth and large financing needs in Emerging Asia offer new business opportunities for Japanese banks. Japan’s proximity to the rest of Asia is an advantage. Major banks have also benefited from the increasing outward FDI and trade links of Japanese firms. Financing needs for infrastructure in Emerging Asia are large (about US$8t), according to the Asian Development Bank. These generate demand for cross-border financial activity between Japan and various FDI destinations (Figure 3). Moreover, the deleveraging of European banks since 2010 has accelerated the pace of overseas expansion. Japanese banks, among other local Asian banks, have stepped up financing to gain market share against the scaling back of European banks in the region.

The current trend of overseas expansion appears to have some differences from previous episodes, though new challenges are likely to emerge. In the past, Japanese banks have largely expanded abroad to support the corporate expansion of Japanese firms. But over the past few years, financial institutions have also moved toward extending loans to non-Japanese entities, now reaching 70% of overseas loans. Second, over time Japanese banks have accumulated net external foreign assets while the short-term

---

7 Estimates suggest that business capital expenditures for fixed investments only account for less than 30 percent of the total credit demand. External financing through capital increases has been negative (on average about ¥5 trillion per year) for the past decade due to weak equity markets. Bank lending has somewhat picked up recently, reflecting reconstruction and housing loans demand.
Japanese financial institutions expanding abroad: Opportunities and risks

liabilities have remained stable. While part of the increase is attributed to higher foreign assets held by trust banks, the increase may suggest that banks have relied on domestic yen-denominated funds to finance long-term overseas loans.

Empirical analysis in cross-border activity of Japanese banks
To analyze the role of these factors in contributing to the rising cross-border bank lending, the paper conducts an empirical analysis to assess determinants of banks’ overseas expansion. The analysis also provides insights into whether the current trend is different from previous episodes. Several other studies also looked into the factors contributing to cross-border banking (Shirota (2013), and Focarelli and Pozzolo (2005)) through factor analyses and institutional features.

The empirical analysis here uses the quarterly consolidated year-on-year growth of foreign claims on an immediate-borrowers basis published by the Bank of International Settlement (BIS). The sample is from 1984 to 2012, spanning across a panel of banking systems consisting of both origination and destination of cross-border claims. The origination countries/regions are mostly advanced economies, including Australia, Japan, France, Germany, Italy, Switzerland, Sweden, the U.K., the U.S., and developed Europe. The destination countries or regions include emerging Asia (China, India, ASEAN 5), developing Europe, and Latin America. The foreign claims (FC) are in U.S. dollar terms and are subject to valuation changes driven by exchange rate movements, which could be partly controlled by including the weighted exchange rates as an explanatory variable.8 The explanatory variables are broadly classified into three categories with the specification as follows:

\[
FC_{i,t} = \beta_0 + \gamma_i + \beta_{\text{GF}}^{FC} + \beta_{\text{RF}}^{FC} + \beta_{\text{HF}}^{FC} + \epsilon_{i,t}
\]

where \( i \) and \( j \) stand for origination and destination countries/regions of foreign claims, respectively. A fixed effect coefficient \( \gamma_i \) is included for each group. The explanatory variables include:

- **Global factors (GF)** - consisting of the VIX index, and the Fed Fund rate.
- **Regional factors (RF)** - consisting of growth differentials and real effective exchange rate movements between destination and origination countries/regions. The regression also considers alternative indicators of growth differentials using the change of fixed investments.
- **Home factors (HF)** - consisting of domestic interest rates, real effective exchange rates, growth of domestic credit to GDP ratio, several indicators for the soundness of the banking systems in origination countries/regions that include Tier-1 capital ratios, nonperforming loan ratios and the return on assets.

The empirical results underscore the role of various factors in contributing to the growth of banks’ foreign claims in Japan and other advanced countries (Table 2). First, higher global uncertainty (measured by the VIX index) tends to reduce banks’ activities abroad, though the net adverse impact on Japanese banks is relatively less than other countries. Second, in terms of regional factors, interest rates at the destination, as a proxy for the tightness of financing conditions, also play some role. Third, the growth differential is also an important driver for banks’ foreign claims. For instance, a 1 percentage-point increase in the real growth differential could increase the foreign claims by about 0.3–1.6 percentage-points. While a currency appreciation in the origination countries tends to increase banks’ activity overseas, the coefficients are not statistically significant across all specifications. Regarding home factors, higher domestic credit growth is generally associated with slower growth overseas, possibly suggesting some substitution in banks when extending credit between home and abroad. Moreover, the soundness of banking systems at home is statistically significant in banks’ overseas activities. Stronger banks’ balance sheets, such as higher capital adequacy ratios and lower nonperforming loan ratios, are often associated with higher cross-border activity.

The empirical results also provide insights into whether the current trend would continue, and how it may be affected by the new policy framework in Japan. Applying the estimated results to Japan would imply that global and regional factors play a key role in explaining the rise of foreign claims. As an illustration, Japanese banks’ foreign claims on Asia have grown by 103% since the end of 2008, of which about 40 percentage-points is attributed to a decline in global uncertainty, as proxied by the VIX index, while regional factors contributed another

8 Including the exchange rate as an explanatory variable controls partly for valuation changes in the BIS data. Strictly speaking, the exchange rate to be included should reflect the composition of foreign claims of origination countries/regions. By using the real effective exchange rate based on external trade weights as a proxy would imply an assumption that those weights are identical to those of foreign claims composition.
20–25 percentage-points. Regarding home factors, the resilience of the Japanese major banks, particularly the strengthening of capital adequacy and low NPLs during the global financial crisis, contributed to about one-third of foreign claims growth. The substitution between domestic and foreign credit contributed modestly, by about 5 percentage-points.

Outlook and policy implications

Expansions abroad by Japanese financial institutions is welcome, though a gradual and cautious approach in overseas strategies is warranted. Financial institutions’ expansion overseas helps improve their profitability by better allocating their liquidity and developing local markets in the Asian region. Banks may also favor a gradual expansion to maintain their balance sheets under

<table>
<thead>
<tr>
<th>Dependent variable: Foreign claims (year-on-year percentage change $i,j$)</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origination</td>
<td>European banks only</td>
</tr>
<tr>
<td>Destination countries/regions</td>
<td>All (1)</td>
</tr>
<tr>
<td>Fixed effects $i,j$</td>
<td>Y</td>
</tr>
<tr>
<td>Constant</td>
<td>171.1$^*$ (54.8)</td>
</tr>
<tr>
<td>Lagged foreign claims (yoy pct. change)</td>
<td>0.45$^*$ (0.05)</td>
</tr>
<tr>
<td>Global factors</td>
<td></td>
</tr>
<tr>
<td>Lagged VIX index</td>
<td>-1.01$^*$ (0.21)</td>
</tr>
<tr>
<td>Lagged U.S. Fed-fund rates</td>
<td>0.27$^*$ (0.09)</td>
</tr>
<tr>
<td>Regional factors</td>
<td></td>
</tr>
<tr>
<td>Real growth differential $i,j$</td>
<td>0.78</td>
</tr>
<tr>
<td>dln (REER differential $i,j$)</td>
<td>-4.62</td>
</tr>
<tr>
<td>Interest rates $i$</td>
<td>-1.19$^*$ (0.64)</td>
</tr>
<tr>
<td>Home factors</td>
<td></td>
</tr>
<tr>
<td>Interest rates $i$</td>
<td>-1.08</td>
</tr>
<tr>
<td>Lagged domestic credit/GDP $i$</td>
<td>-24.2$^*$ (11.0)</td>
</tr>
<tr>
<td>Soundness of banking system $i$</td>
<td></td>
</tr>
<tr>
<td>Capital adequacy (Tier 1 ratio)</td>
<td>17.45$^*$ (4.51)</td>
</tr>
<tr>
<td>Nonperforming loans ratios</td>
<td>-7.57$^*$ (2.05)</td>
</tr>
<tr>
<td>Returns on assets</td>
<td>-15.6</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.40</td>
</tr>
<tr>
<td>Number of observations</td>
<td>535</td>
</tr>
<tr>
<td>Number of origination countries or regions</td>
<td>5</td>
</tr>
<tr>
<td>Number of destination countries or regions</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2: Empirical analysis on bank’s foreign claims 1/2

Source: author’s estimates.
1. Notation $i$ and $j$ refer to origination and destination countries or regions of the foreign claims, respectively.
2. “a,” “b,” and “c” denote the statistical significance at 1%, 5%, and 10%, respectively.
Japanese financial institutions expanding abroad: Opportunities and risks

A rapid expansion could lead to buying foreign assets at high prices or entering unfamiliar local markets that could eventually result in heavy losses, as was the case in the late 1980s and 1990s.

Higher overseas exposures may add to funding risks that would require continued close monitoring by supervisory authorities. Securing stable and long-term dollar funding has remained a risk for Japanese financial institutions. Supervisors should encourage banks to further improve their resilience against shocks by strengthening their funding sources and risk management, such as by closely monitoring the overseas maturity mismatch and foreign currency-denominated loans-to-deposits ratios. At the same time, overseas activities add to the challenges of cross-border supervision of financial institutions. Cross-border risk monitoring arrangements with foreign supervisory authorities can help monitor risks from cross-border activities, including foreign exchange funding risks. In that regard, the supervisory agencies in Japan have signed the Multilateral Framework for sharing information on global systemically important banks (G-SIBs) collected through the FSB Data Gap Initiatives in early 2013 based on discussions at the Financial Stability Board (FSB).

According to the empirical results, the key determinants for Japanese banks to expand abroad are mostly driven by global and regional factors. Policies under Abenomics would mostly affect domestic outlook and exchange rates [IMF (2013)] in exiting deflation and lifting growth. Successful policies may also reinforce the trend of going overseas by financial institutions if there are positive spillovers globally and to the region. Recovery in domestic opportunities may slow the expansion pace but empirical estimates suggest that the substitution effect between domestic and overseas lending contributed modestly to the trend (about 5% in the growth of foreign claims in Japan).

Ensuring a complete package of Abenomics would be important for financial stability. An incomplete set of policies that fails to raise growth and reverse the rising public debt-to-GDP ratios, however, could pose risks for domestic financial stability, which could have implications for foreign exposures, as was the case in the late 1980s and in the early 2000s.

Conclusions
Cross-border activities of Japanese financial institutions have increased over the past few years, particularly in the Asian region. The relative resilience of Japanese banks during the global financial crisis has allowed them to take on further foreign exposures. Stagnant growth and low interest margins in Japan have added to incentives to seek opportunities abroad. Outside Japan, higher growth in Asia and deleveraging of European banks in the region also contributed to a rise in cross-border lending. The recent expansion has, in some aspects, been broader than previous overseas expansion waves of the mid 1980s and 1990s. Japanese financial institutions have broadened their funding sources and extended finance not only to Japanese corporates abroad but also to local firms.

As the global recovery takes hold and growth in Asia is expected to remain robust over the medium term, Japanese banks will likely continue the trend. Stronger domestic growth in Japan could slow the pace, but is unlikely to reverse a long-standing trend, because empirical estimates suggest that global and regional factors play a more prominent role in the growth of Japanese cross-border claims. An incomplete set of policies under Abenomics, however, could pose risks for financial stability that could halt overseas expansion.
But higher overseas exposures may add to funding risks that would require continued close monitoring by supervisory authorities. Securing stable and long-term dollar funding has remained a risk for Japanese financial institutions. At the same time, overseas activities add to the challenges of cross-border supervision of financial institutions. Overseas expansion by Japanese financial institution is welcome, but would warrant a gradual and cautious approach in light of earlier episodes in the late 1980s and 1990s.

References
Fitch Ratings, 2013, “Japanese mega banks’ offshore growth continues”, Fitch Ratings Special Report, February
International Monetary Fund, 2013, “Spillover report,” August
Appendix 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia Pacific</td>
<td>21</td>
<td>20</td>
<td>21</td>
<td>14</td>
<td>16</td>
<td>15</td>
<td>88,119</td>
</tr>
<tr>
<td>Japan</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>13,596 7,479 3,540 4,538 7,392</td>
</tr>
<tr>
<td>Australia</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>9,138 12,892 5,198 6,748 8,077</td>
</tr>
<tr>
<td>India</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>19,897 42,597 52,063 28,352 16,279</td>
</tr>
<tr>
<td>Greater China</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1,946 1,886 14,743 - - -</td>
</tr>
<tr>
<td>ASEAN</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4,061 1,616 2,298 1,419 3,767</td>
</tr>
<tr>
<td>Korea</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4,788 4,539 1,855 603 3,294</td>
</tr>
<tr>
<td>Europe</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>11</td>
<td>9</td>
<td>9</td>
<td>-</td>
</tr>
<tr>
<td>Eurozone</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>985 1,678 2,379 6,833 8,757</td>
</tr>
<tr>
<td>U.K.</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3,994 3,440 816 1,110 4,172</td>
</tr>
<tr>
<td>Others</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>

Table A1: Project finance in Asia by top mandated arrangers, by parent nationality

1. Average for top arrangers that were among the top 25 list for at least two consecutive years.
<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>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top 25 Mandated Arrangers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia Pacific</td>
<td>23</td>
<td>19</td>
<td>21</td>
<td>19</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>24,419</td>
<td>26,323</td>
<td>18,170</td>
<td>7,839</td>
<td>14,749</td>
</tr>
<tr>
<td>Australia</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>4,378</td>
<td>5,680</td>
<td>4,514</td>
<td>1,878</td>
<td>-</td>
</tr>
<tr>
<td>India</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>40,488</td>
<td>50,725</td>
<td>63,822</td>
<td>37,999</td>
<td>25,946</td>
</tr>
<tr>
<td>ASEAN</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>28,893</td>
<td>36,233</td>
<td>16,089</td>
<td>11,085</td>
<td>14,700</td>
</tr>
<tr>
<td>Greater China</td>
<td>9</td>
<td>4</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>54,256</td>
<td>31,773</td>
<td>63,762</td>
<td>33,502</td>
<td>24,996</td>
</tr>
<tr>
<td>Korea</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>9,672</td>
<td>5,002</td>
<td>-</td>
<td>-</td>
<td>14,145</td>
</tr>
<tr>
<td>Europe and N. America</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eurozone</strong></td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>-</td>
<td>7,614</td>
<td>9,610</td>
<td>6,310</td>
<td>14,763</td>
</tr>
<tr>
<td>U.K.</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>16,426</td>
<td>28,470</td>
<td>20,270</td>
<td>9,130</td>
<td>19,519</td>
</tr>
<tr>
<td>U.S. and Canada</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5,706</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Average ranking**

<table>
<thead>
<tr>
<th>Parent Nationality</th>
<th>Average ranking 1/</th>
<th>Amount of proceeds (in percent of industry total)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asia Pacific</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>ASEAN</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Greater China</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td><strong>Europe and N. America</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eurozone</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>U.K.</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>U.S. and Canada</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

**Table A2:** Syndicated loans in Asia Pacific (ex Japan) by top mandated arrangers, by parent nationality

1. Average for top arrangers that were among the top 25 list for at least two consecutive years.
Safe to fail

Thomas F. Huertas
Partner, EY LLP

Abstract
Banks cannot be made fail-safe. But they can be made safe to fail, so that the failure of a bank need not disrupt the economy at large nor pose cost to the taxpayer. In other words, banks can be made resolvable, and “too big to fail” can come to an end. To do so, the authorities, banks and financial market infrastructures need to prepare in advance for what amounts to a pre-pack reorganization of the bank that the resolution authority can implement over a weekend, if the bank reaches the point of non-viability in private markets (fails to meet threshold conditions). This pre-pack consists of two principal elements: (i) a recapitalization of the bank through the bail-in of investor instruments, and (ii) the provision of liquidity to the bank-in-resolution. Creating such a pre-pack solution should form the core of the resolution plans that authorities are developing for global systemically important financial institutions (G-SIFIs).

1 The paper draws extensively on the analysis presented in (Huertas (2013)) and has benefited from comments by Stefan Walter, Eva Huepkes, Maria Nieto, Markus Ronner, Wilson Erwin, John Whittaker, and David Schraa, as well as from a discussion at a seminar organized by the Financial Markets Group at the London School of Economics. The opinions expressed here are the author’s personal views and do not in any way represent the views of EY LLP, its partners, or any associated organization. Any remaining errors are the author’s responsibility.
Banks cannot be made fail-safe. But they can be made safe to fail, so that the failure of a bank need not disrupt the economy at large nor pose cost to the taxpayer. In other words, banks can be made resolvable, and “too big to fail” can come to an end. To do so, the authorities, banks and financial market infrastructures (FMIs) need to prepare in advance for what amounts to a pre-pack reorganization of the bank that the resolution authority can implement over a weekend, if the bank reaches the point of non-viability in private markets/fails to meet threshold conditions. Creating such a pre-pack solution should form the core of the resolution plans that authorities are developing for global systemically important financial institutions (G-SIFIs).

We start off by setting out the conditions that must be met for a bank to be resolvable. This paper then outlines that this “safe-to-fail” test can be met under a variety of banking structures under a so-called Single Point of Entry approach, where the home country resolution authority acts as what amounts to a manager of a global resolution syndicate (Annex A deals with the Multiple Point of Entry approach). How banks are organized matters less than what banks, authorities, and financial market infrastructures do to prepare for the possibility that resolution may be required. That agenda for action concludes the paper.

Resolvability

Resolution reform aims to make feasible the resolution of financial institutions without severe systemic disruption and without exposing taxpayers to loss, while protecting vital economic functions through mechanisms that make it possible for shareholders and uninsured and uninsured creditors to absorb losses in a manner that respects the hierarchy of claims in liquidation (FSB [2011]).

An institution is, therefore, resolvable, if three conditions are met: (1) it can be readily recapitalized without recourse to taxpayer money; (2) in resolution it can continue to conduct normal transactions with customers, ideally from the opening of business on the business day following the initiation of the resolution; and (3) the resolution process itself does not significantly disrupt financial markets or the economy at large.

The resolution timeline

Resolution falls into three phases: pulling the trigger, stabilizing the institution and restructuring the institution (Figure 1).

Pulling the trigger initiates resolution: for the purpose of this discussion, we assume that the trigger is pulled upon a finding (usually by the bank’s supervisor) that the bank has reached the point of non-viability (it no longer meets threshold conditions). We also assume that the trigger is pulled at the end of the business day, ideally on a Friday, so that the resolution process takes place over a weekend, when markets are closed. As a practical matter, for a G-SIFI, the end of the business day is in all likelihood the end of the business day in the U.S., for it is when the U.S. market closes that there is a period of hours before the next market opens in Asia.

Once the trigger has been pulled, resolution begins. In line with the requirements set out by the FSB in its Key Attributes paper (FSB [2011]), we assume that the resolution regime has designated a resolution authority for the jurisdiction and empowered such an authority to make decisions with respect to the bank-in-resolution without prior judicial review.

The work of the resolution authority falls into two distinct phases: stabilization and restructuring. The stabilization phase covers the period between the point at which the trigger is pulled (e.g., close of business Friday) and the opening of the next business day.

---

2 For a description of the Single and Multiple Point of Entry approaches see FSB (2012).

3 Normal transactions would include payments and settlement of securities trades and various other ‘non-investment’ transactions with both individual and institutional customers. In contrast, investment obligations would be subject to a stay (e.g., on the payment of interest and dividends or the repayment of capital instruments), as outlined below.

4 The importance of this assumption cannot be overstated. Pulling the trigger during the course of the business day greatly compounds the potential disruption to financial markets that the bank’s failure could cause. Although much has been done to improve the robustness of financial market infrastructures (e.g., introduction of real time gross settlement in payment systems and introduction of delivery versus payment in securities settlement systems), allowing a major bank to fail during the course of a business day could still cause significant disruption, a phenomenon known as Herstatt risk, in reference to the disruption caused by the failure of Herstatt Bank in 1974 while markets were still open.

5 The phrase ‘when markets are closed’ requires some qualification. It is common for banks to offer customers (especially consumers) 24-hour access to their accounts seven days a week via internet banking and/or automated teller machines. Such access may need to be temporarily halted over the resolution weekend in order to effect the stabilization of the failed bank. Thought also needs to be given to how so-called ‘in-flight’ transactions are to be handled, particularly if the resolution does not provide for continuity.

6 This timing factor gives the U.S. a disproportionate influence in determining when the trigger should be pulled to put a G-SIFI into resolution. In particular, if the U.S. were to decide to put the U.S. operations of a G-SIFI into resolution on the grounds that the U.S. operations did not meet U.S. standards for capital and liquidity, it is highly likely that the rest of the group would quickly follow into resolution. The recent U.S. proposal [FRB (2012)] for the regulation of foreign banking organizations (FBOs) in the U.S. further heightens such concerns, as the U.S. proposes to impose requirements on the U.S. operations of FBOs that are higher than those imposed in the Basel Accord and makes no reference to cooperation with the host country authorities.
In practical terms, this means that the stabilization phase for a G-SIFI lasts no more than 36 to 48 hours, from close of business in North America on a Friday to opening of business in Asia on Monday. If the stabilization succeeds, customers will continue to be able to transact with the bank-in-resolution, much the same as airline passengers who are able to continue flying on airlines that are in bankruptcy.

The restructuring phase is open-ended. It can take months, or even years, but the objective will be to return the bank to the private sector as soon as possible. The resolution authority will act in the same capacity as an administrator in a bankruptcy proceeding and may take decisions to sell assets (including subsidiaries, lines of business and individual assets), reconfigure businesses or discontinue them entirely.

This paper focuses on the stabilization phase. It makes the assumption that the supervisor pulls the trigger when the bank reaches “the point of non-viability,” i.e., the point at which the bank is no longer able to finance itself in private markets, and that this point corresponds to the point at which the bank no longer meets threshold conditions. In other words, the authorities do not exercise forbearance.

Meeting condition 1: the institution can be readily recapitalized without recourse to taxpayer money

Bail-in can enable banks to meet condition 1. This effectively creates what amounts to reserve capital and allows the resolution authority to utilize instruments other than common equity to absorb loss. This should be done in accordance with strict seniority, so that common equity bears first loss, then non-common equity Tier-1 capital (e.g., preferred stock), then Tier-II capital (e.g., subordinated debt), then other ‘investor’ obligations such as senior debt. Such investor instruments should be subject to mandatory bail-in immediately upon the bank entering resolution.

Four caveats are in order:

1. The mandatory bail-in must generate enough capacity to absorb loss and recapitalize the bank to at least the minimum required level: this implies that the total reserve capital, the investor instruments subject to mandatory bail-in (non-core Tier-I capital, Tier-II and senior debt subject to mandatory bail-in), should be at least equal to the required common equity Tier-I capital. If this is the case, the mandatory bail-in would effectively recapitalize the bank, even if the entire amount of common equity Tier-I capital had to be written off. Note that pulling the trigger promptly (i.e., at the point at which the bank fails to meet threshold conditions — reaches the point of non-viability) greatly enhances the probability that the mandatory bail-in of investor instruments will be sufficient to recapitalize the bank, for such intervention will occur at a point where the bank still has positive net worth as it enters resolution.

---

7 The term “bank-in-resolution” also includes successor institutions, such as bridge banks, that may be created during the course of resolution by the resolution authority.

8 A second paper [Huertas (forthcoming)] will focus on the restructuring phase.

9 Note that the “waterfall” described here assumes (see caveat (4)) that senior debt subject to mandatory bail-in is subordinated to deposits. If senior debt is pari passu with deposits, bailing-in senior debt whilst keeping deposits whole will give rise to potential compensation payments from the resolution fund under a “no creditor worse off” criterion (Huertas (2013)). Note as well that the “waterfall” does not stop as a matter of law (and should not stop) with senior debt. If losses exceed the total amount of investor obligations, then bail-in should extend to more senior obligations, such as deposits. From a policy standpoint, the question then arises as to whether insured deposits should have preference over uninsured deposits (as is proposed for the ring-fenced retail and commercial bank under the UK ICB legislation) or whether all deposits should be pari passu with one another, as well as whether deposits should have preference over other obligations. This has implications for the risk to the deposit guarantee scheme and to the contribution that such schemes could be expected to make to loss absorption in the event of resolution.
2. The legal and contractual framework should be in place to allow the resolution authority to execute the mandatory bail-in of investor instruments immediately upon the entry of the bank into resolution: to assure that this will be the case, the relevant law(s) should give the resolution authority the statutory power to implement mandatory bail-in, and the bank should complement this with contractual provisions and information disclosures to investors that make clear that the instrument will be subject to mandatory bail-in, if the bank goes into resolution.

3. The implementation of mandatory bail-in should not in and of itself trigger cross-default clauses in customer obligations, such as derivatives or repurchase agreements: the mandatory bail-in of investor obligations should recapitalize the bank and enable the bank to meet its customer obligations. It would be counterproductive to allow mandatory bail-in itself to be an event of default that would allow derivative counterparties to trigger close out and/or allow derivative and repo counterparties to liquidate collateral that such counterparties may have received from the bank at the point at which the bank goes into resolution. If mandatory bail-in effectively recapitalizes the bank, this should provide sufficient immediate protection to counterparties. They should only be allowed to invoke close-out and/or liquidate collateral if the bank-in-resolution defaults on a payment due.

4. Implementing mandatory bail-in will be easier if such instruments are explicitly subordinated to other obligations of the bank: this will already be the case for obligations of a parent holding company, as they are structurally subordinated to obligations of the subsidiary bank. This will also be the case for non-core Tier-I and Tier-II capital instruments issued at the bank level. However, this will not be the case for senior debt issued at the bank level. This is pari-passu with customer obligations, such as deposits and derivatives. To the extent that debt senior to Tier-II capital would be counted toward the minimum amount of reserve capital under caveat (1), such “senior” debt should really be a mezzanine facility, senior to subordinated debt, but junior to customer obligations and to debt pari-passu with such customer obligations.

Meeting condition 2: the institution in resolution can continue to transact with customers from the opening of business on the business day following the initiation of the resolution. For the stabilization phase to be successful, the bank-in-resolution needs to be able to continue to meet customer obligations. If the bank enters resolution at the close of business in North America on a Friday evening, it needs to be able to reopen for business as usual in Asia on Monday morning Asia time. In particular, it will need to be able to meet the demand of customers (e.g., holders of current accounts, repo providers, holders of maturing time deposits) who have an immediate claim on the bank.

For the purpose of this discussion, we assume that the bank-in-resolution has met condition 1. Mandatory bail-in has recapitalized the institution without recourse to taxpayer money. As a result, the bank-in-resolution is solvent and can potentially remain in operation while its capital is being restructured. However, this will require:

1. The bank-in-resolution to continue to be authorized to operate as a bank: the resolution regime should assure that the bank-in-resolution receives immediate authorization to operate as a bank, and that the resolution authority has the power to continue the operations of the failed bank.

2. The bank-in-resolution to retain capability to continue to operate: if the bank-in-resolution is to continue to transact with customers, provision should be made to assure that the entry of the bank into resolution does not cause suppliers of operational and technological support services to cut off provision of these services to the bank-in-resolution. To assure continuity in the event of resolution, the bank should conclude service-level agreements (SLAs) with suppliers (including other affiliates in the banking group) that continue in force even if the bank enters resolution. Note that achieving this objective may require the

---

10 In contrast, the resolution authority should have a reserve power to bail-in non-investor instruments, such as deposits and derivatives, upon a finding that losses are likely to exceed the total amount of investor instruments. See comments on valuation below.

11 The one to two day stay on the ability of counterparties to close out derivative transactions included in some legislation (e.g., U.S. Dodd Frank Act) only partially addresses this caveat, for it does not preclude the counterparty from initiating close out after the stay has expired. There is a presumption that the counterparty will accept an assignment of the contract to the bridge institution (bank-in-resolution), but there is no requirement that it does so. Nor does such a stay apply to contracts that are concluded outside the U.S. under non-U.S. law.

12 Note that there may be a broader range of instruments subject to bail-in than those subject to mandatory immediate bail-in. Ideally, these would also be subordinated to customer obligations, but that need not be the case.

13 Some have suggested that it might be acceptable for the bank-in-resolution to reopen after a one to two day stay or suspension of operations (IF (2011)). However, such an interruption to continuity could create complications at financial market infrastructures and cause contagion to other financial institutions, financial markets, and to the economy at large.
3. The bank-in-resolution to have access to financial market infrastructures: if the bank-in-resolution is to continue transacting with customers, it will need access to financial market infrastructures (FMIs), such as payment systems, securities settlement systems and central counterparties. Accordingly, authorities responsible for the regulation of FMIs should take measures to assure that the mere entry of a bank into resolution does not automatically end its access to the FMI. As long as the bank-in-resolution continues to meet its obligations to the FMI, the FMI should continue to allow the bank-in-resolution access to the FMI. This continued access should follow two precepts: (i) no acceleration of obligations due from the failed bank at the point at which it enters resolution, unless the bank-in-resolution fails to meet its obligations to the FMI at the close of business on the day the bank entered resolution (see condition 1), but (ii) freedom of the FMI to insist on risk-limitation measures (such as the provision of collateral or the requirement to make payments to the FMI in central bank money) for new transactions of the bank-in-resolution with the FMI.

4. The bank-in-resolution to have access to adequate liquidity: most importantly, the bank-in-resolution will need to have access to adequate liquidity if it is to be able to meet customer obligations from the opening of business on the business day following the entry of the bank into resolution. This is akin to the debtor-in-possession financing that banks provide in connection with restructurings under bankruptcy proceedings for non-financial corporations. In all likelihood, the central bank(s) or resolution authorities will be the only source of such a liquidity facility in the amount and with the speed that a bank-in-resolution is likely to require.14 According to central bank doctrine, a central bank should lend to a solvent but temporarily illiquid bank secured by sound collateral.15 The mandatory bail-in of investor obligations should assure that the bank-in-resolution is solvent and that the door to the central bank and/or resolution authority is open to provide the liquidity facility. The actual facility should be on a super-senior basis and secured by the bank's unencumbered assets. As a practical matter, the provider of such a liquidity facility will want to assure that it can track and take a charge over the bank's unencumbered assets, and banks' resolution plans will need to reflect this. Banks and central banks will also want to assure that the central bank can smoothly take over any collateral released by counterparties, such as repo providers, that demand repayment from the bank-in-resolution. Resolution planning should also give consideration to the contract that central bank(s) might wish to be used for such a facility (but stop short of the central bank actually giving a commitment to a bank that such a facility would actually be granted so as to not fetter the discretion of the central bank). Finally, central bank(s) will want assurances that they will not be ultimately responsible for bearing any losses that might be incurred on the provision of such a liquidity facility to the bank-in-resolution, should it fail to repay the facility and liquidation of the collateral provided by the bank prove insufficient to do so. This assurance should come from a resolution fund, financed by a levy on all banks, that would compensate the central bank for any losses that the central bank might incur through the provision of liquidity to the bank-in-resolution.16

Meeting condition 3: the resolution process itself does not significantly disrupt financial markets or the economy at large. Finally, the resolution process should not, in and of itself, significantly disrupt financial markets or the economy at large. To achieve this result:

1. The resolution process should not come as a surprise to the market: the shift from bailout to bail-in should be well advertised to investors, not sprung on them by surprise, as it was arguably done in the case of Lehman Brothers in 2008 [Huertas (2011)]. The revision of resolution regimes, the introduction of resolution planning and the conduct of resolution policy all point in this direction, as does the increased dependence of pricing and

---

14 In the U.S., under Dodd Frank, the resolution authority (FDIC) is responsible for providing such a liquidity facility to the bank-in-resolution and such a facility is subject to certain quantitative limits. The Federal Reserve is prohibited from extending an institution-specific credit to the bank-in-resolution, but may create a general market facility open to all banks, including the bank-in-resolution. In other jurisdictions (such as the U.K.), the central bank can provide liquidity to the bank-in-resolution under its general powers to act as a lender of last resort.

15 The central bank should certainly charge the bank-in-resolution a penalty rate (i.e., adds a spread or premium to the market rate) in order to induce the bank-in-resolution to replace central bank funding with funding from private sources as soon as possible, the central bank should avoid setting that spread at punitive levels that would undermine the ability of the resolution authority to restructure the institution.

16 For details see FSB (2013a). Note that the obligation to be covered by such a resolution fund differs from that to be covered by a deposit guarantee scheme (the coverage of insured deposits up to a limit). This implies that two separate funds and two separate levies may be required, particularly where deposits have preference (and especially where insured deposits have preference).
Safe to fail

ratings for instruments subject to mandatory bail-in on a bank’s stand-alone risk (and correspondingly reduced reliance on implicit government support).

2. The resolution process should not accelerate fire sales of assets: if the resolution process requires the bank-in-resolution to conduct or empowers its counterparties to conduct fire sales of assets, it can have an adverse knock-on effect on the market as a whole. Although such fire sales enable the seller to raise cash, they depress the price at which assets must be valued in mark-to-market portfolios across the entire market. That will generate losses in such portfolios and reduce capital at banks and other financial institutions, possibly causing one or more such institutions to experience liquidity pressures, even if the institution had no direct exposure to the bank-in-resolution. In other words, fire sales are a possible transmission mechanism for contagion. The likelihood of fire sales will be reduced, if the resolution process meets conditions (1) and (2). In particular, if the entry of the bank into resolution does not trigger close-out of derivatives, it will reduce the adverse impact on that market as well as on the market(s) for any collateral that the bank-in-resolution may have posted with derivative counterparties. Similarly, asset markets will be less volatile if repo providers to the bank-in-resolution are not entitled to simply liquidate the collateral that the bank-in-resolution had pledged. In effect, the resolution process outlined in conditions (1) and (2) enables the bank-in-resolution to continue to meet its obligations to derivative counterparties and repo providers, so that they have no need to close out or liquidate collateral pledged by the bank-in-resolution.

3. The resolution process should not interrupt clients’ access to their assets: once the bank-in-resolution opens for business on Monday, clients should be able to access their accounts and assets as normal. The resolution process should not freeze client assets, restrict client transactions, or limit clients’ access to their money.\textsuperscript{17}

4. The resolution process should not trigger the failure of financial market infrastructures (FMIs): finally, the resolution process should leave FMIs intact and able to continue to fulfill their functions. This will certainly be the case if FMIs are themselves robust, i.e., able to withstand the simultaneous failure of their two largest participants, as called for under the CPSS-IOSCO (2012) principles. But it may also be the case, if the resolution process for a G-SIFI meets conditions (1) and (2) as outlined above, for the bank-in-resolution would continue to fulfill its obligations to the FMI. As far as the FMI is concerned, there would be no participant failure, and the FMI should remain robust.

In summary, if a bank meets the three conditions outlined above it will be resolvable. In other words, the bank will be safe to fail – its failure will not pose solvency costs to the taxpayer nor will its failure significantly disrupt financial markets or the economy at large.

\textbf{Which banking structures can meet the safe-to-fail test?}

We now turn to the question of which banking structures can meet the safe-to-fail test outlined above. We consider two cases: (a) where the parent organization for the group is a bank, and (b) where the parent organization is a holding company that owns one or more banks as operating subsidiaries.

\textbf{Bank as parent company}

We start with the case where the bank itself is the parent company, and this bank operates in a single jurisdiction (A). Here, the conditions outlined above apply directly. If the bank meets those conditions, it will be safe to fail.

As a practical matter, the authorities, banks and financial market infrastructures (FMIs) need to prepare in advance for what amounts to a pre-pack reorganization of the bank that the resolution authority can implement over a weekend, if the bank reaches the point of non-viability in private markets (i.e., fails to meet threshold conditions). This pre-pack consists of two principal elements: a recapitalization of the bank through the mandatory bail-in of investor instruments; and the provision of liquidity to the bank-in-resolution through what amounts to debtor-in-possession financing.

\textbf{Implementation of bail-in}

For bail-in to operate effectively there has to be enough “reserve capital” (instruments subject to mandatory bail-in) to recapitalize the bank. Law and regulation should assure that the aggregate amount of investor instruments subject to mandatory bail-in would be sufficient to recapitalize the bank, even if all of its common equity Tier-I capital had to be written off. In aggregate, therefore, the bank’s non-core Tier-I capital, Tier-II capital and

\textsuperscript{17} An exception to this statement might be made in the event that failures to segregate client money and/or client assets caused the bank to reach the point of non-viability (fail to meet threshold conditions), and therefore be put into resolution.
For mandatory bail-in to operate smoothly and efficiently:

- The resolution authority should have the statutory authority to impose bail-in. This should be anchored in the resolution regime as a matter of law or regulation, and specify the instruments to which mandatory bail-in would apply. The statute should empower the resolution authority to implement bail-in immediately upon the entry of the bank into resolution without prior judicial review and without the ability of investors in instruments subject to mandatory bail-in to seek injunctive relief.
- This statutory provision for bail-in should be reinforced by contractual provisions in the instrument itself, especially where the instrument is issued in a jurisdiction other than jurisdiction A (where the bank is headquartered) and/or issued to investors resident outside jurisdiction A.
- The statutory provision for bail-in should also be reinforced by disclosure. The bank should disclose to investors in instruments subject to mandatory bail-in that they are so subject should the bank enter resolution. This disclosure should be ongoing, including without limitation any prospectus that accompanies new issues of instruments subject to bail-in as well as ongoing communications (e.g., websites, annual reports) with investors and rating agencies.
- There should be a clear separation between customer obligations and obligations subject to mandatory bail-in immediately after the trigger for resolution is pulled. In particular, investor obligations subject to mandatory bail-in should be subordinated to deposits, the quintessential customer obligation. This will be the case for non-core Tier-I capital and Tier-II capital (subordinated debt), and can be done as a matter of regulation and contract for senior debt subject to mandatory bail-in. Note that depositor preference alone will not assure that there is a sufficient amount of non-deposit liabilities available to bail-in should the bank fail to meet threshold conditions and need to be recapitalized. There should be an explicit requirement that the bank issues a minimum amount of instruments subject to mandatory bail-in and subordinated to deposits. Without such a minimum requirement, the introduction of depositor preference would induce banks to fund on a collateralized or secured basis, so that wholesale funding is again on a par or even senior to deposits (once the collateral backing the facility is taken into account).
- The resolution authority should effectively conduct its activities as a trustee for the creditors of the bank-in-resolution, especially for the investors who have been subjected to mandatory bail-in. At a minimum, the resolution regime should assure investors in instruments subject to mandatory bail-in that they will be no worse off than they would have been, had the bank been liquidated.
- Bail-in should be consistent with the principles of strict seniority. Losses should be apportioned according to a waterfall, with common equity absorbing first loss, then non-core Tier-I capital (e.g., preferred stock), then Tier-II capital (e.g., subordinated debt) and finally senior debt subject to bail-in. Any proceeds from the bank-in-resolution should be paid to investors in reverse order (i.e., senior debt first). Provision should also be made to allow holders of instruments subjected to mandatory bail-in to make an offer to convert such claims into common equity Tier-I capital in the bank, as a means of returning the bank to the private sector. Note that the waterfall does not necessarily end with senior debt subject to mandatory bail-in. It is possible that the losses at the bank-in-resolution may be so great as to burn through all of the “reserve capital” (instruments subject to mandatory bail-in), so that other investor obligations, such as senior debt not subject to mandatory bail-in as well as customer obligations, such as deposits, would also be subject to loss. Unless the deposit guarantee scheme assumes such loss and provides for continued access of depositors to their funds, bailing-in deposits greatly diminishes the likelihood that the bank can be resolved in a manner that assures continuity.

---

18. This is consistent with the total capital requirements for Swiss headquartered banks under the so-called ‘Swiss finish’ as well as with the requirements for the U.K. ring-fenced retail and commercial bank to hold primary loss absorbing capacity of 17% of risk weighted assets. It is also consistent with the ECFIN common position on the E.U. Bank Recovery and Resolution Directive (BRR (2013)).
19. For a further discussion of disclosure under bail-in see Huertas (2012).
20. In other words, only senior debt subordinated to deposits would count toward the requirement to keep outstanding a minimum amount of instruments subject to mandatory bail-in.
21. Note that it may be sensible to accept departures from strict seniority (as there are in the bankruptcies of a non-financial corporation) if there is a conversion of obligations into new equity during the restructuring phase. It may also be sensible to allow the junior creditors as a class to buy out the claims of the next most senior class at par plus accumulated interest and to take over the rights of that senior class (including the right to convert such claims into equity in the ‘new’ bank). For further discussion see Huertas (forthcoming).
22. In the U.S., the FDIC has resolved banks in a manner that protects all deposits, including uninsured deposits. An example is the purchase and assumption transaction used to resolve Washington Mutual (WaMu) in 2008. This assured continuity for WaMu depositors.
Safe to fail

Figure 2 illustrates the way in which bail-in could work. When the bank reaches the point of non-viability, the supervisor declares that the bank fails to meet threshold conditions and puts the bank into resolution. The resolution authority immediately bails-in the non-core Tier-I capital, the Tier-II capital and the senior debt subject to mandatory bail-in. This expands the immediate loss-bearing capacity of the bank and effectively recapitalizes it. In exchange for their original instruments, investors subject to mandatory bail-in obtain receivership certificates that entitle them to the proceeds that the resolution authority may, over time, realize from restructuring the bank-in-resolution. Such proceeds are distributed in accordance with strict seniority. Proceeds go first to holders of certificates (senior proceeds note) representing the claims of holders of senior debt subject to bail-in. Once these claims have been fully satisfied, any remaining proceeds are distributed to more junior creditors, again according to strict seniority. To the extent that a creditor receives less than it would have done had the bank been liquidated, the creditor has a claim for compensation for the difference on the resolution fund [IIF (2011); BRR (2013)].

Figure 2: Bail-in via stay on investor capital

Liquidity

As emphasized above, implementing mandatory bail-in of investor instruments is only the first step in the stabilization process. Successful stabilization requires not only recapitalization of the bank-in-resolution, but also provision of liquidity to the bank-in-resolution. Only the two measures taken together can assure continuity, and therefore minimize any adverse impact on the financial markets and the economy at large.

The framework for such a liquidity facility needs to be put in place well in advance of the bank being put into resolution. The framework should cover four factors:

1. The priority of the liquidity facility relative to other liabilities on the bank-in-resolution. As a practical matter, liquidity facilities to the bank-in-resolution will need to be on a super-senior basis so that they would have priority in liquidation over all other unsecured creditors.
2. The pool of collateral backing the facility. As a practical matter, this should be a charge over the unencumbered assets of the bank-in-resolution, including without limitation the investments of the parent bank in its subsidiaries. Any proceeds from asset sales should go toward repaying the facility.
3. The allocation of loss should the bank-in-resolution fail to repay the facility and the liquidation of the collateral prove insufficient to repay the facility. As noted above, provision should be made to recoup from the industry any loss that the resolution authority/central bank might suffer.
4. How and where the bank-in-resolution might draw on such a liquidity facility.

Note that the issuance of such proceeds notes greatly reduces the need to conduct an immediate valuation of the bank-in-resolution for the purpose of apportioning ultimate loss. Provided the authorities do not engage in forbearance (allow banks that fail to meet threshold conditions to continue in operation), losses should be less than the amount of the bank’s primary loss-absorbing capacity (common equity plus instruments subject to mandatory bail-in). Consequently, the valuation immediately required at the point of resolution is: (1) an assessment that the bank has reached the point of non-viability (so that the trigger to resolution is pulled); (2) an assessment that losses will not be greater than the amount of investor capital (primary loss absorbing capacity); and (3) an assessment of the advance rate that the central bank is willing to make on the unencumbered assets that the bank-in-resolution will pledge to the central bank as collateral for the liquidity facility that the central bank provides to the bank-in-resolution.

23 Ideally the framework would also be disclosed, certainly to host country central banks and resolution authorities (see international considerations below), to the bank itself, and to investors. Such disclosure would also go some way to surfacing and addressing the political objections that might be made to such a facility, particularly if the facility is a global one.
International considerations

We now turn to the situation where the bank is active in more than one jurisdiction, and start with the simplest scenario—a bank headquartered in jurisdiction A with a branch in jurisdiction B. Such a bank will be safe to fail if the resolution process follows the same principles as outlined above for a bank that operates solely within a single jurisdiction.

Briefly put, this will be the case if resolution is a unitary process, i.e., there is a single resolution process initiated and implemented by the home country resolution authority (see Figure 3), and such a process follows the principles outlined above for the case of a bank operating in a single jurisdiction. In such a unitary process, the assets and liabilities of the foreign branch are treated as an integral part of the bank as a whole. In such a unitary process the home country central bank would have to arrange for a liquidity facility that would also cover the bank’s foreign branch (indeed, if the foreign branch actually opens before the head office on the day after the bank enters resolution, the first draw on the liquidity facility is likely to be in the foreign jurisdiction).

This will require that the home country central bank make arrangements with the foreign country central bank(s) as to the role that the foreign central bank will play in such a liquidity facility to the bank-in-resolution. Two approaches are possible:

1. The foreign central bank acts as an agent of the home country central bank, so that any losses from extending the facility (if the proceeds from liquidating the collateral are insufficient) would accrue to the home country central bank (before it recouped such losses from assessments on the home country resolution fund). Such an agency approach enables the liquidity facility to be based on a single global collateral pool and for such collateral to support drawings on the facility wherever they might occur.

2. The foreign central bank acts as principal and extends credit solely on the basis of the collateral that the bank-in-resolution pledges to it. This implies that each central bank (the home country and the foreign central bank(s)) has access to a separate pool of collateral and has a separate lending agreement with the bank-in-resolution.

Although either of the approaches to liquidity provision is technically possible, the first, a unitary approach to liquidity provision, is more consistent with a unitary approach to resolution.

In contrast, under a territorial approach, resolution occurs separately within each jurisdiction (see Figure 4). In particular, the host country has the right to ring-fence the assets and liabilities of the branch in the host country, and liquidate the assets of the branch and use the proceeds to meet the liabilities of the branch to host country creditors (so that creditors of the host country branch have a preferential claim on the assets of the branch in the host country). Note that under the territorial approach, the host country may also have (take) the right to initiate resolution. Such a case may be envisioned if the bank fails to meet net asset requirements (equivalent to branch capital) and/or fails to meet local (branch) liquidity requirements. In the event that the host country puts the foreign branch into resolution, the home country may have no choice but to put the rest of the bank into resolution.

From the standpoint of investors in instruments subject to mandatory bail-in, the territorial approach creates a class of assets (the assets of the host country branch) that are segregated for the benefit of a specific class of liability holders (in this case the creditors (e.g., depositors) of the branch in the host country). If the host country authorities have the unrestricted right to sell such assets, they may have an incentive to do so at a discount so as to effect a quick sale. Indeed, one of the motives for the host country’s imposing a net asset requirement on the host country branch of a foreign bank is precisely to afford the host country resolution
The territorial approach is likely to impose higher losses on instruments subject to mandatory bail-in, than a unitary approach.

More importantly, the territorial approach creates a bias toward liquidation, with a greater loss of value to creditors and a greater possibility of disruption to financial markets and the economy as a whole. The territorial approach breaks the bank into pieces and effectively creates two separate banks in resolution, not one. Indeed, if the host country decides to liquidate separately the host country branch of the foreign bank-in-resolution, it will be difficult, if not impossible, for the bank-in-resolution to conduct new international transactions and difficult, if not impossible, for the home country bank-in-resolution to avoid the triggering of cross-default clauses in derivative and repo contracts. This will make it difficult, if not impossible, for the home country bank-in-resolution to preserve continuity with respect to its operations. Indeed, losses under the territorial approach are likely to be disproportionately greater for creditors of head office (as they do not benefit from the assets segregated behind the host country’s ring fence for the benefit of depositors in/creditors of the branch in the host country).

Banking organizations with holding company as parent

We now consider the case where the banking organization is structured as a parent holding company with a bank subsidiary. Can such an organization be safe to fail? Briefly put, the answer is yes, provided certain conditions are fulfilled.

We start with the simplest case, where the banking organization consists solely of a parent holding company and a single bank subsidiary, wholly owned by the parent holding company (see Figure 5), both headquartered in jurisdiction $A$. Assume, as is likely to be the case, that the loss causing the group to reach the point of non-viability originates in the subsidiary bank. This leads to a write-down of the equity in the subsidiary bank and a reduction in the value of the parent holding company’s investment in the subsidiary bank. This may be sufficient to wipe out the equity of the parent holding company.

Bail-in at the parent holding company can recapitalize the holding company, but it will not recapitalize the subsidiary bank. This requires supplemental measures, such as bail-in at the bank subsidiary and/or the issuance of new equity by the bank subsidiary to the parent holding company in exchange for cash from the parent. Liquidity facilities for the bank-in-resolution will also need to be arranged. Without such supplemental measures, stabilization will fail and continuity will not be achieved. Table 1 illustrates how bail-in could work in the situation where a parent holding company owns a domestic bank subsidiary. At the point of intervention, the bank subsidiary writes down its loan portfolio from 700 to 600. This loss of 100 wipes out the bank’s common equity of 100. It also causes a write-down of 100 in the value of the parent holding company’s asset, “equity in bank subsidiary.”

Bail-in should occur at two levels: the subsidiary bank and the parent holding company. The former is actually more important. In the example, bail-in at the parent converts preferred stock,
subordinated debt, and senior debt issued by the parent to third-party investors into primary loss absorbing capacity (in a manner similar to that depicted in Figure 2). Following bail-in at the parent level, PLAC is 200, corresponding to assets of 100 in marketable securities, 50 in senior debt issued by the subsidiary bank, 25 of subordinated debt, and 25 of preferred stock.

Without bail-in at the bank level, nothing changes at the bank level. The write-down in the loan portfolio has wiped out the equity of the bank. If the bank is to be stabilized, the bank must be recapitalized. This can occur either through the issuance of new equity by the bank to the parent (e.g., the parent would exchange its 100 of marketable securities for 100 of new equity in the bank subsidiary) or through a bail-in of instruments at the bank level. Such a bail-in process will work most smoothly where the parent holding company owns all of the instruments that are subject to mandatory bail-in at the bank level.24 If this is not the case, some simplicity may be preserved, if the parent holding company agrees as a matter of contract to subordinate its holdings of an instrument subject to bail-in to those held by third parties. This is arguably consistent with the fact that the parent will, or should have, greater and/or timelier information about the state of the bank subsidiary than the third-party investor, as well as by the fact that such subordination facilitates the retention of control of the bank subsidiary by the investors in the parent holding company. This concept is illustrated in Table 1. The parent holding company owns the entire amount of preferred stock (25) and subordinated debt (25), but only a portion (50) of the senior debt (200) issued by the bank subsidiary. The rest (150) is held by third-party investors. If such debt held by third-party investors is bailed in, control over the bank subsidiary will effectively pass to such investors. In the example, the senior debt of the bank issued to the parent holding company is assumed to be contractually subordinated to the senior debt issued by the bank to third parties. The senior debt issued to the parent holding company is subjected to bail-in; that issued to third parties is not.

International considerations
We now look at the case where the banking organization consists of a parent holding company headquartered in the home country with subsidiaries banks in both the home and the host country (Figure 6). In such a situation, the banking organization will be safe to fail if the home country resolution authority takes a unitary approach to resolution and treats foreign subsidiaries the same way as it would treat domestic subsidiaries. This implies that the home country resolution authority would take measures to assure that the foreign bank subsidiary could and would be as promptly recapitalized as a domestic bank subsidiary, in the event that the banking organization required resolution.

As a practical matter, this is only likely to be the case if the subsidiary bank in the host country has issued to the parent holding company instruments subject to bail-in in an amount sufficient to recapitalize the host-country bank, should losses at the host-country bank wipe out its common equity Tier-I capital. Such an arrangement would provide to the host country resolution authority the up-front assurance that the parent holding company will, in fact, have acted as a source of strength to the host-country bank, should the host country bank experience severe losses.25

Without such up-front assurance, the host country authorities would have to be concerned that either the parent holding company or the home country resolution authority would exercise their option to walk away from a failed subsidiary in the host country. With such up-front assurance, the host country resolution authority could be reasonably confident that the home country resolution authority would have an incentive to take the interests of the host country into account in formulating resolution plans for the group as a whole.

Two further matters require consideration. The first is what might be called a self-denying ordinance, namely a limitation on the ability of the host-country resolution authority to seize or sell the host-country subsidiary to a third party without the approval of the parent holding company or home country authorities, if the home country puts the home country bank and/or parent holding company into resolution. Without such a constraint on the host country resolution authority, the host country authority could potentially sell the (healthy) host country subsidiary to a third party for a nominal amount. This would

24 The assumption that the parent holding company wholly owns the bank subsidiary also simplifies matters. It abstracts from any rights that minority shareholders may have.

25 For smaller subsidiaries that are non-material to the group and non-systemic to the host country authority the host country authority may be satisfied with a parental guarantee, particularly if this is a legally binding, first demand guarantee, where the failure to perform would constitute an event of default for the parent holding company. However, the home country authority may be uncomfortable with the parent holding company’s giving such a guarantee and/or seek to insert clauses in domestic statute and/or regulations that would empower the home country resolution authority to suspend such guarantees if the banking group went into resolution.
cause significant additional losses to the parent (its investment in the common equity of the host country subsidiary bank would have to be written off) and additional losses to the holders of parent company obligations subject to bail-in.

Certainly, such a self-denying ordinance will be easier for host country authorities to give, if the home country authorities make some provision for a global liquidity facility to be provided to the group in resolution.\textsuperscript{26} Without such a global facility, there is a risk that the entry of the domestic bank subsidiary into resolution could cause the host country subsidiary bank to experience liquidity pressures sufficiently great to cause it to fail to meet threshold conditions in the host country. (That would allow the host country authorities to trigger resolution of the bank subsidiary in the host country.)

With such a global liquidity facility and with the up-front issuance of bail-in instruments to the parent holding company, the way should stand clear for the home country resolution authority to run what amounts to a single global resolution process. This is the solution most likely to make the bank resolvable, or safe to fail.

\textbf{The road to resolution}

If the above accurately portrays what would be required to make banks safe to fail, what steps need to be taken by policymakers and by banks to reach resolvability so that banks will be safe to fail? Three steps

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|c|}
\hline
\textbf{Assets} & \textbf{Prior to intervention} & \textbf{At intervention} & \textbf{Bail-in at parent only} & \textbf{After bail-in at parent and bank} \\
\hline
Marketable securities & 100 & 100 & 100 & 100 \\
Senior debt at bank subsidiary & 50 & 50 & 50 & 0 \\
Subordinated debt at bank subsidiary & 25 & 25 & 25 & 0 \\
Preferred stock at bank subsidiary & 25 & 25 & 25 & 0 \\
Common equity in bank subsidiary & 100 & 0 & 0 & 100 \\
\hline
Total & 300 & 200 & 200 & 200 \\
\hline
\textbf{Liabilities} & & & & \\
Senior debt & 100 & 100 & 0 & 0 \\
Subordinated debt & 25 & 25 & 0 & 0 \\
Preferred stock & 25 & 25 & 0 & 0 \\
Common equity & 150 & 50 & 0 & 0 \\
\hline
Total & 300 & 200 & 200 & 200 \\
\hline
\end{tabular}
\caption{Operation of bail-in with parent holding company structure: bail-in at parent must be accompanied by bail-in at the bank subsidiary}
\end{table}

\textsuperscript{26} The form for such a facility might be as follows: each central bank would be responsible for extending credit to the bank headquartered in its jurisdiction, and such credit would be collateralized by a pledge of assets from that bank to the central bank in its jurisdiction. Should the bank in question be unable to repay its obligation to its central bank and should the liquidation of the collateral be insufficient to repay the obligation in full, the home country resolution authority would make up the shortfall, and it would in turn recoup any loss that it suffered through a levy on the industry and/or recourse to the home country resolution fund.
stand out. First, authorities need to finish the reform of resolution regimes. Second, banks need to change their funding arrangements to accommodate bail-in. Third, financial market infrastructures need to take steps to coordinate their own recovery and resolution planning with that of their principal participants.

To complete the reform of resolution regimes “authorities” need above all to:

- Create the legal basis for bail-in at both the parent holding company and the operating bank subsidiary levels. Here, the enactment of the proposed E.U. Bank Recovery and Resolution Directive (BRR (2013)) would represent a critical step forward (Huertas and Nieto (2013)).
- Require that banks maintain a minimum amount of instruments subject to mandatory bail-in. This should be sufficient to recapitalize the bank, even if the bank’s common equity Tier-I capital is wiped out. These instruments subject to mandatory bail-in should be subordinated to customer obligations, such as deposits, on a statutory and contractual basis.
- Arrange adequate facilities for the provision of liquidity to the bank-in-resolution.
- Set out the basis on which home and host countries will cooperate with one another. As outlined above, a single point of entry, global approach to resolution can make banks resolvable. But such a global approach can only work, if (i) the home country is willing and able to take on the direction and leadership of a global resolution process, and (ii) the host countries are willing to accept the leadership of the home country and refrain from unilateral action to initiate and/or conduct a separate resolution process for the banking group’s subsidiaries or branches in the host country.

Bail-in holds the key to resolution, and “banks” to be resolved under the Single Point of Entry approach will need to rearrange their funding arrangements to accommodate immediate bail-in at both the parent holding company (if they are so organized) and at the level of the operating bank. This involves:

- Establishing a target funding model with the requisite amount of instruments subject to mandatory bail-in in issue to third-party investors. Note that such instruments will include non-core Tier-I and Tier-II capital instruments. These are likely to form the base of any funding subject to immediate bail-in, as Basel III requires such instruments to be subject to write-down or conversion at the point of non-viability, if they are to continue to qualify as capital. Senior debt subject to mandatory bail-in should be senior to non-core Tier-I and Tier-II capital, but subordinated to customer obligations, such as deposits, as a matter of contract and, ideally statute. As noted above, debt obligations of parent holding companies are structurally subordinated to obligations of the operating bank subsidiaries, and it should be feasible for operating bank subsidiaries to issue instruments subject to mandatory bail-in to their parent holding companies, that is contractually subordinated to senior debt issued to third parties, as well as contractually subordinated to deposits and other customer obligations.
- Eliminating the entry into resolution as an event of default in instruments (such as deposits and derivatives) that are not subject to mandatory bail-in and are senior to instruments subject to mandatory bail-in. In particular, revisions will need to be made to netting contracts including the standard ISDA agreement, and to repurchase agreements. Such contracts should not be subject to acceleration, and counterparties should have no right to close out or sell collateral pledged by the bank-in-resolution against such contracts, unless the bank-in-resolution fails to make payments as due.

---

27 For a further discussion of the importance of international cooperation see IIF (2012).
28 For a discussion of banks under the Multiple Point of Entry approach see Annex A.
Eliminating cross-guarantees or other forms of support (such as repurchase commitments and/or liquidity backstops) from the operating bank subsidiaries to the obligations of the parent holding company. Default on such parent company obligations should not trigger payments from the operating bank subsidiary, either to the parent holding company or third-party investors.

Disclosing to investors and counterparties whether the instrument in which they have invested is subject to mandatory bail-in, and where they stand in the queue to receive payments, should the bank enter resolution. To this end, the banking organization may find it helpful to conduct and keep up-to-date what might be called an entity priority analysis. This documents the order in which an investor has claims on the cash flows from specific assets (in the case where the obligation is secured) as well as directly or indirectly from entities within the group, in the event that the immediate obligor fails to pay.

In addition, banks will need to monitor and make available to central banks (and possibly private investors) information concerning what might be called a “collateral budget.” This relates to:

- Uses, or the encumbrance that the banking group has granted to creditors (assets pledged by the bank to creditors, noting whether such assets are owned outright or borrowed (and re-hypothecated by the bank to the lender)). Such information should include assurance that, if the borrowing bank repays the obligation the borrowing bank can rapidly and smoothly regain possession of the collateral previously pledged to the lender. Such information should also include estimates of the amount of additional collateral that the borrowing bank might be required to post under different scenarios, including without limitation deterioration in general market conditions and/or in the credit rating of the borrowing bank.

- Sources, or the amount of unencumbered assets that the banking group retains, the legal vehicles in which such assets are held, the eligibility of such assets for discount at central bank(s), either under ordinary discount window facilities or under emergency liquidity assistance, whether such assets are pre-positioned with the central bank, and some estimate of the terms (e.g., haircut) on which central bank and/or private lenders might be willing to provide funds. Note that the ability to repossess collateral from one lender (e.g., a repo counterparty seeking repayment) in order to provide it to another (e.g., a central bank providing a liquidity facility) is likely to be especially important in assuring that the bank-in-resolution can gain access to sufficient funding liquidity at the close of the resolution weekend/opening of business on Monday.

Finally, FMIs have to take steps to integrate their own recovery [CPSS-IOSCO (2013)] and resolution [FSB (2013b)] planning with that of the G-SIFIs who are their principal members (see Figure 7). In particular, FMIs should take steps to assure that:

- The entry of a participant into resolution does not automatically exclude the bank-in-resolution from access to the FMI. If the resolution process succeeds in stabilizing the bank so that the bank-in-resolution can continue operation, it should retain access to the FMI.29

- There is a clear understanding on how the FMI will handle “in-flight” transactions if a participant to the FMI enters resolution, and there should be a bias toward completing such transactions. Indeed, that is the purpose of the margin requirements and default funds that FMIs require participants to post.

29 However, the terms on which the bank-in-resolution transacts with the FMI may differ from the terms on which the bank was able to transact prior to its entry into resolution. In particular, it is unlikely that the FMI would be willing to grant credit (even on an intraday basis) to the bank-in-resolution.
Participants’ margin requirements and default funds at FMIs are liquid, i.e., they are either in cash or in instruments readily convertible into cash (even during the weekend).

Participants’ obligations to replenish a FMI’s default fund are limited and capable of being fulfilled rapidly (even during the weekend).

The FMI itself has sufficient capital to bear the loss that might arise as a result of the default of at least one of its largest participants.

In addition, both the authorities and FMIs will need to take steps to create a framework for resolution of an FMI, should the recovery measures outlined above prove insufficient. This would include designating a resolution authority for each FMI and empowering the resolution authority to take measures, such as the transfer of the FMI’s business to an alternative provider or to a bridge institution, and/or the imposition of a hair-cut on the initial margin provided by the surviving members, to allow the FMI to continue operations or conduct an orderly wind-down (Tucker (2013)).

Conclusion

These steps together constitute a massive agenda. But it is an agenda that is possible for authorities, banks and financial market infrastructures to achieve. Indeed, important steps have already been taken toward this end.

There is a way to make banks safe to fail, so that they can be resolved without taxpayer solvency support and without significant disruption to the economy. And, this can be done without compromising the contribution that global banks can make to growth in the global economy. What is required is cooperation among the authorities, realignment of funding at banks to accommodate bail-in and reform at FMIs. This constitutes a single, global approach to resolution under the direction of the home country resolution authority.

In contrast, national, “go-it-alone” approaches to resolution will impose significant costs and reduce the capability of global banks to contribute to global growth (see Annex A). More significantly, it is likely that the pursuit of financial stability in one jurisdiction would cause instability elsewhere. If one does not consider the coordination that a single, global approach to resolution would also require, it is difficult to see how the multiple point of entry approach could succeed in making banks resolvable. As Bill Dudley (2013), President of the Federal Reserve Bank of New York, recently remarked, “We can do better through international cooperation and coordination both on macro policy and on regulation and supervision, rather than trying to ‘go it alone’.”

In summary, “too big to fail” is not too tough to solve. Now is the time to finish the job.

References


Huertas, T. F., forthcoming, “Restructuring the bank-in-resolution”


Annex A: A note on resolution via multiple point of entry

The above discussion suggests that banks can be made resolvable via what amounts to a pre-pack reorganization — a single point of entry, global approach to resolution under the firm direction of the home country resolution authority accompanied by a global liquidity facility arranged by the home country central bank/resolution authority.

A multiple points of entry approach is also feasible, at least for banking groups organized as “archipelagos” or collections of independent, separately capitalized and separately funded bank subsidiaries owned by a common parent holding company. Each of these separate bank subsidiaries would be resolved (if that particular bank subsidiary reached the point of non-viability [failed to meet threshold conditions]) in the jurisdiction in which the subsidiary was headquartered without reference to the parent holding company or affiliates in other jurisdictions. Each such resolution process should proceed along the lines outlined above for the case where the bank is the parent entity.

In general, the caveats outlined in the main text also apply to the multiple points of entry approach. In particular, if an operating bank subsidiary has branches in a foreign country, the resolution of that bank can be seriously compromised if the host country takes a territorial approach to resolution and attempts to resolve the foreign branch of the bank separately from the rest of the bank. Indeed, if the host country were to take such a step without prior consultation or warning to the home country authorities (supervisor, central bank, and resolution authority), such a step would practically assure financial instability in the home country and in the other jurisdictions in which the bank conducted a significant amount of business and/or played a significant role in financial markets. Multiple points of entry should not mean two uncoordinated attempts to resolve the same legal vehicle at the same time.

Similarly, chaos can result if resolution authorities have, and take the option to implement what amounts to a “cross-resolution” clause [entry of a subsidiary (the failed affiliate) into resolution in one country entities any other resolution authority elsewhere in the world to put into resolution affiliates in its jurisdiction]. As outlined in the main text, such powers could result in the host country authority’s selling a healthy affiliate in the host country to a third party for a nominal sum to the detriment of the creditors of the parent holding company and to the detriment of the creditors of the other operating subsidiaries of the group (who would be denied access to the capital resources that the parent holding company might otherwise have had available to recapitalize such subsidiaries).30

Consequently, for the MPE approach to work some limits will need to be placed on the ability of host countries to take unilateral action. Without some type of coordination and without some type of limitation on unilateral action, the multiple point of entry approach runs the risk of creating, not a race for the courthouse (for there is no international court house to go to) but simply a race for assets, where speed, opportunity and might make right.

At a minimum, the multiple point of entry approach implies that host countries will agree to be blind — when it comes to resolution — to the fact that a bank in its country is owned by a group headquartered in another country. Concretely, it implies that: (1) the host country central bank is willing to extend liquidity facilities to a subsidiary bank owned by a foreign banking group on the same terms and conditions as it would employ for a domestic bank; (2) the host country authorities are willing to stay their hand until such point as the subsidiary in the host country reaches the point of non-viability (fails to meet threshold conditions in the host country); and (3) the home and host country authorities are willing to allow a group to simply walk away from a subsidiary in the host country, without some type of coordination and without some type of limitation on unilateral action.

It is not clear whether the authorities have given any such assurances. If anything, authorities in many countries, notably the U.S.,31 seem to be going out of their way to emphasize that they retain the power of discretion to act as local law empowers them to do to protect local creditors regardless of the impact that such actions may have on the rest of the group or on international financial stability.

Nor is it clear whether authorities are willing to follow through to the logical consequence of a multiple point of entry approach: the

---

30 Such behavior would be close to or even tantamount to expropriation or nationalization without compensation and is arguably a risk that banking groups already run in their normal course of business. However, in a resolution situation the barriers to host countries taking such a step are lower as is the likelihood that a court would rule against the host country supervisor.
31 For example, the recent Federal Reserve Board (2012) proposal for remediation and resolution of foreign banking organizations in the U.S. makes no provision for the Fed to consult or coordinate with the home country supervisor.
removal of capital requirements on the parent holding company. Under a multiple point of entry approach, the banking group is expected to put in up front all the strength required to keep each subsidiary bank well capitalized and well funded. Each subsidiary is required to be self-sufficient. Should a subsidiary fail to remain so, the supervisor of that subsidiary can put the subsidiary into resolution. That is the supervisory remedy, not a call on the parent to provide more capital (presumably the parent would have injected such capital already, if it had the capital available and if it considered it in its commercial interest to make such an injection).

Removal of parent company capital requirements would underline that under a multiple point of entry approach the focus of supervisors is exclusively on the operating bank subsidiaries, not the group as a whole. It would also underline to the market that there is no support for the group at the group level. And the market, rather than the regulator, would determine the most efficient capital structure (balance of equity and debt) for the parent holding company. This could present an effective means of marrying a very high degree of protection at the bank level (thereby assuring the safety of deposits) with the freedom for financial firms to manage their overall cost of capital in an efficient manner.
Market risk of real estate: Using indirect data to understand direct risks

Felix Schlumpf
Head of Strategic Asset Allocation, Investment Management, Zurich Insurance Company Ltd.

Genene Tessera
Senior Quantitative Analyst, Investment Management, Zurich Insurance Company Ltd.

Catalina Martínez
Research Assistant, Investment Management, Zurich Insurance Company Ltd.

Abstract
Even if the market capitalization of direct real estate is comparable to that of equities and fixed income, the data on direct real estate is very poor. It is, therefore, difficult to estimate the market risk of this important asset class. Moreover, risk systems from most vendors cover equities and fixed income, but do not cover direct real estate. We propose a simple methodology that uses widely available data on indirect real estate to estimate the market risk of direct real estate. In particular, we use data on Real Estate Investment Trusts (REITs) returns, determine their factor exposures to other asset classes and deleverage these exposures according to REITs' balance sheets. We show that direct real estate can be considered as a portfolio of equities, fixed income and credit combined with idiosyncratic risk. We find that the existing direct indices understate the risk of the real estate market. In addition, with our methodology, the correlations to other asset classes become materially different and higher.
Market risk of real estate: Using indirect data to understand direct risks

Introduction
Direct real estate is the largest asset class without readily available prices. It is difficult to observe direct real estate data because transactions have low frequency and are often private. It is also difficult to interpret the few data that can be found because properties are heterogeneous and often difficult to compare. There exist many methodologies that try to solve these problems, such as the appraisal-based indices, the transaction-based indices and the hedonic approach. Nevertheless, all these methodologies present important challenges and are difficult to use for practical applications. These data issues make the market risk of real estate poorly understood. This paper proposes a simple model that allows using widely available data on indirect real estate to achieve a better understanding of the direct real estate market risk.

We propose to use widely available data on real estate investment trusts (REITs) to acquire a more accurate knowledge of real estate market risk. REITs are closed-end funds that invest in direct real estate and are financed with equity and debt. REITs are openly traded in the market and provide abundant and frequent data. They are more transparent than direct real estate indices, and also relatively more price-efficient. Indeed, REITs are required to prepare standardized public financial statements and accurate performance measurements, which make it easier to re-price assets on a timely fashion.

There is an abundant literature supporting a close link between REITs and direct real estate. Our paper proposes a simple model to apply these widely accepted findings to the practical problem of real estate risk management and portfolio allocation. Building on the work documented in Schlumpf and Tessera (2010), our model is based on the simple observation that a REIT is a real estate company that can be described as a portfolio of direct property investments financed by mortgages or bonds. With this observation, we are able to build a bridge between direct and indirect real estate that we use to better understand the market risk of direct real estate.

Using REIT balance sheet data, we are able to translate indirect real estate risk factor decompositions into direct real estate risk factor decompositions. This approach enables us to represent direct real estate as a portfolio of equities, fixed income and credit combined with an idiosyncratic risk. This is useful for practical risk and asset management decisions given that risk systems from most market vendors cover equities and fixed income as asset classes, but not direct real estate. Our methodology offers a simple, transparent and accurate description of the real estate asset class and of its market risk.

We find that the estimations of real estate volatility that can be derived from the existing methodologies considerably understate the risk of the real estate asset class. Indeed, taking into account the case of the U.K., which has good availability of both direct and indirect real estate indices, we are able to compare the volatility resulting from our model to that of direct real estate indices, both appraisal-based and transaction-based. We find that the volatility that results from our model is almost three times bigger than that of direct real estate appraisal-based indices, and two times bigger than transaction-based ones.

Moreover, we find that the correlations between real estate and risky assets resulting from our methodology are substantially higher than those resulting from the traditional approaches. This has important implications for risk management and asset allocation. Our results are particularly important in the light of the role that real estate played in the recent financial crisis. Indeed, in our model the price movements that took place in the recent crisis correspond to less than a two standard error move.

Recent literature
The majority of benchmarks for direct investments in real estate are appraisal-based. They are usually smoothed and tend to lag the returns in the property markets. Moreover, appraisal-based indices understate the volatility of real estate investments and their correlation to other asset classes, making them inadequate to describe market risk and to guide portfolio diversification. Moreover, the history of most direct indices is quite short and the data frequency is low.

A more accurate way for constructing direct real estate indices would be to use the hedonic method. Nevertheless, this methodology requires remarkable datasets containing information such as physical characteristics, neighborhood, area, distance to the city center, etc. Since this kind of data is often not available, the hedonic method is difficult to apply in practice.
A vast amount of research is devoted to providing solutions to the problems presented by direct appraisal-based indices. In particular, several authors have attempted to provide techniques to unsmooth these indices, which tend to be particularly autocorrelated because appraisers usually keep their estimations fairly constant over time. Indeed, even if there exists an established standardized appraisal methodologies, appraisal-based indices tend to lag the market and understate volatility because appraisers usually smooth their valuations over time. Suryanarayanan and Stefek (2010) argue that the lag between appraisal and market values is between three months, when transaction volumes are high, and one year, when they are low.

General unsmoothing techniques apply a statistical filter to the appraisal-based returns to remove the autocorrelation in the series. Nevertheless, these procedures present several problems. In particular, Bond et al. (2006) argue that autocorrelation filters tend to overstate the smoothing that takes place at the individual property level. They argue that while idiosyncratic shocks are cancelled out when individual property level appraisals are aggregated, common shocks are not, and they tend to be highly persistent. This persistency is due to the illiquidity of the real estate market that impedes arbitrageurs to exploit shocks. Moreover, common factors may be related to macroeconomic fundamentals that tend to evolve slowly over time. The bias of the unsmoothing techniques arises because autocorrelation filters are applied to the aggregate indices and not to the individual ones. As a result, these procedures overstate the smoothing and may result in misleading information.

Transaction-based indices are often proposed as an alternative to appraisal-based indices. Nevertheless, transaction based indices may contain a lot of noise, as the baseline information often presents outliers. Moreover, transactions have low frequency, are often private, and it is unclear if the observed ones can be considered representative of the market.

Some authors have developed techniques to address the fact that real estate property transactions do not occur very often. For example, Sheharyar and Geltner (2012) propose a two-stage frequency conversion methodology to increase the frequency of direct real estate data. Nevertheless, this approach tends to lag behind the estimation. Other econometric models aimed at solving the problems posed by the transaction approach have been proposed by Fisher et al. (2007) in the MIT Centre for Real Estate for the case of the U.S., and Devaney and Martinez Diaz (2010) for the case of the U.K.

A growing literature has departed from the transaction-based approach to consider how REIT data can serve as a proxy for the direct real estate market. For example, Suryanarayanan (2011) describes the methodology implemented by MSCI Barra, which by unsmoothing private index returns finds a close relationship between direct and indirect real estate investments. The correlations (between REITs and direct real estate) in the long run model for the U.S. are between 40% and 60%, while for the U.K. they are between 50% and 65%. Similarly, Suryanarayanan and Stefek (2010) find a strong link between annual returns to public and private real estate in the U.K. and U.S. markets. These authors correct for appraisal smoothing and take into account the lead-lag relationship between the two asset returns. They find that the relationship between direct and indirect real estate is stronger as the time horizon increases.

Sebastian and Schätz (2009) include economic fundamentals in the analysis of the relationship between direct and indirect real estate indices. Including macroeconomic variables is very important because it helps to control the effects that these variables may have on the comovement between REITs and direct real estate. These authors focus on the overall real estate market and do not consider the specificities of different sectors. Pavlov and Wachter (2011), on the other hand, take into account economic fundamentals and different real estate sectors, but not the influence of lead-lag relations or the long-run behavior of the assets.

Hösi and Oikarinen (2012) propose the most rigorous and complete study to our knowledge. Using quarterly data for U.S., U.K. and Australia, and a Vector Error Correction Model (VECM), these authors derive impulse response functions of the asset returns, which allow estimating the reaction patterns that REITs and direct real estate returns have to shocks in economic fundamentals and in the asset returns themselves. The variance decompositions proposed show that, in the long term, the forecast error variance of direct real estate returns is explained mostly by shocks to REITs. This implies that securitized and direct real estate returns are driven by a common “real estate factor,” given that the fundamental asset is essentially the same in both
Market risk of real estate: Using indirect data to understand direct risks

markets. Consequently, in the long run, direct and indirect real estate are closely related.

Moreover, Hösli and Oikarinen (2012) find that in the long run shocks in the stock market do not have an influence in the variance of REITs’ shocks. REITs, therefore, provide the same diversification benefits that direct real estate does, making direct and indirect real estate good substitutes in the long term. Furthermore, the authors find that even if REITs shocks help in predicting the variance of direct real estate shocks in the long run, the opposite does not hold.

This suggests that real estate shocks take place first in the REIT market, and only then the direct market adjusts. Consequently, using REIT data may be more useful for predicting crisis and recovery periods. This is similar to the findings of Cotter and Roll (2011), who argue that REITs can provide useful information for predicting collapses in the real estate market, such as the one that was closely associated with the recent financial crisis.

We rely on this evidence to use REIT data for performing an accurate risk factor analysis of the real estate asset class. Building on the model proposed by Schlumpf and Tessera (2010), we focus in this paper mostly on the risk management implications of using REIT data to improve our understanding of the real estate asset class. We do so by estimating the risk factor exposure of REIT returns and deleveraging it to obtain the risk factor exposure of direct real estate.

A similar approach to ours is developed by Horrigan et al. (2009), who construct property market returns indices by using data on REITs’ returns, bonds and property holdings. According to these authors, the beneficial characteristics of REITs allow us to detect the price movements of the underlying property markets on a daily basis without having to observe property transactions, which occur with low frequency. To achieve this, they combine information on REIT asset holdings and stock returns using the weighted average cost of capital (WACC) accounting identity. The resulting so-called “pure play” indices can be used to make targeted investments in the commercial real estate market, while keeping the advantages of REITs, such as liquidity, transparency and pricing efficiency.

Our approach is different from that of Horrigan et al. (2009) because we focus on the practical applications of using the WACC identity and REIT data for achieving a better risk analysis of the real estate asset class, and therefore producing better risk estimates. Indeed, while Horrigan et al. (2009) want to build new and more accurate data, as well as to propose the grounds for new investment instruments, our approach is aimed at deriving real estate risk factor exposures in a more accurate way, so as to provide a better understanding of the real estate risk that is useful in empirical applications.

Some studies have focused on the insights regarding the risk of real estate that can be derived by comparing direct and indirect real estate risks. In particular, Cotter and Roll (2011) propose an exhaustive comparison of the returns, risk, and distributional characteristics of REITs and residential real estate indexes, using S&P Case Shiller monthly data. They find that smoothed residential real estate series tend to have lower volatilities than REITs. According to these authors, direct real estate volatility is around 1/5 of REITs volatility.

Nevertheless, Cotter and Roll (2011) point to the problem that it is not clear whether the smoothed residential real estate series that they used really reflected the direct real estate prices. This problem is taken into account by Devaney and Martinez Diaz (2010), who argue that transaction-based indices help to find a more precise estimation of real estate market risk, as compared to appraisal-based indices. We propose to go one step further and take into account the widely available data on REITs to produce simpler, and more accurate and transparent estimations of the direct real estate market risk.

Our approach is of particular interest given that, after the protagonist role that real estate had in the recent financial crisis, a growing attention has been given to extreme risks in real estate markets. Most of the interest in the extreme risks in real estate has been reflected in papers that study the volatility of REITs but not how this volatility can help us better understand the volatility of the direct real estate markets.

For example, Lu et al. (2012) examine the diversification effects across international REITs and find that U.S. REITs contribute to most of the risks within international REITs. Moreover, Zhou and Anderson (2012) explore extreme risks in REITs and find
that they tend to be higher than those of stocks. Other papers providing tools to better understand the risks of REITs include Springer and Cheng (2006), Pagliari et al. (2005) and Riddiough et al. (2005). We contribute to this literature by disentangling the direct real estate risk from the REIT risk, which could have potential implications for practitioners who wish to improve their understanding and management of real estate risk.

**Model**

We propose a simple model to better understand the risk of the real estate asset class, building on the work documented in Schlumpf and Tessera (2010). This model is based on two pillars: 1) a risk factor analysis of REIT returns, and 2) a simple accounting identity that relates REIT returns with direct real estate returns. Combining the first and the second pillars, we are able to perform a risk factor analysis of direct real estate that is useful for practical risk management considerations. This section explains in detail how our simple model works and how it can be easily applied to practical problems.

Our purpose is to estimate the risk factor exposures of the direct real estate. We would, therefore, like to produce estimates for the beta exposures of following equation:

\[ r_{\text{direct}} = \alpha_{\text{direct}} + \sum \beta_{\text{direct}i} f_i + \varepsilon_{\text{direct}} \]  \hspace{1cm} (1)

Where \( r_{\text{direct}} \) represents the returns on direct real estate, \( \alpha_{\text{direct}} \) represents the alpha returns on direct real estate, \( \beta_{\text{direct}i} \) represents the beta exposures of direct real estate to factor \( i \), with \( f_i \) being the factor considered and \( \varepsilon_{\text{direct}} \) representing the shocks to direct real estate returns.

Since direct real estate data presents several problems, as we argued in the previous section, the estimates for \( \beta_{\text{direct}i} \), that we can achieve using this data can be strongly biased, leading to wrong risk predictions. Nevertheless, given that REITs are listed on major exchanges and traded continually, there is enough availability of indirect real estate data that allows us to produce consistent and efficient estimates for the following equation:

\[ r_{\text{indirect}} = \alpha_{\text{indirect}} + \sum \beta_{\text{indirect}i} f_i + \varepsilon_{\text{indirect}} \]  \hspace{1cm} (2)

Where \( r_{\text{indirect}} \) represents the returns on indirect real estate, \( \alpha_{\text{indirect}} \) represents the alpha returns on indirect real estate, \( \beta_{\text{indirect}i} \) represents the beta exposures of indirect real estate to factor \( i \), with \( f_i \) being the factor considered and \( \varepsilon_{\text{indirect}} \) representing the shocks to indirect real estate returns.

We estimate \( \beta_{\text{indirect}i} \) using a simple OLS approach in which the main factors under consideration are equities, fixed income, and credit. We can then go from (1) to (2) by considering the characteristics of REITs’ balance sheets. Indeed, REITs are portfolios of direct property investments financed by mortgages or bonds.

Taking this information into account, we can show that once we have modeled the behavior of indirect real estate using time series data on REITs’ returns (equity in the graph), we can approximate the behavior of direct real estate (property holdings in the graph) by taking into account the behavior of the debt of REIT companies. Indeed, since the returns on direct real estate is the sum of indirect real estate returns and debt returns weighted by the amount of leverage, i.e.,

\[ r_{\text{direct}} = r_{\text{indirect}} (1 \cdot \text{lev}) + r_{\text{debt}} (\text{lev}) \]  \hspace{1cm} (3)

we can recursively use our risk factor model for REITs of equation (2) to show that the following relationships hold:

\[ \beta_{\text{direct}i} = \beta_{\text{indirect}i} (1 \cdot \text{lev}) + \beta_{\text{debt}} (\text{lev}) \]  \hspace{1cm} (4)

\[ \text{dur}_{\text{direct}} = \text{dur}_{\text{indirect}} (1 \cdot \text{lev}) + \text{dur}_{\text{debt}} (\text{lev}) \]  \hspace{1cm} (5)

\[ \varepsilon_{\text{direct}} = \varepsilon_{\text{indirect}} (1 \cdot \text{lev}) + \varepsilon_{\text{debt}} (\text{lev}) \]  \hspace{1cm} (6)

Consequently, with data on each market’s leverage and duration we can calculate \( \beta_{\text{indirect}i} \), i.e., the risk factor exposures of direct real estate. With unbiased estimates of \( \beta_{\text{indirect}i} \) we are then be able to express direct real estate in terms of \( f_i \), build predicted time series, and, in this way, have a better understanding of direct real estate risk.

**Data**

We analyze six local indirect real estate indices for the markets of Australia, Canada, Switzerland, E.U., U.K. and U.S. Each REIT index (REITAULC, G250CALC, DBCHREF, EPEU, G250GBLC, and REITUSLC, respectively) represents the returns of a set of public real estate companies belonging to each particular market. In order to cover a full economic cycle, we consider the timeframe from January...
Market risk of real estate: Using indirect data to understand direct risks

1994 (February 1995 for Switzerland) to July 2012. Appendix A shows the details of the time series used in each case and Table 1 shows their summary statistics.

The leverage figures that we consider, shown in Table 2, have been calculated bottom-up by looking up all the annual reports of the REITs in the indices used. It has been found that the funding duration is quite constant over time at the index level.

For each of the REIT indices that we consider, we propose relevant factors to estimate our OLS regression. Each model consists of six factors (See further data details in Appendix A): local equity index, world equity index (MSCI world), local short term fixed income (1-3 year maturity), local medium term fixed income (5-7 year maturity), local long term fixed income (10+ year maturity) and credit.

In order to compare the risk of the real estate asset class implied by our model with that of direct real estate indices, we take into account data of direct real estate in the UK, using the IPD index. We choose this market because it presents the longest time series available and also has a relatively good periodicity, as it is available on a monthly basis.

### Results

#### Factor risk modeling

The results of the OLS regressions on indirect real estate indices are summarized in Table 3 and further details can be found in Appendix B. The different exposures across countries can partially be explained by the different sector splits in the REIT indices used for our analysis. Table 3 also reports the REIT duration that can be derived using the exposure of each series to the relevant fixed income time series and the respective EFFAS duration.

As in Schlumpf and Tessera (2010), we tested for multicollinearity in the regression design matrix (i.e., the independent variables), using the variance inflation factor (VIF). None of the estimated regression coefficients showed multicollinearity.

---

1 We are grateful to Stefan Weber, Ramona Reinert, and Philipp Langenegger for providing us with this data.

2 The VIF is derived from the variance of the estimated regression coefficients and tells us how much the variance of the estimated coefficient is “inflated” by the existence of correlation among the predictor variables in the model.

---

### Tables

#### Table 1: Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>AU</th>
<th>CA</th>
<th>CH</th>
<th>EU</th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.008</td>
<td>0.007</td>
<td>0.005</td>
<td>0.003</td>
<td>0.004</td>
<td>0.009</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.045</td>
<td>0.051</td>
<td>0.022</td>
<td>0.045</td>
<td>0.061</td>
<td>0.063</td>
</tr>
<tr>
<td>Median</td>
<td>0.008</td>
<td>0.013</td>
<td>0.005</td>
<td>0.006</td>
<td>0.007</td>
<td>0.016</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>6.727</td>
<td>4.211</td>
<td>0.838</td>
<td>5.410</td>
<td>3.040</td>
<td>10.982</td>
</tr>
<tr>
<td>Skewness</td>
<td>-1.446</td>
<td>-0.759</td>
<td>-0.048</td>
<td>-1.135</td>
<td>-0.662</td>
<td>-1.749</td>
</tr>
<tr>
<td>Minimum</td>
<td>-0.259</td>
<td>-0.238</td>
<td>-0.068</td>
<td>-0.258</td>
<td>-0.265</td>
<td>-0.392</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.146</td>
<td>0.212</td>
<td>0.083</td>
<td>0.146</td>
<td>0.222</td>
<td>0.268</td>
</tr>
<tr>
<td>Observations</td>
<td>222</td>
<td>222</td>
<td>210</td>
<td>210</td>
<td>222</td>
<td>222</td>
</tr>
</tbody>
</table>

#### Table 2: Leverage and duration

<table>
<thead>
<tr>
<th>Market</th>
<th>Leverage (%)</th>
<th>Debt duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUD</td>
<td>28.6</td>
<td>1.5</td>
</tr>
<tr>
<td>CAD</td>
<td>64.6</td>
<td>5.0</td>
</tr>
<tr>
<td>CHF</td>
<td>13.4</td>
<td>3.1</td>
</tr>
<tr>
<td>EUR</td>
<td>46.7</td>
<td>3.5</td>
</tr>
<tr>
<td>GBP</td>
<td>44.3</td>
<td>7.9</td>
</tr>
<tr>
<td>USD</td>
<td>52.3</td>
<td>3.8</td>
</tr>
</tbody>
</table>

#### Table 3: Regression summary — indirect real estate modeling

<table>
<thead>
<tr>
<th></th>
<th>AU</th>
<th>CA</th>
<th>CH</th>
<th>EU</th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local equity</td>
<td>0.63a</td>
<td>0.56a</td>
<td>0.43a</td>
<td>0.81a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global equity</td>
<td>0.16a</td>
<td>0.51a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FL 1-3</td>
<td>1.80a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FL 5-7</td>
<td>0.73a</td>
<td>0.68a</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FL 10-30</td>
<td>0.32a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Credit</td>
<td>1.15b</td>
<td>2.08a</td>
<td>4.51a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual volatility</td>
<td>11.9%</td>
<td>14.8%</td>
<td>6.9%</td>
<td>11.9%</td>
<td>17.2%</td>
<td>16.6%</td>
</tr>
<tr>
<td>Duration</td>
<td>3.8</td>
<td>3.7</td>
<td>3.6</td>
<td>1.1</td>
<td>5.1</td>
<td>3.2</td>
</tr>
</tbody>
</table>

#### Table 4: Regression summary — direct real estate modeling

<table>
<thead>
<tr>
<th></th>
<th>AU</th>
<th>CA</th>
<th>CH</th>
<th>EU</th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local equity</td>
<td>45.0%</td>
<td>20.0%</td>
<td>23.0%</td>
<td>45.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global equity</td>
<td>13.6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FL 1-5</td>
<td>6.0%</td>
<td>7.0%</td>
<td>56.0%</td>
<td>74.0%</td>
<td>39.0%</td>
<td></td>
</tr>
<tr>
<td>FL 5-10</td>
<td>49.0%</td>
<td>73.0%</td>
<td>30.4%</td>
<td>3.0%</td>
<td>26.0%</td>
<td>36.5%</td>
</tr>
<tr>
<td>FL 10-30</td>
<td>28.7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Credit</td>
<td>82.2%</td>
<td>111.0%</td>
<td>215.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual volatility</td>
<td>8.5%</td>
<td>5.2%</td>
<td>6.0%</td>
<td>6.4%</td>
<td>9.6%</td>
<td>7.9%</td>
</tr>
<tr>
<td>Duration</td>
<td>3.2</td>
<td>4.5</td>
<td>3.5</td>
<td>2.2</td>
<td>6.4</td>
<td>3.5</td>
</tr>
</tbody>
</table>
Using our figures for leverage and duration and applying our simple model, we achieve the estimation of the exposures of direct real estate returns summarized in Table 4. In particular, we use equation (4) for the beta coefficients of local equity, global equity and credit. For the residual volatility, we use equation (6). In addition, for fixed income we consider coefficients that would add up to 100% leaving credit as an overlay, and seek to distribute the fixed income exposures in order to comply with equation (5), while using the respective EFFAS durations.

The duration figures that we calculate for direct real estate are in line with the estimations of the recent literature. Indeed, several papers show that real estate has low interest rate sensitivity [Rauh and Rieder (2004); Iacoviello (2000); Sutton (2002); Tsatsaronis and Zhu (2004)]. This applies also for the Swiss market, as it is argued by Constantinescu (2009).

Market risk estimation
We would now like to use these results to achieve a better understanding of risk in the real estate markets. In this sense, we would like to be able to provide some explanation of extreme events such as the one depicted in Figure 1, and to provide insights about the implications on risk management.

We use the results of our direct real estate model reported in Table 4 and build a predicted series. This allows us to compare our model with the traditional approaches, as we can compare the behavior of this predicted series with that of traditional direct real estate series. We do this exercise using data from the U.K.

We find that the volatility of the predicted series for the studied period is 11.9%. This is considerably higher than the volatility of the simple IPD U.K. appraisal-based index for the same period, which is 4%. Consequently, with our modeled direct real estate time series we are able to make extreme events, such as the one depicted in Figure 1, not so extreme, as they would correspond to a 2 standard deviation move, and not to a 5 standard deviation move. Indeed, this kind of events could be expected under a fat tail distribution.

According to the model proposed by Devaney and Martinez Diaz (2010), the volatility of the transaction-based index that they propose, which is based on the same IPD data for the U.K. that we consider, is 6.3% for the period between Q1 2002 and Q2 2009. Even if this volatility is higher than the IPD appraisal-based data, it is almost half of our estimation, which implies that this methodology would still considerably underestimate risk. Moreover, if we consider the same time frame as these authors do, then our result is even higher and equal to 12.2%.

Furthermore, we also take into account the volatility that would result from the common practice of unsmoothing direct real estate data by considering an autocorrelation parameter. We find that this volatility will be much higher than the one predicted by our model. Indeed, for the period under study it would be 16.8%. This reflects the findings of Bond et al. (2006), who argue that the autocorrelation filter tends to overstate the smoothing taking place in direct real estate data.
These empirical results indicate that our approach proposes a middle-way solution, which does not suffer from the problems of the autocorrelation filter, but is still able to provide a higher volatility for the real estate market. Moreover, we find that both unfiltered and filtered direct real estate data present negative correlations with fixed income in the short- and long-run, while our model predicts positive correlations, as can be seen in Table 5. In this sense, our results are similar to those of Solvency II. Indeed, according to the fifth Quantitative Impact Study (QIS 5) for Solvency II, the volatility for equities is 15.5% and for real estate it is 9.7%, with a correlation of 0.75 between the two.

We also consider a representative portfolio that allocates 25% to real estate, 25% to equity, 30% to short-term fixed income and 20% to long-term fixed income, to see how these findings would affect overall portfolio risk. As can be seen in Table 6, we find that when we consider direct real estate IPD indices, the volatility of the portfolio is the lowest at 4.6%. Using the autocorrelation filter leads to a result of 5.9%. Finally, when we consider our modeled real estate series we have a higher portfolio risk, amounting to 6.7%.

The fourth row of Table 6 shows how our model can be implemented and used in vendors’ risk systems that cover equities and fixed income as an asset class, but do not cover direct real estate. Indeed, our model can easily be implemented in a simple two-step process: 1) In a portfolio of equities and fixed income, add to each of these asset classes the direct real estate exposures to those asset classes resulting from our model. 2) Add the residual risk of real estate as an idiosyncratic risk.

For our U.K. example, in the first step we do the following: to the 25% equity, the 30% of 1-5 fixed income and the 20% of 10+ fixed income, we add the 25% of real estate allocation multiplied by the exposures of the direct regression for U.K. This gives an allocation to EQ of 36.3%, to FI 5-10 of 36.5% and to FI 10+ of 27.2% and results in a portfolio volatility of 6.1%. By adding the residual volatility of 9.6% with a weight of 25% and zero correlation we get a higher portfolio volatility, close to the one of the direct calculation.

**Conclusion**

We show that direct real estate can be thought of as a portfolio of equities, fixed income and credit combined with idiosyncratic risk. The composition of this portfolio is derived by regressing REITs data and deleveraging the resulting asset exposures according to the leverage in the considered REITs. Following this simple model, the market risk for direct real estate is higher than that of appraisal or transaction-based data. Moreover, the correlations to risky assets are higher in our model than in the appraisal-based models. This is important for both risk management and asset allocation. Further studies could consider more countries and analyze the impact of the different real estate sectors.

### Table 6: Representative portfolios’ volatility comparisons

<table>
<thead>
<tr>
<th>EQ</th>
<th>FI 5-10</th>
<th>FI 10+</th>
<th>RE IPD</th>
<th>IPD filter</th>
<th>RE model</th>
<th>Volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>30%</td>
<td>20%</td>
<td>25%</td>
<td>25%</td>
<td>25%</td>
<td>6.1%</td>
</tr>
<tr>
<td>25%</td>
<td>30%</td>
<td>20%</td>
<td>25%</td>
<td>25%</td>
<td>25%</td>
<td>6.7%</td>
</tr>
<tr>
<td>36.3%</td>
<td>36.5%</td>
<td>27.2%</td>
<td></td>
<td></td>
<td></td>
<td>4.6%</td>
</tr>
</tbody>
</table>

**References**


Lu, C., Y. Tse, and M. Williams, 2012, “Returns transmission, value at risk, and diversification benefits in international REITs: evidence from the financial crisis,” Review of Qualitative Finance and Accounting, 40(2), 293-318

Morningstar, 2011, “Commercial real estate investment: REITs and private equity real estate funds,” September

Northfield, 2012, “REIT risk model”, Models and Analytics


Schlumpf, F., and G. Tessera, 2010, “Factor analysis for real estate;” Available at SSRN
Sebastian, S., and A. Schütz, 2009, “Real estate equities - real estate or equities?”, EPRA Research, International Real Estate Business School, Universität Regensburg
Suryanarayanan, R., 2011, “The Barra Private Real Estate Model (PRE1),” MSCI Model Insight, MSCI Research
Wilshire, 2012, “The role of REITs and listed real estate equities in target date fund asset allocations,” National Association of Real Estate Investment Trusts

Appendices
A. Data description
The following data series are used as the explanatory variables in the regressions: MSCI equity indices and local equity indices, EFFAS government bond indices for fixed income, and credit spread data based on CDS investment grade index (5 years) for credit.

<table>
<thead>
<tr>
<th>Fixed income</th>
<th>Equities</th>
<th>Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3 years XXG1TR index</td>
<td>MSCI World GDDUWI index</td>
<td>IBOXUMAE index</td>
</tr>
<tr>
<td>3-5 years XXG2TR index</td>
<td>S&amp;P500 SPI Index</td>
<td>EDGISP index before 2003</td>
</tr>
<tr>
<td>5-7 years XXG3TR index</td>
<td>FTSE 100 UKX index</td>
<td></td>
</tr>
<tr>
<td>7-10 years XXG4TR index</td>
<td>ASX 200 AS51 index</td>
<td></td>
</tr>
<tr>
<td>10+ years XXG5TR index</td>
<td>TSX SPTSX index</td>
<td></td>
</tr>
</tbody>
</table>

Table A1: The specific Bloomberg tickers

B. Regression details

<table>
<thead>
<tr>
<th>Country</th>
<th>Index</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia – REITAULC Index</td>
<td>Constant</td>
<td>-0.001</td>
<td>0.003</td>
<td>-0.340</td>
<td>0.735</td>
</tr>
<tr>
<td></td>
<td>ASX</td>
<td>0.631</td>
<td>0.073</td>
<td>8.586</td>
<td>0.000*</td>
</tr>
<tr>
<td></td>
<td>FI_m</td>
<td>0.732</td>
<td>0.169</td>
<td>4.323</td>
<td>0.000*</td>
</tr>
<tr>
<td></td>
<td>Credit</td>
<td>1.151</td>
<td>0.561</td>
<td>2.053</td>
<td>0.000*</td>
</tr>
<tr>
<td></td>
<td>Mean dependent variable</td>
<td>0.005</td>
<td>S.D dependent variable</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sum squared residual</td>
<td>0.260</td>
<td>S.E. of regression</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R-squared</td>
<td>0.404</td>
<td>Adjusted R-squared</td>
<td>0.396</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F(3,218)</td>
<td>49.318</td>
<td>P-value (F)</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Log-likelihood</td>
<td>434.110</td>
<td>Akaike criterion</td>
<td>-860.220</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schwarz criterion</td>
<td>-846.609</td>
<td>Hannan-Quinn</td>
<td>-854.725</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rho</td>
<td>0.060</td>
<td>Durbin-Watson</td>
<td>1.873</td>
<td></td>
</tr>
</tbody>
</table>

OLS, using observations 1994:02-2012:07 (T=222); significance at the 1% confidence level is denoted with a, and significance at the 10% confidence level is denoted with b.

<table>
<thead>
<tr>
<th>Country</th>
<th>Index</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada – G250CALC Index</td>
<td>Constant</td>
<td>-0.001</td>
<td>0.003</td>
<td>0.292</td>
<td>0.771</td>
</tr>
<tr>
<td></td>
<td>TSX</td>
<td>0.565</td>
<td>0.062</td>
<td>9.110</td>
<td>0.000*</td>
</tr>
<tr>
<td></td>
<td>FI_m</td>
<td>0.680</td>
<td>0.231</td>
<td>2.948</td>
<td>0.004a</td>
</tr>
<tr>
<td></td>
<td>Mean dependent variable</td>
<td>0.007</td>
<td>S.D dependent variable</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sum squared residual</td>
<td>0.410</td>
<td>S.E. of regression</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R-squared</td>
<td>0.295</td>
<td>Adjusted R-squared</td>
<td>0.289</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F(3,218)</td>
<td>45.856</td>
<td>P-value (F)</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Log-likelihood</td>
<td>386.000</td>
<td>Akaike criterion</td>
<td>-766.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schwarz criterion</td>
<td>-755.793</td>
<td>Hannan-Quinn</td>
<td>-761.879</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rho</td>
<td>0.168</td>
<td>Durbin-Watson</td>
<td>1.655</td>
<td></td>
</tr>
</tbody>
</table>

OLS, using observations 1994:02-2012:07 (T=222); significance at the 1% confidence level is denoted with a, and significance at the 10% confidence level is denoted with b.

<table>
<thead>
<tr>
<th>Country</th>
<th>Index</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switzerland – DBCHREF Index</td>
<td>Constant</td>
<td>0.005</td>
<td>0.002</td>
<td>2.154</td>
<td>0.032*</td>
</tr>
<tr>
<td></td>
<td>MSCI</td>
<td>0.157</td>
<td>0.031</td>
<td>5.142</td>
<td>0.000*</td>
</tr>
<tr>
<td></td>
<td>cash</td>
<td>-3.433</td>
<td>1.607</td>
<td>-2.136</td>
<td>0.340*</td>
</tr>
<tr>
<td></td>
<td>FI_short</td>
<td>1.800</td>
<td>0.381</td>
<td>4.727</td>
<td>0.000*</td>
</tr>
<tr>
<td></td>
<td>Mean dependent variable</td>
<td>0.005</td>
<td>S.D dependent variable</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sum squared residual</td>
<td>0.083</td>
<td>S.E. of regression</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R-squared</td>
<td>0.177</td>
<td>Adjusted R-squared</td>
<td>0.396</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F(3,218)</td>
<td>10.746</td>
<td>P-value (F)</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Log-likelihood</td>
<td>1203.053</td>
<td>Akaike criterion</td>
<td>-860.220</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schwarz criterion</td>
<td>-1020.917</td>
<td>Hannan-Quinn</td>
<td>-854.725</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rho</td>
<td>-0.112</td>
<td>Durbin-Watson</td>
<td>1.873</td>
<td></td>
</tr>
</tbody>
</table>

OLS, using observations 1995:02-2012:07 (T = 210); significance at the 1% confidence level is denoted with a, and significance at the 10% confidence level is denoted with b.
### European Union – EPEU Index

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.000</td>
<td>0.003</td>
<td>0.190</td>
<td>0.850</td>
</tr>
<tr>
<td>SXE</td>
<td>0.432</td>
<td>0.051</td>
<td>8.488</td>
<td>0.000*</td>
</tr>
<tr>
<td>FI_m</td>
<td>0.201</td>
<td>0.224</td>
<td>0.897</td>
<td>0.371</td>
</tr>
<tr>
<td>credit</td>
<td>2.083</td>
<td>0.552</td>
<td>3.775</td>
<td>0.000*</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.002</td>
<td>S.D dependent variable</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>0.262</td>
<td>S.E. of regression</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.442</td>
<td>0.434</td>
<td>2.237</td>
<td>0.026a</td>
</tr>
<tr>
<td>F(3,218)</td>
<td>57.531</td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>433.195</td>
<td>Akaike criterion</td>
<td>-858.390</td>
<td></td>
</tr>
<tr>
<td>Schwarz criterion</td>
<td>-844.779</td>
<td>Hannan-Quinn</td>
<td>-852.895</td>
<td></td>
</tr>
<tr>
<td>Rho</td>
<td>0.185</td>
<td>1.628</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OLS, using observations 1994:02-2012:07 (T=222); significance at the 1% confidence level is denoted with a, and significance at the 10% confidence level is denoted with b.

### United Kingdom – G250GBLC index

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.000</td>
<td>0.003</td>
<td>-0.089</td>
<td>0.929</td>
</tr>
<tr>
<td>UKX</td>
<td>0.814</td>
<td>0.080</td>
<td>10.198</td>
<td>0.000*</td>
</tr>
<tr>
<td>FI_long</td>
<td>0.318</td>
<td>0.142</td>
<td>2.237</td>
<td>0.026a</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.003</td>
<td>S.D dependent variable</td>
<td>0.061</td>
<td></td>
</tr>
<tr>
<td>Sum squared residual</td>
<td>0.542</td>
<td>S.E. of regression</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.332</td>
<td>0.326</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(3,218)</td>
<td>54.388</td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>352.630</td>
<td>Akaike criterion</td>
<td>-699.329</td>
<td></td>
</tr>
<tr>
<td>Schwarz criterion</td>
<td>-689.051</td>
<td>Hannan-Quinn</td>
<td>-695.138</td>
<td></td>
</tr>
<tr>
<td>Rho</td>
<td>0.081</td>
<td>1.835</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OLS, using observations 1994:02-2012:07 (T=222); significance at the 1% confidence level is denoted with a, and significance at the 10% confidence level is denoted with b.

### United States – REITUSLC Index

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.005</td>
<td>0.003</td>
<td>1.498</td>
<td>0.136</td>
</tr>
<tr>
<td>MSCI</td>
<td>0.514</td>
<td>0.089</td>
<td>5.744</td>
<td>0.000*</td>
</tr>
<tr>
<td>FI_m</td>
<td>0.568</td>
<td>0.243</td>
<td>2.332</td>
<td>0.021a</td>
</tr>
<tr>
<td>Credit</td>
<td>4.508</td>
<td>0.823</td>
<td>5.481</td>
<td>0.000*</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.009</td>
<td>S.D dependent variable</td>
<td>0.063</td>
<td></td>
</tr>
<tr>
<td>Sum squared residual</td>
<td>0.505</td>
<td>S.E. of regression</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.418</td>
<td>0.410</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(3,218)</td>
<td>52.150</td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>360.567</td>
<td>Akaike criterion</td>
<td>-713.134</td>
<td></td>
</tr>
<tr>
<td>Schwarz criterion</td>
<td>-699.523</td>
<td>Hannan-Quinn</td>
<td>-707.638</td>
<td></td>
</tr>
<tr>
<td>Rho</td>
<td>-0.033</td>
<td>2.060</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OLS, using observations 1994:02-2012:07 (T=222); significance at the 1% confidence level is denoted with a, and significance at the 10% confidence level is denoted with b.
Insurance risk transfer and categorization of reinsurance contracts

Eugene N. Gurenko
Lead Insurance specialist, World Bank Capital Markets and Non-Bank Finance Practice, World Bank

Alexander Itigin
Consultant, World Bank, and Senior Actuary at a global reinsurance company

Renate Wiechert
Consultant, World Bank, and Accounting Policy Consultant at a global reinsurance company

Abstract
Despite the existence of numerous quantitative approaches to the categorization of financial reinsurance contracts, insurance regulators may often find the practical implementation of the task to be technically challenging. This paper develops a simple, affordable and robust regulatory method that can help insurance regulators categorize whether financial reinsurance contracts classify as reinsurance. By reviewing real examples of different categorization methods, this paper explains how the proposed method standardizes such categorization. It also summarizes the existing pertinent literature on the subject with the view to help insurance regulators to first apply some simple indicators to flag the main issues with financial reinsurance contracts that may need further reviews.

Having identified the suspicious reinsurance contracts, supervisors may consider several solutions provided by the authors and, in some cases, requiring further quantitative testing of risk transfer contracts for categorization purposes; supervisors may also consider adopting the Standardized Expected Reinsurer's Deficit (SERD) approach to contract testing presented in this paper. The approach advocates the use of a simple standardized stochastic method that would allow market participants and regulators to perform robust quantitative tests quickly, and at an affordable cost. Besides addressing the obvious drawbacks of the “10-10” test, the proposed alternative method makes it possible to greatly reduce the technical challenges posed to the users of the Expected Reinsurer's Deficit approach based on full stochastic models with only a minimum loss of predictive accuracy.

1 We are very grateful for useful comments provided by our colleagues Charles Michael Grist, Senior Insurance Specialist, World Bank, and John Daniel Pollner, Lead Financial Officer, World Bank. The comments made in this article are those of the authors and are in no way representative of the views of the World Bank, or any of their associate organizations.
Introduction

Financial reinsurance, often referred to as “finite risk reinsurance” or just “finite reinsurance,” is a form of transaction between a reinsurance company and its client, which focuses mainly on purely financial effects such as capital management, solvency relief, influencing financial and earnings position, etc. rather than on the transfer of insurance risk, as is the case in traditional reinsurance. Very often financial reinsurance contracts combine features of both, a financial effect oriented financial instrument and a risk transfer oriented traditional reinsurance transaction. Categorization of such mixed financial reinsurance contracts and their eventual acceptance for supervisory purposes often becomes rather challenging and technically complex.

Despite the existence of several approaches to categorize financial reinsurance contracts, their technical complexity makes them practically unusable for insurance supervision purposes, particularly in developing countries. As a consequence, proper categorization of reinsurance contracts continues to remain a major challenge for many market players and insurance regulators, leaving ample room for mistakes, potential abuses and malpractice.

The paper aims to address this problem by first providing a thorough overview of the existing categorization methods, and then, presenting a simple and easy-to-implement quantitative regulatory method. The proposed method is much simpler and less costly than the current ones used in the market.

Why a correct categorization of reinsurance contracts matters

Besides transferring insurance risk, usually to a limited extent, financial reinsurance contracts also address other objectives, like financing or smoothing the profit and loss results. Consequently, from the outset, one has to determine whether the contract is transferring insurance risk at all, or only deals with financial risks. For example, a contract that protects only the investment results from the outset, one has to determine whether the contract is transferring insurance risk, as is the case in traditional reinsurance. In such cases, the contract may have to be separated into individual elements (so-called “unbundling”) to enable the application of the guidelines.

If it can be demonstrated that a contract transfers insurance risk, then an assessment has to be made of the amount of insurance risk transferred to a reinsurer. Depending on the contractual content and the accounting environment, those contracts that fail to provide enough risk transfer are categorized as financial instruments or accounted for under “deposit accounting” with no effect on underwriting results.

In most countries with well-developed insurance regulatory regimes, insurance companies are required by law to differentiate between traditional and financial reinsurance contracts in their statutory financial reporting. Those contracts that do not transfer enough risk to a reinsurer disqualify the cedant from obtaining solvency capital relief. Such adverse practical implications of regulatory categorization decisions underscore the importance of developing an accurate and easy to administer categorization method.

As has been already mentioned, in the use of financial reinsurance there might be cases of malpractice. The most common are the attempts to disguise self-financing (or borrowings) as a legitimate risk transfer for obtaining solvency relief or for smoothing the underwriting results. For regulatory purposes, therefore, it is important to define methods which would provide for adequate categorization of such contracts as either reinsurance or financial instrument/deposit accounting.

While the existing common international accounting guidelines, like International Financial Reporting Standards (IFRS) or the U.S. Generally Accepted Accounting Principles (US-GAAP), already provide guidance on how to distinguish between financial and insurance contracts, sometimes a contract may contain both elements, i.e., financial and insurance risks, thus making the application of existing guidance more difficult. In such cases, the contract may have to be separated into individual elements (so-called “unbundling”) to enable the application of the guidelines.

While the existing common international accounting guidelines, like International Financial Reporting Standards (IFRS) or the U.S. Generally Accepted Accounting Principles (US-GAAP), already provide guidance on how to distinguish between financial and insurance contracts, sometimes a contract may contain both elements, i.e., financial and insurance risks, thus making the application of existing guidance more difficult. In such cases, the contract may have to be separated into individual elements (so-called “unbundling”) to enable the application of the guidelines.

The E.U. Directive 2005/68/EC on reinsurance established a definition and some general criteria for financial reinsurance. It allows home member states to lay down specific provisions concerning the pursuit of financial reinsurance activities, like mandatory contract components, administrative and accounting procedures, internal control mechanisms, risk management requirements, accounting, prudential, and statistical information requirements, or rules relating to the available solvency margin (E.U. Council of Ministers (2005)).

2 Financial reinsurance is also known as “non-traditional reinsurance,” “limited risk reinsurance,” or “structured reinsurance.” Historically, the term “financial reinsurance” was used; however, as some fraudulent transactions took place in the past, the term “financial reinsurance” got a negative reputation. Consequently, the industry had established the new term “finite risk” or just “finite” to make clear that these contracts have sufficient risk transfer.

3 An U.S.-GAAP accounting method for the recognition of (re)insurance contracts without significant insurance risk transfer. Only the reinsurer’s margin is reported in the profit and loss statement, such that no premiums, commissions, or incurred losses are recorded. In the balance sheet also only a net asset or liability is reported.

4 For example, the E.U. Directive 2005/68/EC on reinsurance established a definition and some general criteria for financial reinsurance. It allows home member states to lay down specific provisions concerning the pursuit of financial reinsurance activities, like mandatory contract components, administrative and accounting procedures, internal control mechanisms, risk management requirements, accounting, prudential, and statistical information requirements, or rules relating to the available solvency margin (E.U. Council of Ministers (2005)).
The key to the categorization of reinsurance contracts lies in estimating the proportion of risk transfer obtained by a cedant through a given reinsurance transaction. To determine the extent of risk transfer in a reinsurance treaty, regulators may need to conduct quantitative tests. The main problem with the application of quantitative tests is their technical complexity, which considerably impairs their wide-spread use by the market and insurance regulators. Often, (re)insurance companies and regulatory bodies cannot perform quantitative tests on their own due to the lack of specialized technical resources. As a result, they may have to employ specialized actuarial consulting companies, which may be expensive or, in the case of developing countries, simply not available.

According to the International Association of Insurance Supervisors (IAIS), reinsurance is a key instrument for insurers and supervisors to conduct risk management [IAIS (2006)]. The insurance industry has significantly benefited from many product innovations. However, these innovations often bring along complex definitions of reinsurance contracts that not only affects insurers’ risk management practices and capital management but also their economic solvency and the regulatory treatment of solvency relief typically sought by insurers under reinsurance contracts [IAIS (2006)].

To date, there is no globally accepted definition of financial reinsurance. However, a typical transaction in many cases would include risk-limiting features like a self-retention of some risk by the insurer structured as a loss corridor, loss cap, or an aggregate limit of liability. Adjustable commissions, like sliding scale or profit commissions, or experience refunds in case of a positive or negative performance are also often used for aligning the interests of insurers and reinsurers. Financial reinsurance contracts are also often set up for multiple years and/or multiple lines of business to reduce volatility and aggregate risk. Although the IAIS lists these and some other characteristics of financial reinsurance contracts, it also points out that these could also be present in traditional reinsurance, thus rendering this guidance impractical in some cases [IAIS (2006)].

Under the E.U. directive 2005/68/EC Article 2, paragraph 1 (q) financial reinsurance is defined as follows: “finite reinsurance” means reinsurance under which the explicit maximum loss potential, expressed as the maximum economic risk transferred, arising both from a significant underwriting risk and timing risk transfer, exceeds the premium over the lifetime of the contract by a limited but significant amount, together with at least one of the following two features: (i) explicit and material consideration of the time value of money; (ii) contractual provisions to moderate the balance of economic experience between the parties over time to achieve the target risk transfer” [E.U. Council of Ministers (2005)].

As there can never be an exclusive list that defines the contract’s categorization, risk transfer assessments should always be conducted on a case by case basis. In the absence of a standardized approach there is ample room for misinterpretations, misrepresentations, and abuses (e.g., avoiding ratings downgrades by employing financial smoothing instruments, saving taxes, circumventing creditor lending conditions by disguising new borrowings as reinsurance, or avoiding or delaying supervisory intervention to prevent breaches of solvency) [IAIS (2006)]. As a result, there is an urgent need for a standardized practical approach to a robust regulatory classification of risk transfer contracts.

**How to categorize contracts as reinsurance**

As a first step, it has to be determined whether the reinsurance contract transfers insurance risk (e.g., a book of motor or liability insurance policies is reinsured) or financial risks. If under the relevant local statutory and accounting provisions, a contract can be seen as transferring financial risk only, then no further categorization is required. Such a contract cannot be considered as reinsurance and has to be accounted for under the investment result.

---

5 The IAIS list of most common features of risk financing instruments versus reinsurance contains the following characteristics: a) insurance risk transfer and financing are combined, b) assumption of limited risk by the reinsurer (e.g., aggregate limit of liability, blended cover, sliding scale and other adjustable commissions, loss corridors, and limits or caps), c) transfer of volatility (e.g., multiple lines of business, multiple years of account, and multiple year contract terms), d) inclusion of future investment income in price of contract (recognition of time value of money with funds withheld), e) potential profit sharing between parties (e.g., profit-sharing formulas, experience accounts), f) pricing determined by ceding insurers’ results and not reinsurance pricing cycle, and g) terms and pricing are typically determined in advance for the whole a block of new or in-force business (i.e., administration of reinsurance is done on a bulk basis rather than on a traditional seriatim policy-by-policy basis) [IAIS (2006)].

6 For example IFRS 4 Appendix A defines financial risks as: “The risk of a possible future change in one or more of a specified interest rate, financial instrument price, commodity price, foreign exchange rate, index of prices or rates, credit rating or credit index or other variable, provided in the case of a non-financial variable that the variable is not specific to a party to the contract.” [IASB (2004)]
Insurance risk transfer and categorization of reinsurance contracts

If the contract transfers insurance risk it has to be determined whether the risks transferred are sufficient to allow for reinsurance accounting. However, not every reinsurance contract must be tested. Under U.S.-GAAP and the NAIC regulations, the so-called “reasonably self-evident exception” or “safe harbor exception” has become common practice. It says that reinsurance contracts that contain only traditional contractual components do not have to be tested for a sufficient amount of risk transfer. Only in cases where a contract includes some characteristic risk transfer limiting features will a quantitative testing have to be performed.

Reasonably self-evident contracts
While the notion of “reasonably self-evident” is often not explicitly referenced in the accounting standards, it has become common practice. In essence, “reasonably self-evident” means that the given reinsurance contract can be seen as traditional reinsurance. For such contracts an explicit evaluation of risk transfer is not required. They are seen as “per definition” transferring a sufficient amount of risk. Obviously, evaluating risk transfer for each contract, irrespective of whether it is classified as traditional or financial reinsurance, would be costly and time consuming.

“Reasonably self-evident” only requires a minimum level of technical analysis and documentation to demonstrate compliance with the accounting standards. This opinion is widely shared by the insurance industry, auditors, and regulatory authorities.

Currently, there are no uniform criteria for categorizing a reinsurance contract as “reasonably self-evident” or, in other words, for distinguishing between traditional and financial reinsurance. Usually, the regulators or accounting standards only stipulate some minimum criteria and the companies have to apply their own more detailed list of internal criteria. However, this also gives companies some room for abuse.

Besides the criteria of the IAIS (IAIS (2006)) and the E.U. directive (E.U. Council of Ministers (2005)), the document “Reinsurance Attestation Supplement 20-1: Risk Transfer Testing Practice Note” prepared by the American Academy of Actuaries’ Committee on Property and Liability Financial Reporting (AAA (2007)), provides some more detailed examples on how a checklist for financial reinsurance could look like. EIOPA has also recently published a draft paper on the Solvency II treatment of financial reinsurance, which also includes a list of eligibility criteria

There are several important characteristics of such contracts where risk transfer is “reasonably self-evident.” According to the AAA (2007), there are reinsurance contracts that, in effect, comply with the “reasonably self-evident” principle by virtue of their belonging to a particular classification of reinsurance contracts. For example, risk transfer is reasonably self-evident for straight quota shares with fixed terms, e.g., no risk-limiting or any other variable terms like sharing positive or negative contract experience and with a fixed reinsurance commission that adequately compensates the ceding company for all acquisition costs. For most traditional per-risk or per-occurrence excess of loss reinsurance contracts (both treaty and facultative), risk transfer is reasonably self-evident if for a predetermined amount of premium the reinsurer assumes all or nearly all of the potential variability in the underlying losses, and it is evident from reading

7 The concept of “reasonably self-evident” was codified by the NAIC in 2005 as part of the Reinsurance Attestation Supplement 20-1. It addresses contracts for which a risk transfer test has to be carried out and also where detailed risk transfer testing is not required in order to conclude that the contract allows for reinsurance accounting (“For each such reinsurance contract entered into, renewed, or amended on or after January 1, 1994, for which risk transfer is not reasonably considered to be self-evident, documentation concerning the economic intent of the transaction and the risk transfer analysis evidencing the proper accounting treatment, as required by SSAP No. 62—Property and Casualty Reinsurance, is available for review”).

8 The CEIOPS’ Advice for Level 2 Implementing Measures on Solvency II: SCR standard formula - Article 111f Allowance of Reinsurance Mitigation Techniques includes a list of possible criteria: “Some of the following characteristics may be present within reinsurance contracts: • Insurance risk transfer, for example: – excess of loss reinsurance, which provides indemnification to the ceding insurer for each covered risk up to a predetermined limit. The ceding insurer is required to meet the obligations of the claim up to a preset amount before the reinsurer becomes liable; or - the insurer and the reinsurer share in an agreed ratio, all premiums, losses, and loss expenses arising out of the original business covered under the reinsurance agreement. There are two forms of proportional reinsurance: quota share and surplus share; or - catastrophe bonds issued to manage peak risks and embedded value securitization to help undertakings manage their capital more efficiently; • assumption of significant but limited risk by the reinsurer (e.g., aggregate limit of liability, blended cover, sliding scale and other adjustable commissions, loss corridors and limits or caps); • transfer of volatility (e.g., multiple lines of business, multiple years of account and multiple year contract terms); • inclusion of future investment income in price of contract (recognition of time value of money); • potential profit sharing between parties (e.g., profit-sharing formulas, experience accounts); • bulk reinsurance or treaty reinsurance (i.e., administration of reinsurance is done on a bulk basis rather than on a traditional policy-by-policy basis; a block of new or in-force business). Certain features can sometimes reduce the effective risk transfer considerably under the reinsurance contract. For example, this may be the case for certain finite reinsurance arrangements.” [CEIOPS (2009)]

9 A less restrictive, but generally accepted exception is the case of a straight quota share reinsurance contract with no risk-limiting features, other than a very high loss ratio cap with negligible effect on the economics of the transaction.
the basic terms of the contract that the reinsurer can incur a significant loss.\textsuperscript{10} In addition, for single year property catastrophe and casualty clash covers it can be demonstrated that these contracts fall under “reasonably self-evident” if no risk limiting features, like sub-limits, retrospective premium adjustments or other exclusions apply.

In general, it is less likely that risk transfer is reasonably self-evident, when most risk is retained by the ceding company and if certain experience-based contractual features, such as experience accounts, variable commissions, or premium adjustments, are included in the contract. When considering whether non-proportional contracts fall under the “reasonably self-evident” exemption it is also worth mentioning the “rate-on-line-criterion.”\textsuperscript{11} Even if no risk limiting features are included in the contract, a high premium (rate on line) can disqualify the contract from meeting the exemption rule. If the premium approaches the present value of the limit of coverage, risk transfer is usually no longer deemed to be reasonably self-evident, even if a contract has no risk-limiting features [AAA (2007)]. However, these kinds of contracts usually have other characteristic features like contingent commissions (to allow the ceding company to participate in the positive experience of the contract) and which would violate “reasonably self-evident” anyway.

Even though the topic has been in discussion for several decades, the definition of risk transfer still remains ambiguous. This discovery only underscores the importance of formulating a simple standardized approach to categorize financial reinsurance contracts.

**Not-reasonably self-evident contracts**

In case a contract does not meet the criteria of “reasonably self-evident” exemption, the contract falls into the “not-reasonably self-evident” category. For such contracts there are some commonly used quantitative criteria that help determine the amount of insurance risk transferred under a reinsurance contract. Below, we will describe the most relevant of such quantitative criteria.

It is important to note that a failure to satisfy the “reasonably self-evident” standard does not necessarily mean that a contract does not qualify as reinsurance contract. It simply means that more analysis, usually a quantitative risk transfer test, may be required to arrive at a contract categorization. The categorization approach under these quantitative methods can be summarized as follows. For a given reinsurance transaction to be categorized, each method calculates the value of a specific parameter that provides a quantitative measure for the amount of risk transferred by the reinsurance transaction. The resulting value is then compared with a specific parameter threshold value, which corresponds to the minimum level of risk transfer required by the method. If the parameter value is higher than the threshold, i.e., the transaction is transferring more risk than minimally required, then the transaction is categorized as reinsurance and accounted for as such.

Although there may be exceptions, contracts that falls under “not-reasonably self-evident,” (i.e., not automatically qualify for reinsurance) would typically have the following characteristics [AAA (2007)]: (a) non-proportional per risk, per occurrence or aggregate excess of loss contracts if the premium approaches the present value of the coverage provided and/or the contracts contain significant risk-limiting features or other variable features (e.g., profit commission); (b) contracts with experience accounts, i.e., sharing positive or negative experience of the contract, or similar provisions with a significant impact on the contract’s economics; (c) multi-year contracts with such provisions and/or provisions that adjust the contractual terms in later years, based on contractual experience in earlier years; and (d) proportional quota share contracts with risk-limiting features such as loss ratio caps, loss participations/corridors, or sub-limits or other variable features sharing positive or negative contract experience, like sliding scale commissions.

In such cases, for accounting and regulatory purposes, the company will need to evaluate the amount of risk transferred by the underlying contract. Risk transfer analyses may range from very simple premium to limit approaches, to highly sophisticated stochastic methods. In most cases, the rigor of the analysis is likely to be inversely correlated with the amount of risk transferred under the contract (e.g., the less risk that is transferred the more technical effort is required to determine the true extent of such transfer). In addition, internal processes

\textsuperscript{10} In many cases, there is no aggregate limit on the reinsurer’s loss. However, few practitioners would feel the need for a detailed probabilistic cash flow analysis to reach the conclusion that risk transfer is reasonably self-evident.

\textsuperscript{11} “Rate on line” is defined as the premium paid to reinsurer divided by the amount of reinsurance coverage.
have to be established as the regulator would usually request the company to present supporting documentation for the business rationale of the contract categorization.

Contracts which meet “virtually equivalent” condition
Under some regulatory regimes, contract categorization is entirely driven by the quantitative methods mentioned above. However, some more advanced accounting or regulatory regimes practice a more nuanced approach, which maintains that if the economic positions of the cedant (before risk transfer) and reinsurer (after risk transfer) are virtually equivalent for the ceded part of the underlying risk exposure, then the contract can be accounted for as reinsurance even if a quantitative risk transfer test is not fulfilled. In essence, virtual equivalence means that substantially all insurance risk relating to the reinsured portions of the underlying contracts has been assumed by the reinsurer. This condition is met only if some insignificant amount of insurance risk remains with the ceding enterprise on the reinsured portions of the underlying insurance contracts and the economic position of the reinsurer is equivalent to having written the underlying policies directly.12 In such contracts, the reinsurer de facto acts as the original insurer. For example, this exception can be found under U.S.-GAAP and U.S.-Statutory.13

Consequently, even if the chosen quantitative methods fail to provide the sufficient level of risk transfer, the contract can be accounted for as reinsurance given the economic positions of the reinsurer and the cedant are virtually equivalent. In this case, the low level of risk transfer provided just means that the original cedant’s risk for the ceded part of the underlying exposure had been already low before the transfer to a reinsurer.

The virtual equivalence-based categorization analysis is typically supplemented with a transaction study to establish whether the contract includes some specific features aimed at limiting the extent of risk transfer to reinsurers, such as loss caps, loss participations, loss corridors, sliding scale commissions, experience accounts, etc. Proving virtual equivalence for “not-reasonably self-evident” contracts is a challenging task, which mainly entails comparing the risk retained by the cedant with the risk transferred to the reinsurance company. An approach is suggested in the Casualty Actuarial Society Research Working Party paper on risk transfer testing [CAS (2005)].14 Further to comparing the risk exposure, the profit position of the cedant and the reinsurance company under the reinsurance contract also needs to be compared as well. For this, an analysis needs to be performed if significant positive contract experience is shared with the cedant, as mentioned in the Risk Transfer Practice Note by AAA’s COPFLER [AAA (2007)] or in the AICPA document “Evaluating Risk Transfer in Reinsurance of Short-Duration Contracts” [AICPA (2003)]. For some reinsurance contracts, such as non-proportional contracts, virtual equivalence is difficult to demonstrate. If the virtual equivalence cannot be demonstrated, the contract would not qualify under this exception.15

Summary of relevant reinsurance contracts categorization steps:
1. Reasonably self-evident: the first step in the categorization approach is to determine whether the contract belongs to the “reasonably self-evident” category. For contracts from this category, no quantitative risk transfer test is required. Such contracts are considered to transfer a sufficient amount of insurance risk by virtue of the class and/or individual contract characteristics. Such contracts are classified as “traditional reinsurance.”16

---

12 However, note that the introduction of risk limiting features to a quota-share contract, such as a loss ratio cap (other than one that is so high it’s effect on the economics of the contract is de minimis), a loss retention corridor, or a sliding scale commission, often prevents the contract from qualifying for the exception.

13 FAS 113 par. 11 and 67 states (FASB (1992)); 11. If, based on this comparison, the reinsurer is not exposed to the reasonable possibility of significant loss, the ceding enterprise shall be considered indemnified against loss or liability relating to insurance risk only if substantially all of the insurance risk relating to the reinsured portions of the underlying insurance contracts has been assumed by the reinsurer. 4 Footnote 4: This condition is met only if insignificant insurance risk is retained by the ceding enterprise on the reinsured portions of the underlying insurance contracts. 67. Under very limited circumstances, the reinsurer need not be exposed to the reasonable possibility of significant loss for a contract to meet the conditions for reinsurance accounting. For example, applying the “reasonable possibility of significant loss” condition is problematic when the underlying insurance contracts themselves do not result in the reasonable possibility of significant loss to the ceding enterprise. The Board concluded that, when the reinsurer has assumed substantially all of the insurance risk in the reinsured portions of the underlying policies, even if that risk does not result in the reasonable possibility of significant loss, the transaction meets the conditions for reinsurance accounting. In this narrow circumstance, the reinsurer’s economic position is virtually equivalent to having written the insurance contract directly. The risks retained by the ceding enterprise are insignificant, so that the reinsurer’s exposure to loss is essentially the same as the insurer’s. For easier reference the new FASB Accounting Standard Codification is not mentioned. FAS 113 par. 11 and 67 can now be found under FASB ASC 944-20-15-51 through 15-53 and FASB ASC 944-20-55-55.

14 “Because “substantially all” is less than “all,” if the EUD faced by the reinsurer is within a small tolerance of the expected underwriting deficit faced by the cedent, say, within 0.1%, then we would also say the “substantially all” test is met.” [CAS (2005)]

15 See U.S.-GAAP: EITF D-34 Q&A 24; FASB ASC 944-20-55-56.

16 Depending on the local accounting or regulatory bodies, companies might have to implement internal criteria for defining “reasonably self-evident.” Some examples can be found in the appendix of the Risk Transfer Practice Note by AAA’s COPFLER [AAA (2007)].
2. **Not reasonably self-evident**: some type of quantitative cash flow analysis must be performed to assess the extent of risk transfer.

3. **Substantially all/virtually equivalent**: even if not reasonably self-evident or a significant risk transfer test is met, reinsurance accounting might be applicable depending on the economic impact of the contractual features.\(^\text{17}\)

By placing contracts in one of these three categories one can considerably reduce the amount of technical work by focusing only on those contracts that require further testing.

**Existing quantitative risk transfer methods**

While most reinsurance contracts are designed to protect the ceding company from adverse financial effects of one or more insured events by transferring the risk to a reinsurer, there are situations when the positive economic effect of a reinsurance contract on the ceding company cannot be easily determined. In such cases, a quantitative testing of a reinsurance contract must be performed to prove the existence and the extent of risk transfer, assuming relevant data is available.

Quota share reinsurance with its characteristic risk transfer limiting elements, such as sliding scale commission, loss participation, loss ratio cap, experience refund, etc., is an example of a financial reinsurance contract that needs to be subjected to further quantitative testing for the purposes of categorization. Computer models that perform scenario testing may be required in such cases to perform the tests. Relevant data are either based on historical results of the business in question or similar businesses. Scenario testing can either be deterministic or stochastic. Stochastic approaches might comprise a comprehensive set of possible stochastic scenarios (often called stochastic models) or its subsets.

Risk transfer testing is essentially a discounted cash flow test for the assumed scenarios [Vendetti and Freihaut (2008)].

Most of the risk transfer testing approaches can be broken down into the following three steps:

**Step 1.** Analysis of the underlying risk exposure and defining the loss scenarios: In this step, the underlying risk exposure needs to be analyzed with the objective to define the characteristic loss scenarios.

**Step 2.** Analysis of the reinsurance transaction and modeling the resulting cash flows: all terms and conditions of the contract relevant for the cash flow between the cedant and the reinsurer should be taken into account. Based on this analysis and scenarios defined in Step 1, all cash flows between the cedant and the reinsurer, as defined by the reinsurance contract, should be modeled for each scenario. These include claims, premiums, commissions, loss participations, experience based premium provisions, etc. The timing of any cash flow should be taken into account as well. In addition, all amounts should be deemed to be paid, i.e., the analysis should be performed irrespective of whether funds are transferred or not (so-called deposit retained or funds withheld).

**Step 3.** Cash flow analysis and deriving a quantitative measure of risk transfer: in this step, the cash flow modeled under Step 2 should be analyzed. This analysis includes discounting with suitable interest rates, often risk-free rates for different maturities depending on the timing of each cash flow. Alternatively, one single risk free rate can be used instead of different rates for different maturities. In this case, the duration of the interest rate should be chosen approximately equal to that of the net cash flows. Then, for each scenario, all considered cash flows should be summed up to the total discounted positive and negative cash flows from the reinsurer’s perspective. Finally, through taking into account all considered scenarios with their total positive and negative cash flows, each method will determine whether there is a reasonable possibility of significant loss to the reinsurer based on the method’s specific underlying quantitative risk transfer parameter and threshold.

Below, we describe the most common quantitative methods used by reinsurance practitioners to assess the extent of risk transfer in a financial reinsurance contract.

**Most common quantitative risk transfer methods**

The most common quantitative risk transfer testing methods are the “10-10” rule and the expected reinsurer’s deficit (ERD). Both have been well described in the pertinent literature [CAS (2002),

---

\(^{17}\) Depending on the local accounting or regulatory bodies, companies might have to implement internal thresholds for defining “substantially all risks” and “virtually equivalent.”
AAA (2005), Vendetti and Freihaut (2008)). These are briefly summarized below. Another, less common method is the “premium to limit of coverage ratio.”

**Premium to limit of coverage ratio:** under this criterion, the contract is unlikely to be a risk transfer, if the value of the ratio approaches or exceeds 1. This easy-to-apply method, which is not scenario-based and therefore does not require an in-depth analysis of the underlying exposure, might be used as a quick first test. However, in many cases the method may generate inaccurate results when the ratio is slightly below 1, indicating that the degree of risk transfer is insufficient. Also, where the premium to limit of coverage ratio is substantially under 1, the objective of the contract might still be purely financial even though according to the method the contract should be categorized as reinsurance.

The “10-10” rule: this rule was loosely derived from the accounting standard language that required that a reinsurer faced a “reasonable possibility of a significant loss.” According to the “10-10” rule, a reinsurance contract exhibits significant risk transfer characteristics if there is at least a 10% probability of an at least 10% loss relative to the cash inflows of the reinsurer (usually reinsurance premiums). Even though the rule does translate risk transfer into an easy-to-apply benchmark, its potential to be applied in practice has substantial shortcomings (as described in CAS (2005); AAA (2005)). For instance, some conventional reinsurance contracts that normally would be classified by regulators and reinsurance practitioners as traditional reinsurance do not pass the “10-10” rule. This is, for example, the case with excess of loss property catastrophe contracts which fail to pass the rule because the frequency of major catastrophes is so low that there may not be a 10% chance of a loss for a reinsurer. Yet, despite much lower odds of the loss, the reinsurer may potentially end up paying the full amount of the treaty limit. Moreover, even ordinary quota share reinsurance treaties designed for transferring high frequency but low severity losses from insurers’ portfolios may also fail the “10-10” test. However, it was common practice not to disqualify catastrophe excess of loss contracts or favorable quota share treaties as reinsurance because “10-10” was not met. For low severity-high frequency contracts the aforementioned “substantially all/virtually equivalent” criterion was applied. For high severity-low frequency contracts, however, no accounting guidance was applicable.

Furthermore, we must mention at least two other major shortcomings of the “10-10” rule. First, the “10-10” rule ignores the risk in the tail of distribution beyond the 90th percentile. Only the present value loss at the 90th percentile (VaR) is taken into account. However, in the right tail of the distribution the loss potential for catastrophe covers can be significant. Second, we must point out that the selected parameters for frequency and severity in the “10-10” test are completely arbitrary and can be replaced by a “5-20” test or a “1-100” test. Because of these shortcomings a new test, the so-called “product rule,” was developed. Every combination that would lead to an at least 1% (10%*10%) threshold could also be applied for testing sufficient risk transfer. This “product test” was shortly replaced by the ERD, although the “product test” did solve the shortcomings for high severity-low frequency or high severity-low frequency contracts. However, in fairness to the “10-10” rule or the “product rule,” we must point out that these tests were never intended to become universally applicable risk transfer tests. These tests had emerged as informal quantitative methods of experienced practitioners to determine whether reinsurance contracts contained sufficient risk transfer.

Expected reinsurer’s deficit method (ERD): another common measure of risk transfer that has gained acceptance among regulators and reinsurance practitioners is the ERD. In 2002, the CAS Valuation, Finance, and Investments Committee published the paper “Accounting rule guidance statement of Financial Accounting Standards No. 113 – considerations in risk transfer testing,” which discussed the shortcomings of the “10-10” rule and introduced the ERD (CAS (2002)). The ERD overcomes the shortcomings of the “10-10” rule by including the right tail of

---

18 Also known as 9a/b-test under US-GAAP. FAS 113 par 9a/ b states: “a. The reinsurer assumes significant insurance risk under the reinsured portions of the underlying insurance contracts. b. It is reasonably possible that the reinsurer may realize a significant loss from the transaction.” Guidance also available under FASB ASC 944-20-15-41.

19 FAS 113 par. 10 states: “The ceding enterprise’s evaluation of whether it is reasonably possible for a reinsurer to realize a significant loss from the transaction shall be based on the present value of all cash flows between the ceding and assuming enterprises under reasonably possible outcomes, without regard to how the individual cash flows are characterized. The same interest rate shall be used to compute the present value of cash flows for each reasonably possible outcome tested.” FAS 113 par 11 states: “Significance of loss shall be evaluated by comparing the present value of all cash flows, determined as described in paragraph 10, with the present value of the amounts paid or deemed to have been paid to the reinsurer.” Guidance also available under FASB ASC 944-20-15-49 through 15-54.
the distribution in the risk transfer test. In addition, one single measure was developed that allowed the same treatment for low frequency-high severity, high frequency-low severity, and moderate frequency-moderate severity contracts. However, conducting an ERD test represents a considerable technical challenge due to the need to generate a realistic distribution of reinsured losses that are likely to be incurred by the cedant during the life-time of the reinsurance contract. In the case of reinsurance contracts that provide coverage for more than one line of business, the technical challenge of drawing numerous loss distributions and combining them into one single loss distribution at the portfolio level can be even more daunting.

Mathematically, the ERD test can be defined as follows [see Ruhm and Brehm (2007); CAS (2005); AAA (2005) for more details]:

\[
ERD = \frac{p \times T}{P} \geq A
\]

Where, \( p \) = probability of net economic loss; \( T \) = expected (average) severity of net economic loss (present value), when it occurs; \( P \) = expected reinsurance premium (present value) or, more general, the cash inflows to the reinsurer; and \( A \) = a threshold above which a contract is considered to have provisionally “passed” the “significant” risk transfer test and below which it is considered to have “failed”; usually, \( A \) is set at 1%, which has become an international best practice standard.

Since the ERD incorporates information about both the frequency and severity of the reinsurer’s downside risk into one single measure, it allows utilizing a combined numeric threshold for significant risk transfer (\( A \)) rather than defining it separately in terms of frequency and severity.

Although the above definition of ERD test provides a good intuitive understanding of the motivation behind the ERD test, in the scenario-based ERD calculation framework, it is often more convenient to rephrase the definition in the form of the following equation:

\[
ERD = \frac{(\sum \pi \cdot p \cdot S_i)}{P} \geq A
\]

Where, the expression in brackets calculates the expected severity of the net economic loss to the reinsurer among all considered scenarios 1 to \( N \). The probability \( p_i \) denotes the occurrence probability of the scenario \( i \), and \( S_i \) the severity of the net economic loss to the reinsurer in the corresponding scenario (present value). If, for a given scenario, net economic result of the reinsurer is positive, i.e., this scenario produces a net economic profit, then for this scenario \( S_i \) is equal to zero. Obviously, for \( S_i \) we can state:

\[
S_i = \min (0; \text{Present value of the reinsurer’s net result})
\]

Please notice that in the above definitions the severity of net economic loss, and therefore also ERD, is always positive or zero. In the last formula, this is achieved by setting the minus sign before the min function. Further, please note that each contract that qualifies as risk transfer under the “10-10” rule also fulfills the ERD test at 1% threshold of ERD (a 10% loss multiplied by a 10% probability is a 1% ERD). However, not every contract which fulfills the ERD test (say, a 100% loss multiplied by 1% probability is also a 1% ERD) would also meet the “10-10” rule, cf. the discussion of the shortcomings of the “10-10” rule presented above.

A simple example below illustrates the use of ERD test (Table 1). Assume, for a reinsurance contract, the following holds: Reinsurance premium \( P = 10 \) million. The aggregate probability of net economic loss is calculated as: \( p = 2.5% + 1% + 0.5% = 4% \). The expected severity of net economic loss under the condition that it occurs: \( T = (2.5\% \times 30 \, \text{m} + 1\% \times 72 \, \text{m} + 0.5\% \times 200 \, \text{m}) / 4\% = 61.8 \text{m} \). Then, according to the first variant of the calculation formula given above, we obtain ERD = \( \frac{p \times T}{P} = 4\% \times 61.8 \, \text{m} / 10 = 24.7\% \). We also could apply the second variant of the calculation formula which of course leads to the same result:

\[
ERD = \frac{(\sum \pi \cdot p \cdot S_i)}{P} = 0\% \cdot 0 + 2.5\% \cdot 30 + 1\% \cdot 72 + 0.5\% \cdot 200 \} / 10 = 24.7\%
\]

Obviously, in this example the resulting ERD is well above

<table>
<thead>
<tr>
<th>Probability of net result (p)</th>
<th>Present value of net result (m)</th>
<th>Severity of net economic loss (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>96.0%</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>2.5%</td>
<td>-30</td>
<td>30</td>
</tr>
<tr>
<td>1.0%</td>
<td>-72</td>
<td>72</td>
</tr>
<tr>
<td>0.5%</td>
<td>-200</td>
<td>200</td>
</tr>
</tbody>
</table>

Table 1: Example results of a reinsurance contract
Insurance risk transfer and categorization of reinsurance contracts

the threshold of 1% and thus according to the ERD test, the reinsurance contract transfers enough risk to be classified as a reinsurance transaction.

The simplified example above only uses premium in the denominator as a fixed variable. However, in practice, premiums sometimes are not fixed upfront but depend on the contract’s loss experience or result. In such cases, the premiums in the ERD calculation should be considered at their expected values for which different approaches are possible and used by the industry. Among other aspects, the choice of an approach may also depend on contract conditions, especially if with the increasing contract loss the premium increases or decreases (the latter case is rather rare but can be seen in practice as well; such contracts are often called contracts with inverse character).

In developed markets, market participants utilize either full or partial stochastic models of the underlying loss exposure for calculating the ERD. Both the partial and full stochastic models are based on the loss scenarios, with occurrence probability assigned to each scenario. While a full stochastic model tries to capture a possibly comprehensive set of loss scenarios, a partial stochastic model works with its suitable subset. Obviously deriving a partial stochastic model is easier. However, in practice, often partial stochastic models may be insufficient for proving the desired ERD threshold, thus necessitating the use of a full stochastic method.

A full stochastic model of the underlying loss exposure will usually differentiate between the basic loss exposure (losses below a certain threshold), individual large loss exposure and event-based catastrophe loss exposure. If, for a reinsurance transaction, one of the three loss categories does not seem relevant, then the loss exposure model can be abandoned. The full stochastic method, nevertheless, must still account for the total annual, individual or per event loss scenarios with the probabilities assigned to each loss scenario. These loss scenarios are then tested through a mathematical model of the underlying reinsurance transaction that produces a stochastically generated set of cash flows between the reinsurer and the cedant resulting from the underlying reinsurance transaction. These cash flows are then discounted to the present point in time, and recalculated into the values of the discounted reinsurer’s deficit per each scenario.

To illustrate how this method works, we provide a more detailed numeric illustration of the ERD method for a typical quota share reinsurance treaty with elements of finite risk transfer in Appendix 1.

In addition to the “10-10” rule and the ERD, other risk transfer methods have been developed; however none of them has become as well accepted as the 10-10 rule and later the ERD test. For more details on these other methods please refer to Wang (2008), Vendetti and Freihaut (2008), and Ruhm (2001).

To summarize, despite the existence of numerous approaches to categorization of reinsurance contracts, including ERD, which is currently viewed as best practice, none of them can be easily performed. Even for international (re)insurers, performing the quantitative risk transfer tests represents a substantial technical burden, let alone small insurance companies and regulatory bodies in developing markets. This creates room for mistakes, misuse, and malpractice. Hence, the need for a simple but robust risk transfer categorization method still remains.

SERD method: standardized ERD test
As described in the previous section, the main drawback of the ERD is the technical complexity of its application. To address this problem, we developed a method that can be followed by insurance regulators to carry out the ERD test without employing complex actuarial techniques.

The proposed Standardized ERD (SERD) method aims to help regulators and insurers in developing countries to apply the ERD method at an affordable cost. The SERD relies on the ERD method applied to proportional reinsurance contracts. In a nutshell, the SERD represents a simplified approach to exploiting a full stochastic model of the underlying risk exposure for calculating ERD. However, instead of developing an individual model for each given case, the SERD utilizes a standardized model template which needs to be fed with only a few relatively easy-to-obtain parameters, without imposing a significant technical burden on a user. With this approach, the SERD method follows the logic of the European Solvency II approach that allows an insurance company to apply a “simplified” approach by using the Standard Formula instead of a “full” or “partial internal model,” if the latter represents a heavy technical (and cost) burden for the company. Hence, the SERD can be seen as the “standard formula” method of the ERD calculation framework.
The level of method customization can be adjusted depending on how much information about the individual risk exposure to be modeled is available to the user. For example, the volatility parameters of the chosen probabilistic loss distributions or the loss development patterns can either be used at their default values, which will be suggested by the SERD method, or, if available, at their unique values for a given individual case. The default parameter values have been taken from the Solvency II (QIS5) and Swiss Solvency Test frameworks (see Appendix 2 for more details).

On the basis of the user inputs as well as the default values of risk exposure parameters, the SERD method then provides a comprehensive set of probabilistic scenarios (stochastic model) for the severity of the reinsurer's net economic loss (similar to the format of Table A2 in Appendix 1) and finally automatically calculates the ERD based on the obtained probabilistic scenarios.

The SERD method comprises the following four modules:

1. **Loss modeling module** – The user will be given the possibility of choosing the line of business covered by the contract from the list comprising the main standard non-life lines of business. Following the approach of Solvency II and Swiss Solvency Test [European Commission (2010); Federal Office of Private Insurance (2006)], the SERD method allows us to assume some correlation between basic losses in different lines of business. However, no correlation is assumed between basic, large and catastrophe loss exposures.

   In the loss modeling module, the SERD tool allows us to differentiate between basic, large and cat loss burdens. The basic loss modeling is based on the assumption that the annual basic loss ratio is distributed according to the lognorm distribution. With this choice we follow the Solvency II (QIS5) approach [European Commission (2010)]. The user will be asked to enter the expected ultimate loss ratios\(^20\) for all lines of business covered by the contract. Further, the user will be given the choice to use the individual user defined values of the volatility per line of business or to use the default values imported by the SERD method from the Solvency II framework.

   The modeling of the individual large loss burden is based on the recognized international best practice assumption that the number of individual large losses per line of business is distributed according to the Poisson distribution and the loss severity according to the Pareto distribution. The user will be asked to provide the expected number of individual large losses per annum as well as lower and upper loss thresholds for each line of business covered by the contract. For the shape parameter of the loss severity distribution (Pareto alpha), the user will be given the option to provide its own unique value, if available, or to use a standard default value. The default values can be derived from the industry experience in other more mature markets. The choices used in the SERD method draw on the values of Pareto alpha suggested in the Swiss Solvency Test Standard Model [Federal Office of Private Insurance (2006)].

   If there is a catastrophe loss exposure, the user will be asked to provide a loss exceedance curve describing the exposure. At a later stage, the SERD method can be supplemented with catastrophe risk modeling modules for various natural catastrophe scenarios, which will undertake assessments of natural catastrophe risk based on sum insured aggregates (to be provided by the user) instead of a loss exceedance curve. This will further reduce the technical burden on the user while improving the accuracy of calculations.

2. **Reinsurance transaction module** – this module captures the risk transfer effect of the reinsurance transaction, taking into account its main features, such as reinsurance and profit commissions, loss participations and loss corridors, loss caps, etc. For each value of the underlying exposure's loss ratio, this module encapsulates the corresponding value of the reinsurance result as a percentage of the reinsurance premium. Together with modules 1 and 3, module 2 will be the basis for obtaining stochastic scenarios for the severity of the reinsurer's net economic loss and finally for calculating the ERD in module 4.

3. **Loss developments patterns module** – on the basis of the loss development patterns for each line of business covered

\(^{20}\) The term “ultimate loss ratio” is used to describe the ratio between the ultimate loss attached to an annual contract term period after its full development and the insurance premium belonging to this annual term period.
Insurance risk transfer and categorization of reinsurance contracts

by the contract, the SERD method calculates the discounting adjustments in the present value of the reinsurer's net economic loss required for calculating ERD. The user will be asked to provide loss development patterns for each line of business. In those cases where the patterns are not available, the SERD method will provide default market patterns for each line of business. For the reinsurance premium and all kinds of commissions, the SERD method assumes the full payment within one year (no development after the first year).

4. Scenario generating and the ERD calculating module - in this module, the SERD method generates probabilistic scenarios for the severity of the reinsurer's net economic loss and then, based on these scenarios, calculates ERD and cumulative probabilities for different severity values in the same manner as shown in Appendix 1 for the example of “financial quota share.” Behind the scenario generating module is a stochastic Monte Carlo simulation which is carried out automatically based on the user specific and default parameters as defined in modules 1, 2 and 3. This automatic simulation engine represents one of the main strengths of the suggested method. In a conventional ERD calculation framework, stochastic Monte Carlo simulation is usually one of the most technically challenging parts, which is often too heavy a burden for many market participants and regulators. The suggested SERD method fully automates the most difficult part of the calculation so that the user does not need to worry about it, the method does it automatically.

To illustrate our approach, we applied the suggested SERD method to the Financial Quota Share example shown in Appendix 1, where the full-fledged stochastic model was applied, and then compared the ERD results obtained with both methods.

As a more detailed description of the SERD method for this example has been provided in Appendix 2, below we provide only the results of the SERD Module 4 – the probabilistic scenarios for the severity of the reinsurer’s net economic loss and the resulting ERD. Table 2 shows the results of the SERD method in the same format as the results of the conventional ERD calculation.

ERD calculates to

\[ \text{ERD} = \sum_{n=1}^{26} A_a B_n \Rightarrow \text{ERD} = 1.4\% \]

As we can see, with the ERD value of 1.4% resulting from the SERD method, while the contract still fulfills the ERD test, this value deviates from the value obtained with a full stochastic model (1.9%) (Table A2 in Appendix 1). The reason for this deviation is that the default value for the standard deviation of the basic loss ratio deviates from the specific value used when deriving the full stochastic model. Default value is 10% and in the full model we worked with 12%. Also, for the shape parameter of the severity distribution for the individual large loss exposure (Pareto alpha) there has been a difference between the value used in the full stochastic model (2.4) and the default value in the SERD method (2.5). The observed differences in results demonstrate that with the default values selected by the user under the SERD approach the contract shows less risk transfer. This can be seen from the higher probabilities of the lower loss severities in the Table 2, as compared to Table A2, and of

<table>
<thead>
<tr>
<th>Scenario N (aggregated scenarios)</th>
<th>A. Realization probability</th>
<th>B. Severity of reinsurer’s net economic loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>78.6%</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>3.0%</td>
<td>1%</td>
</tr>
<tr>
<td>3</td>
<td>2.6%</td>
<td>2%</td>
</tr>
<tr>
<td>4</td>
<td>2.3%</td>
<td>3%</td>
</tr>
<tr>
<td>5</td>
<td>2.1%</td>
<td>4%</td>
</tr>
<tr>
<td>6</td>
<td>1.8%</td>
<td>5%</td>
</tr>
<tr>
<td>7</td>
<td>1.4%</td>
<td>6%</td>
</tr>
<tr>
<td>8</td>
<td>1.4%</td>
<td>7%</td>
</tr>
<tr>
<td>9</td>
<td>1.2%</td>
<td>8%</td>
</tr>
<tr>
<td>10</td>
<td>0.9%</td>
<td>9%</td>
</tr>
<tr>
<td>11</td>
<td>0.8%</td>
<td>10%</td>
</tr>
<tr>
<td>12</td>
<td>0.7%</td>
<td>11%</td>
</tr>
<tr>
<td>13</td>
<td>0.6%</td>
<td>12%</td>
</tr>
<tr>
<td>14</td>
<td>0.4%</td>
<td>13%</td>
</tr>
<tr>
<td>15</td>
<td>0.4%</td>
<td>14%</td>
</tr>
<tr>
<td>16</td>
<td>0.3%</td>
<td>15%</td>
</tr>
<tr>
<td>17</td>
<td>0.3%</td>
<td>16%</td>
</tr>
<tr>
<td>18</td>
<td>0.2%</td>
<td>17%</td>
</tr>
<tr>
<td>19</td>
<td>0.2%</td>
<td>18%</td>
</tr>
<tr>
<td>20</td>
<td>0.1%</td>
<td>19%</td>
</tr>
<tr>
<td>21</td>
<td>0.1%</td>
<td>20%</td>
</tr>
<tr>
<td>22</td>
<td>0.1%</td>
<td>21%</td>
</tr>
<tr>
<td>23</td>
<td>0.1%</td>
<td>22%</td>
</tr>
<tr>
<td>24</td>
<td>0.1%</td>
<td>23%</td>
</tr>
<tr>
<td>25</td>
<td>0.3%</td>
<td>24%</td>
</tr>
<tr>
<td>26</td>
<td>0.0%</td>
<td>&gt;=25%</td>
</tr>
</tbody>
</table>

Table 2: SERD method, results for the Financial Quota Share example
course from the lower resulting ERD value of the SERD method compared to the full stochastic method. Obviously, this deviation in results is the price one has to pay for omitting the technical burden of the full stochastic model.

It is worth noting however, that if all unique distribution parameters were available, our proposed SERD method would produce the same result as the full stochastic ERD method. Consequently, we believe that the suggested SERD method is considerably less costly to implement, which makes it a robust alternative to the ERD, which relies on full-fledged unique stochastic models of the risk.

In conclusion, we would like to note that other quantitative risk transfer tests, such as, for example, the “10-10” rule, could also be automated in the same manner as the suggested SERD method automates the ERD test. The approach to calculating the ERD values proposed in this paper can, therefore, be also viewed as a methodological framework applicable to a wide range of risk transfer tests used by insurance regulators and the insurance industry.

Summary

Despite the existence of numerous quantitative approaches to the categorization of financial reinsurance contracts, insurance regulators may often find the practical implementation of the task to be technically challenging. To simplify the categorization process, in this paper, we first conveniently summarize the existing pertinent literature on the subject with the view to help insurance regulators to first apply some simple indicators to flag the main issues with financial reinsurance contracts that may need further reviews.

Some of such obvious “red flags” are [IAIS (2006)]: contracts, including clauses, that change the nature of how risks are transferred; contracts including different and diverse lines of businesses, making it difficult to assess the risk and exposures of the transfer; contracts where cedants do not follow formal processes nor guidelines; signed contracts near financial year end covering past periods or earlier years; or, when contracts are backdated (i.e., replicating a retroactive coverage); and contracts combining financial reinsurance with traditional reinsurance contracts, making it difficult to assess the two contracts separately.

Having identified the suspicious reinsurance contracts, supervisors may consider (a) conducting on-site inspections of reinsurance programs and risk management practices, (b) requesting annual attestations from the management on risk transfers reporting accuracy, (c) further expanding actuaries’ responsibility to cover the analysis of risk transfer content in reinsurance contracts when submitting reinsurers’ system assessments, (d) reviewing companies’ annual reinsurance plans, and (e) implementing reinsurance “whistle blower” programs mainly focusing on actuaries and auditors [IAIS (2006)].

In some cases, requiring further quantitative testing of risk transfer contracts for categorization purposes, supervisors may also consider adopting the SERD approach to contract testing presented in this paper. The approach advocates the use of a simple standardized stochastic method that would allow market participants and regulators to perform robust quantitative tests quickly and at an affordable cost. Besides addressing the obvious drawbacks of the “10-10” test, the proposed alternative method allows the user to greatly reduce the technical challenges posed by the ERD approach based on full stochastic models with acceptable loss of predictive accuracy.

References

Insurance risk transfer and categorization of reinsurance contracts

Taxing Times, 5(1)

Ernst & Young LLP, 2012, “Financial reporting developments – a comprehensive
guide: accounting for reinsurance”

and of the council of 16 November 2005 on reinsurance and amending Council
Directives 73/239/EEC, 92/49/EEC as well as Directives 98/78/EC and 2002/83/EC”

Specifications”, Brussels
http://ec.europa.eu/internal_market/insurance/docs/solvency/qis5/201007/
technical_specifications_en.pdf

Federal Office of Private Insurance, 2006, Schweizerische Eidgenossenschaft,
“Technical document on the Swiss Solvency Test”
http://www.finma.ch/archiv/bvp/download/e/SST_techDok_061002_E_wo_
Li_20070118.pdf

113, accounting and reporting for reinsurance of short-duration and long-duration
contracts”

Freihaut, D., and P. Vendetti, 2009, “Common pitfalls and practical considerations
in risk transfer analysis.” Casualty Actuarial Society EForum

International Accounting Standards Board (IASB), 2004, “International Financial
Reporting Standard 4 Insurance Contracts”

International Association of Insurance Supervisors (IAIS), 2006, “Guidance
paper on risk transfer, disclosure and analysis of finite reinsurance,” (Guidance paper
No. 11)

National Association of Insurance Commissioners (NAIC), 2000, “Statement of
statutory accounting principles No. 62, property and casualty reinsurance”

contracts. A summary of the report by the CAS Research Working Party on Risk
Transfer Testing” Variance, 1(1)

Figure A1: Sliding scale reinsurance commission

Figure A2: Ceded loss ratio with loss cap

Appendix 1: ERD calculation for a financial quota share (FQS)
reinsurance contract.
Contract details: proportional quota share contract for motor
third party liability (MTPL) business; Ceded premium = € 100m;
Sliding scale reinsurance commission:

Loss ratio (LR) <= 50% → Reinsurance commission (RI) = 47%
Loss ratio (LR) >= 87% → Reinsurance commission (RI) = 10%

For LRs between 50% and 87%, RI commission drops by 1% for
any LR increase of 1% (Figure A1).

> Loss ratio is capped at 120% – this feature means that the
reinsurer does not accept any loss arising from the underlying
contract beyond 120%. The original losses beyond this loss
ratio are not ceded into the underlying reinsurance contract
“ceded LR” = min (LR, LR cap) (Figure A2).

> Ceded’s loss participation of 20% between ceded LRs of
90% and 120% – this means that the cedant will reimburse
the reinsurer for 20% of the loss between the ceded LRs of 90%
and 120%; “loss participation” = 20% * min(30%; max( “ceded
LR” – 90%; 0)). For example, if the LR = 80%, then the Ceded
LR = 80% and the loss participation is zero. If on the other
hand, LR = 125%, Ceded LR = 120%, there will be some non-
zero loss participation: “Loss participation” = 20% * min(30%;
max(120% - 90%; 0))= 6%. Please note that with this definition,
the reinsurer’s loss participation is always positive or zero.
When modeling the risk exposure for this contract, we differentiate between basic losses below certain threshold and individual large losses above this threshold. Whereas basic losses are modeled on the annual aggregate basis, large losses are modeled individually through a combination of loss frequency and loss severity probabilistic assumptions. For the sake of simplicity, we assume that there are no loss development patterns, i.e., losses are fully settled within one year. Hence, loss discounting effects do not play any role.

Let us assume that from the analysis of loss and premium statistics from the previous years, we can derive the following assumptions:

Basic losses
- Losses below the threshold of €3m
- Distribution assumption: ultimate LR follows a lognormal distribution
- Expected ultimate basis loss ratio = 80%
- Standard deviation of the ultimate basic loss ratio = 12%

Individual large losses
- Losses above the threshold of €3m
- Distribution assumptions: number of losses p.a. (loss frequency) follows a Poisson distribution, loss severity follows a truncated Pareto distribution
- Expected number of individual large losses above €3m (Poisson Lambda) = 0.5
- Shape parameter of Pareto Distribution (Pareto Alpha) = 2.4
- Lower threshold of Pareto distribution = €3m, upper threshold = €100m

To calculate ERD we carry out a Monte Carlo simulation which generates 100,000 scenarios for the total ultimate loss ratio. The realization probability for each scenario, is equal 1/100,000 = 0.001%. For each scenario we then apply the contract conditions to calculate the reinsurer’s ultimate result and then the severity of reinsurer’s net economic loss (if any).

First we calculate reinsurer’s ultimate result which is the reinsurer’s result after the insured loss has been settled, i.e., has achieved its ultimate value.

Reinsurer’s ultimate result = 100% – “ceded LR” – “RI comm” + “loss participation”

Now, we recalculate the ultimate result into its present value.

Present value of reinsurer’s result = “reinsurer’s ultimate result” + “discounting adjustment”

Please note that according to our previous assumption, the loss will be fully settled within one year, i.e., there will be no loss development over the time and the resulting discounting adjustment can be assumed to be zero. However, for the sake of completeness we provide the formula for a general case.

Finally, we recalculate the present value of reinsurer’s result into the severity of reinsurer’s net economic loss

Severity of reinsurer’s net economic loss (as percentage of the ceded premium) = – min (0; “Present value of reinsurer results”); = – min (0; 100% – “ceded LR” – “RI comm” + “loss participation”)

The following example for both loss scenarios Ultimate LR = 80% and Ultimate LR = 125% will illustrate the above formula.

As explained above, for the Ultimate LR = 80% Ceded LR = 80%, RI Commission = 17% and loss participation = 0; and for the Ultimate LR = 125%, Ceded LR = 120%, RI Commission = 10% and loss participation = 6%. Then, for both scenarios, we calculate for the Reinsurer’s deficit as follows:

Severity of reinsurer’s net economic loss (as percentage of the ceded premium) for the loss scenario LR = 80% = – min (0; 100% – 80% – 17% + 0) = min(0; 3%) = 0.

Severity of reinsurer’s net economic loss (as percentage of the ceded premium) for the loss scenario LR = 125% = – min (0; 100% – 120% – 10% + 6%) = – min(0; – 24%) = 24%.

Out of the resulting scenario set for severities of reinsurer’s net economic loss, we can easily calculate a reliable estimate for the expected reinsurer’s deficit (ERD) (Table A1).
Insurance risk transfer and categorization of reinsurance contracts

This stochastic model has been obtained from the Monte Carlo simulation for the underlying risk exposure (Column A) and applying the terms and conditions of the reinsurance contract (Columns C, D, and E). The reinsurance result for each scenario (Column F) is then recalculated into the severities of reinsurer’s net economic loss (Column G).

Obviously, for calculating ERD we are only interested in the economic loss severities in Column G. Working with 100,000 scenarios might be somewhat inconvenient. To reduce the number of scenarios to be considered, we can now round up the severities to the full percentage digits and aggregate realization probabilities for the scenarios resulting in the same value of the (rounded) loss severity. The results can be found in Table A2.

Then, ERD can be calculated as the total of the row-wise products of the columns A and B

\[
\text{ERD} = \sum_{i=1}^{26} A_i \times B_i = 1.9\%
\]

Obviously, with this result the underlying contract successfully passed the ERD risk transfer test.

<table>
<thead>
<tr>
<th>Scenario N</th>
<th>A. Realization probability</th>
<th>B. Loss ratio</th>
<th>C. Reinsurance commission</th>
<th>D. Ceded LR</th>
<th>E. Loss participation</th>
<th>F = 100% - B - C + D</th>
<th>Reinsurer’s result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001%</td>
<td>77.6%</td>
<td>19.4%</td>
<td>77.6%</td>
<td>0.0%</td>
<td>22.4%</td>
<td>0.0%</td>
</tr>
<tr>
<td>2</td>
<td>0.001%</td>
<td>69.0%</td>
<td>28.0%</td>
<td>69.0%</td>
<td>0.0%</td>
<td>3.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>3</td>
<td>0.001%</td>
<td>104.3%</td>
<td>10.0%</td>
<td>104.3%</td>
<td>2.9%</td>
<td>-11.4%</td>
<td>11.4%</td>
</tr>
<tr>
<td>4</td>
<td>0.001%</td>
<td>90.8%</td>
<td>10.0%</td>
<td>90.8%</td>
<td>0.2%</td>
<td>-0.7%</td>
<td>0.7%</td>
</tr>
<tr>
<td>5</td>
<td>0.001%</td>
<td>65.3%</td>
<td>31.7%</td>
<td>65.3%</td>
<td>0.0%</td>
<td>3.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>6</td>
<td>0.001%</td>
<td>102.1%</td>
<td>10.0%</td>
<td>102.1%</td>
<td>2.4%</td>
<td>-9.7%</td>
<td>9.7%</td>
</tr>
<tr>
<td>7</td>
<td>0.001%</td>
<td>79.4%</td>
<td>17.6%</td>
<td>79.4%</td>
<td>0.0%</td>
<td>3.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>8</td>
<td>0.001%</td>
<td>80.1%</td>
<td>16.9%</td>
<td>80.1%</td>
<td>0.0%</td>
<td>3.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>9</td>
<td>0.001%</td>
<td>109.8%</td>
<td>10.0%</td>
<td>109.8%</td>
<td>4.0%</td>
<td>-15.9%</td>
<td>15.9%</td>
</tr>
<tr>
<td>10</td>
<td>0.001%</td>
<td>74.7%</td>
<td>22.3%</td>
<td>74.7%</td>
<td>0.0%</td>
<td>3.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>11</td>
<td>0.001%</td>
<td>62.8%</td>
<td>34.2%</td>
<td>62.8%</td>
<td>0.0%</td>
<td>3.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>12</td>
<td>0.001%</td>
<td>78.7%</td>
<td>18.3%</td>
<td>78.7%</td>
<td>0.0%</td>
<td>3.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>13</td>
<td>0.001%</td>
<td>89.1%</td>
<td>10.0%</td>
<td>89.1%</td>
<td>0.0%</td>
<td>0.9%</td>
<td>0.0%</td>
</tr>
<tr>
<td>14</td>
<td>0.001%</td>
<td>61.2%</td>
<td>35.8%</td>
<td>61.2%</td>
<td>0.0%</td>
<td>3.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>99,999</td>
<td>0.001%</td>
<td>85.5%</td>
<td>11.5%</td>
<td>85.5%</td>
<td>0.0%</td>
<td>3.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>100,000</td>
<td>0.001%</td>
<td>89.9%</td>
<td>10.0%</td>
<td>89.9%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Table A1: Stochastic model for calculating ERD, 100,000 scenarios

<table>
<thead>
<tr>
<th>Scenario N (aggregated scenarios)</th>
<th>A. Realization probability</th>
<th>B. Severity of reinsurer’s net economic loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75.4%</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>2.8%</td>
<td>1%</td>
</tr>
<tr>
<td>3</td>
<td>2.5%</td>
<td>2%</td>
</tr>
<tr>
<td>4</td>
<td>2.3%</td>
<td>3%</td>
</tr>
<tr>
<td>5</td>
<td>2.0%</td>
<td>4%</td>
</tr>
<tr>
<td>6</td>
<td>1.9%</td>
<td>5%</td>
</tr>
<tr>
<td>7</td>
<td>1.7%</td>
<td>6%</td>
</tr>
<tr>
<td>8</td>
<td>1.5%</td>
<td>7%</td>
</tr>
<tr>
<td>9</td>
<td>1.3%</td>
<td>8%</td>
</tr>
<tr>
<td>10</td>
<td>1.2%</td>
<td>9%</td>
</tr>
<tr>
<td>11</td>
<td>1.1%</td>
<td>10%</td>
</tr>
<tr>
<td>12</td>
<td>1.0%</td>
<td>11%</td>
</tr>
<tr>
<td>13</td>
<td>0.8%</td>
<td>12%</td>
</tr>
<tr>
<td>14</td>
<td>0.6%</td>
<td>13%</td>
</tr>
<tr>
<td>15</td>
<td>0.6%</td>
<td>14%</td>
</tr>
<tr>
<td>16</td>
<td>0.5%</td>
<td>15%</td>
</tr>
<tr>
<td>17</td>
<td>0.4%</td>
<td>16%</td>
</tr>
<tr>
<td>18</td>
<td>0.4%</td>
<td>17%</td>
</tr>
<tr>
<td>19</td>
<td>0.3%</td>
<td>18%</td>
</tr>
<tr>
<td>20</td>
<td>0.3%</td>
<td>19%</td>
</tr>
<tr>
<td>21</td>
<td>0.2%</td>
<td>20%</td>
</tr>
<tr>
<td>22</td>
<td>0.2%</td>
<td>21%</td>
</tr>
<tr>
<td>23</td>
<td>0.2%</td>
<td>22%</td>
</tr>
<tr>
<td>24</td>
<td>0.1%</td>
<td>23%</td>
</tr>
<tr>
<td>25</td>
<td>0.8%</td>
<td>24%</td>
</tr>
<tr>
<td>26</td>
<td>0.0%</td>
<td>&gt;=25%</td>
</tr>
</tbody>
</table>

Table A2: Stochastic model for calculating ERD, aggregated scenarios
Appendix 2: SERD method for the financial quota share example

In this Appendix, we provide details of the suggested SERD method and demonstrate its application to the financial reinsurance quota share example presented above, where the ERD was calculated conventionally with help of a full stochastic model. To carry out the calculations with the suggested SERD method, we applied a prototype SERD tool which was developed by the authors. The figures provided in this Appendix show the input and output interfaces of this tool. Furthermore, we provide the explanation on how to work with these interfaces.

The starting point of the SERD method is to choose one or more line of business contributing to the overall risk exposure covered by the contract (Figure A3). With this set, we follow the Solvency II QIS5 approach for Non-life and non-SLT Health [European Commission (2010)]. For the financial quota share example, we choose the line of business “motor vehicle liability” by marking the corresponding line in the dark grey area (in this and all following figures depicting the tool interfaces, dark grey marks the input fields).

The SERD method consists of the following four modules: (1) loss modeling module; (2) reinsurance transaction module; (3) loss development patterns module; and (4) scenario generating and ERD calculating module. Modules (1) and (3) assess the overall contract risk exposure, module (2) examines the risk transfer effect of the reinsurance transaction, and finally module (4) calculates the ERD value. In modules (1), (2), and (3), the SERD tool requests the user to enter some parameters and specific characteristics of the underlying risk exposure and the reinsurance contract (please see the detailed description of the modules provided below).

1. Loss modeling module

Basic loss – In the loss modeling module, the method allows for differentiating between basic, large and cat loss burdens. The basic loss modeling is based on the assumption that the loss ratio of the basic loss, aggregated over the annual contract term, is distributed according to the lognorm distribution.

---

21 Please note that the approach described below applies for proportional reinsurance contracts. As non-proportional contract usually do not provide any substantial solvency capital relief, their regulatory treatment and categorization as reinsurance or financial instruments usually do not include any solvency capital requirements related aspects.
Insurance risk transfer and categorization of reinsurance contracts

With this choice, we follow the Solvency II (QIS5) approach [European Commission (2010)]. The user is asked to enter the expected ultimate loss ratios for all lines of business covered by the contract. Figure A4 shows the basic loss interface of the ERD tool for the financial quota share example. In the first dark grey column, the user provides the expected ultimate loss ratio of 80% as valid for this example.

In the second input column, the user specifies whether they would use unique parameters of the insurance undertaking (following Solvency II framework we call these parameters USP – Undertaking Specific Parameters) for the standard deviation of the ultimate loss ratio or standard default values. If the USP values are used, these values are provided in the third dark grey column. Otherwise, the tool will use the default standard deviation values, as shown in Table A3. These values are taken from Solvency II QIS5 (QIS5 Technical Specification, Par. SCR.9.25, SCR.8.72) [European Commission (2010)].

Even though, in the financial quota share example presented above, the value of 12% was available for the standard deviation, in the SERD calculation example this USP value was omitted and the default value for the underlying line of business was used instead (Figure A4). This was done to demonstrate the approach for a case when USP values were not available, and also to examine the effect of using default values instead of the USP values on the resulting ERD later on.

Following Solvency II and Swiss Solvency Test [European Commission (2010); Federal Office of Private Insurance (2006)], the SERD method assumes some embedded correlations among different lines of business covered by the contract in their basic loss exposure. However, no correlation is assumed among the loss categories basic, large and catastrophe losses.

Large loss – following the inputs in the basic loss section, the user proceeds to the next large loss section. The user will be asked to provide the expected number of the individual large losses per annum as well as the lower and upper loss thresholds for each line of business covered by the contract (Figure A4). This was done to demonstrate the approach for a case when USP values were not available, and also to examine the effect of using default values instead of the USP values on the resulting ERD later on.

Following Solvency II and Swiss Solvency Test [European Commission (2010); Federal Office of Private Insurance (2006)], the SERD method assumes some embedded correlations among different lines of business covered by the contract in their basic loss exposure. However, no correlation is assumed among the loss categories basic, large and catastrophe losses.

<table>
<thead>
<tr>
<th>Line of business</th>
<th>Lower loss threshold = CHF</th>
<th>Lower loss threshold = CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVL (motor vehicle liability)</td>
<td>2.5</td>
<td>2.8</td>
</tr>
<tr>
<td>MVC-hail (motor vehicle comprehensive hail)</td>
<td>1.85</td>
<td>1.85</td>
</tr>
<tr>
<td>Property</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>Liability</td>
<td>1.8</td>
<td>2</td>
</tr>
<tr>
<td>UVG including UVGZ</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Health collective and individual</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Transport</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Finance and surety</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>Others</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Table A4: Pareto alpha by lines of business (Swiss solvency test standard model)

<table>
<thead>
<tr>
<th>Chosen LoB’s</th>
<th>Expected number of large losses p.a.</th>
<th>Lower loss threshold</th>
<th>Upper loss threshold</th>
<th>Use USP for Pareto alpha?</th>
<th>USP</th>
<th>Pareto alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor vehicle liability</td>
<td>0.5</td>
<td>3000000</td>
<td>100000000</td>
<td>NO</td>
<td>2.5</td>
<td></td>
</tr>
</tbody>
</table>

Figure A5: SERD tool: large loss interface

<table>
<thead>
<tr>
<th>Line of business</th>
<th>Pareto alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor vehicle liability</td>
<td>2.5</td>
</tr>
<tr>
<td>Motor, other classes</td>
<td>1.8</td>
</tr>
<tr>
<td>Marine, aviation, transport (MAT)</td>
<td>1.5</td>
</tr>
<tr>
<td>Fire and other property damage</td>
<td>1.4</td>
</tr>
<tr>
<td>Third-party liability</td>
<td>2</td>
</tr>
<tr>
<td>Credit and suretyship</td>
<td>0.75</td>
</tr>
<tr>
<td>Legal expenses</td>
<td>1.8</td>
</tr>
<tr>
<td>Assistance</td>
<td>1.5</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>1.5</td>
</tr>
<tr>
<td>Medical expense</td>
<td>3</td>
</tr>
<tr>
<td>Income protection</td>
<td>0.75</td>
</tr>
<tr>
<td>Workers’ compensation</td>
<td>2</td>
</tr>
</tbody>
</table>

Table A5: Default values of Pareto alpha by line of business

Individual large losses per line of business is distributed according to the Poisson distribution and the loss severity according to the Pareto distribution. The parameter “Pareto alpha” describes the shape of the loss severity distribution.

Similar to the loss ratio standard deviation in the basic loss section, the user is given an option to provide the USP values for Pareto alpha when available or to use the standard default values.
These default values can be derived from the industry experience in other more mature markets. The choices used in our tool draw on the values of Pareto alpha suggested in the Swiss Solvency Test Standard Model [Federal Office of Private Insurance (2006)]. Also for Pareto alpha, we omitted providing the USP value and worked with the default value instead. The reason is the same as in case of the standard deviation – to demonstrate the approach in the case when USP value were not available and to examine the effect of using default values instead of the USP values on the resulting ERD later on. With this reference from the Swiss solvency test at hand, we selected the following parameters for our line of business classification (Table A5). Please note, that the lower the parameter alpha the more prone the distribution is to higher losses as illustrated by Figure A6.

Catastrophe loss - If there is a catastrophe loss exposure, the user is asked to provide the loss exceedance curve describing the exposure. The corresponding tool interface is shown in Figure A7. For the financial quota share example, we assumed that there is no exposure to natural catastrophe losses. Hence the input field is left empty.

At a later stage, the SERD method can be supplemented with catastrophe risk modeling capabilities to carry out assessments of natural catastrophe risk based on the sum insured aggregates instead of substituting them with a loss exceedance curve. This will further reduce the technical burden of the calculation when using the SERD method.

2. Reinsurance transaction module
This module captures the risk transfer effect of the reinsurance transaction taking into account its main features such as reinsurance and profit commissions, loss participations and loss corridors, loss caps, etc. For each value of the underlying risk exposure loss ratio, the user is asked to provide the corresponding value of the transaction result from the reinsurer’s perspective (as percentage of reinsurance premium), see the first dark grey column in Figure A8.

Further, the user will be asked to provide the value of reinsurance premium, the second dark grey column in Figure A8.

For the financial quota share example, the calculation of the reinsurer’s result depending on the ultimate loss ratio has been already explained above, when the example was introduced. Hence, here we only repeat the calculation formula:

Reinsurer’s ultimate result (as [%] of the reinsurance premium) = 100% - “ceded LR” - “RI comm.” + “loss particip.” + “discounting adjustment”, where “ceded LR”, “RI comm.” and “loss particip” are to be calculated according to the formulae provided above.
Insurance risk transfer and categorization of reinsurance contracts

With help of the above formula, the present values of the reinsurer’s result can be easily calculated for different values of the ultimate loss ratio, cf. the first dark grey column in Figure A8.

Together with Modules 1 and 3, Module 2 will form the basis for obtaining stochastic scenarios for the severity of the reinsurer’s net economic loss and ultimately for calculating ERD in Module 4.

3. Loss development patterns module
In this module, the user is asked to provide loss development patterns for each chosen line of business (on the cumulative paid basis). In cases when the USP patterns are not available, the SERD method will provide default market patterns for each line of business. For the reinsurance premium and all kinds of commission, we assume the full payment within one year.

Based on the development patterns, the SERD tool calculates the discounting adjustment factors for each line of business. These adjustment factors are used to calculate the present value of the reinsurer’s net economic loss required for calculating ERD.

For the financial quota share example, we assume that the loss is fully developed, i.e., reaches its ultimate value within one year and hence there is no discounting adjustment. According to our assumption, 100% is entered for the first year and zeros for all following years (Figure A9).

4. Scenario generating and ERD calculating module
In this module, the SERD method generates the probabilistic scenarios for the severity of the reinsurer’s net economic loss and then, based on these scenarios, finally calculates ERD and cumulative probabilities for different possible severities of the net economic loss. Behind this calculation is a stochastic Monte Carlo simulation that is carried out automatically based on the user specific and default parameters provided in Modules 1, 2 and 3. This automatic simulation engine represents one of the main strengths of the suggested method. In a conventional ERD calculation framework, a stochastic Monte Carlo simulation usually represents one the most technically challenging parts that is often too heavy a burden for many market participants and regulators. The suggested SERD method eliminates the need for having the user perform this most difficult part of the calculation by running it automatically. For example, in the case of the Financial Quota share, the result shown in Figure A10 was obtained automatically from the tool.

As explained above, due to our previous assumption on the development patterns for the Financial Quota Share example, no discounting adjustment has been made. In a generic case, the discounting adjustment, based on the loss development pattern, as calculated in Module 2, would be taken into account when calculating the reinsurer’s net economic loss in Module 4. In this example, the estimated ERD is 1.4%.

In conclusion, Figure A11 provides a flow chart which summarizes all the intermediate steps of the SERD method as implemented in the above described prototype tool.

<table>
<thead>
<tr>
<th>Loss ratio</th>
<th>Reinsurer’s ultimate result as % of the reinsurance premium</th>
<th>Reinsurance premium</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>49%</td>
<td>100,000,000</td>
</tr>
<tr>
<td>2%</td>
<td>48%</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>46%</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>47%</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>48%</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>87%</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>88%</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>89%</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>91%</td>
<td>-0.8%</td>
<td></td>
</tr>
<tr>
<td>92%</td>
<td>-1.6%</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>120%</td>
<td>-24%</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>200%</td>
<td>-24%</td>
<td></td>
</tr>
</tbody>
</table>

Figure A8: SERD tool. Reinsurance transaction interface

<table>
<thead>
<tr>
<th>Chosen LoB’s</th>
<th>Use UP for loss development pattern</th>
<th>YR 1</th>
<th>YR 2</th>
<th>YR 3</th>
<th>YR 4</th>
<th>YR 5</th>
<th>YR 6</th>
<th>YR 7</th>
<th>YR 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor vehicle liability</td>
<td>NO</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Default pattern</td>
<td>40%</td>
<td>30%</td>
<td>20%</td>
<td>5%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td></td>
</tr>
</tbody>
</table>

Figure A9: SERD tool. Loss development patterns interface
### Table

<table>
<thead>
<tr>
<th>A. Realization probability</th>
<th>B. Severity of reinsurer’s net economic loss (as [%] of the RI premium)</th>
<th>ERD = sumproduct (Col A, Col B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>78.6%</td>
<td>0%</td>
<td>1.4%</td>
</tr>
<tr>
<td>3.0%</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>2.6%</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>2.3%</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>2.1%</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>1.8%</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>1.4%</td>
<td>6%</td>
<td></td>
</tr>
<tr>
<td>1.4%</td>
<td>7%</td>
<td></td>
</tr>
<tr>
<td>1.2%</td>
<td>8%</td>
<td></td>
</tr>
<tr>
<td>0.9%</td>
<td>9%</td>
<td></td>
</tr>
<tr>
<td>0.8%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>0.7%</td>
<td>11%</td>
<td></td>
</tr>
<tr>
<td>0.6%</td>
<td>12%</td>
<td></td>
</tr>
<tr>
<td>0.4%</td>
<td>13%</td>
<td></td>
</tr>
<tr>
<td>0.4%</td>
<td>14%</td>
<td></td>
</tr>
<tr>
<td>0.3%</td>
<td>15%</td>
<td></td>
</tr>
<tr>
<td>0.3%</td>
<td>16%</td>
<td></td>
</tr>
<tr>
<td>0.2%</td>
<td>17%</td>
<td></td>
</tr>
<tr>
<td>0.2%</td>
<td>18%</td>
<td></td>
</tr>
<tr>
<td>0.1%</td>
<td>19%</td>
<td></td>
</tr>
<tr>
<td>0.1%</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>0.1%</td>
<td>21%</td>
<td></td>
</tr>
<tr>
<td>0.1%</td>
<td>22%</td>
<td></td>
</tr>
<tr>
<td>0.1%</td>
<td>23%</td>
<td></td>
</tr>
<tr>
<td>0.3%</td>
<td>24%</td>
<td></td>
</tr>
<tr>
<td>0.0%</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>0.0%</td>
<td>26%</td>
<td></td>
</tr>
<tr>
<td>0.0%</td>
<td>27%</td>
<td></td>
</tr>
<tr>
<td>0.0%</td>
<td>28%</td>
<td></td>
</tr>
<tr>
<td>0.0%</td>
<td>29%</td>
<td></td>
</tr>
<tr>
<td>0.0%</td>
<td>30%</td>
<td></td>
</tr>
</tbody>
</table>

### Figure A10: SERD Tool, ERD Output Tableau

### Figure A11: Flow chart SERD method with the prototype tool
Superior information and compensation fees of active mutual funds

Chekib Ezzili
Equity Derivatives, NATIXIS
Patrice Poncet
Distinguished Professor of Finance, ESSEC Business School

Abstract
We posit a fund manager and an individual investor who maximize the expected (log) utility of their respective terminal wealth. The manager possesses more information than the investor does and charges the latter, their would-be customer, a linear compensation fee. The investor will delegate their portfolio decisions to the manager if, and only if, the expected utility of their wealth after fees is larger than the expected utility they can achieve by directly investing in the market. Our framework, which uses a mathematical result by [Amendinger (2000)], allows us to characterize compensation fees in terms of information differential.
Investors who delegate their portfolio decisions to professional managers against compensation fees assume implicitly that the latter have superior skills. Goetzmann et al. (2003) found that, in extreme cases, the global (regular plus performance) fees earned by fund managers could represent up to 30% or even 40% of the fund’s net asset value. According to the same study, rational investors should expect that the fund earns an additional risk-adjusted return (alpha) in the range of 200 to 500 basis points to compensate for these fees. More generally, the required positive alpha is presumably due to better judgment, superior data gathering and processing abilities, and maybe some privileged access to private information permitted by a closer relationship with, say, prime brokers.

The portfolio delegation issue is complex and can be tackled from various angles, which explains why the extant literature is voluminous. For example, many studies focus on the optimal design of the contract binding investors and their fund manager. Admati and Pfleiderer (1997) investigate what would constitute an appropriate benchmark in a one-period setting, and show that the use of a risky benchmark cannot be easily rationalized. Ou-Yang (2003) extends their analysis to a multi-period setting that allows for a dynamic, rather than a static, benchmark. Starks (1987) compares symmetric and asymmetric compensation schemes and shows that symmetric contracts, wherein the manager participates in both gains and losses, dominate asymmetric ones as they induce the manager to choose the portfolio strategy preferred by the investor.1 Grinblatt and Titman (1993) study the case of call-like asymmetric contracts and find that a manager who can hedge their compensation fee will choose investment strategies that increase the fund’s volatility. Goetzmann et al. (2003) use partial equilibrium analysis to evaluate the option granted to the manager in the case of an asymmetric compensation scheme, and express alpha as a function of the manager’s abilities and the contract’s implicit incentives. Carpenter (2000) and Basak et al. (2007) evaluate the impact of incentive fees on fund performance and find in particular that some asymmetric schemes encourage the manager to take reckless risks, since then they have very little to lose and much to gain. Other studies focused on the relationship between fund performance and cash inflows or outflows. For instance, Berk and Green (2004) find that the flow-performance relationship is consistent with both high average amount of skill and strong heterogeneity across managers. Hugonniere and Kaniel (2010) solve the equilibrium problem that arises when investors are allowed to trade dynamically in the fund, which in turn affects the manager’s own trading strategy. Finally, there is some scant literature regarding the general equilibrium analysis of delegated portfolio management. Recently, Cuoco and Kaniel (2011) studied the impact of managers’ decisions on asset allocation by fund investors and direct investors and on asset prices within a general equilibrium dynamic model, where the extent of portfolio delegation by investors and the parameters of the compensation scheme are both endogenous.

All of the aforementioned studies assume away any information asymmetry between the manager and investors. A fund’s outperformance is “explained” by other factors such as its manager’s “intrinsic” abilities. By contrast, the main contribution of this paper is to explicitly introduce an information differential. The latter, which could be translated in terms of management skill, is precisely what motivates portfolio delegation in the first place.

We solve the manager’s and the investor’s optimization programs by combining the classical approaches of Karatzas et al. (1986) and Cox and Huang (1989) in complete markets and the findings of Amendinger (2000) and Grorud and Pontier (2001) that provide a way to deal with an environment where managers detain more information than their customers. For tractability, we assume that the manager’s fees are proportional to the fund’s net asset value. It is optimal for the investor to delegate their portfolio decisions, if by so doing they achieve an expected utility (welfare) – at least as large as the one they could attain through direct investment. We solve for the manager’s problem and characterize the endogenous proportional fee in terms of the information differential. Under the special case of log utility, the solution obtains explicitly.

The economic framework

We consider a financial market that is frictionless and free of arbitrage. Trading is assumed to take place continuously. There are numerous assets available for trade, such that diversification is possible, including a riskless asset (say a money market

---

1 This need not be always the case. Das and Sundaram (2001) in a static setting and Cuoco and Kaniel (2011) in a dynamic framework show examples where asymmetric contracts can dominate.
account) that yields an instantaneous interest rate which is generally time-varying. Participating in the market is, in particular, an investor who must decide whether to invest directly or to delegate their portfolio decisions and a professional fund manager. Their investment horizon is known and fixed (say at future time T). The manager is supposed to be an insider, i.e., to possess some information in addition to the common knowledge possessed by the investor. The extra information consists in knowing initially the outcome of some random variable relevant to some asset prices or returns. This variable may be, for example, the price of a specific asset at time T, or its price at time T distorted somewhat by a random noise, or else the value of some source of risk (we provide different possible examples below).

Although we skip all technicalities, it should be noted that the market prices of risk (or risk premia associated with all risky assets) implicitly used by the direct investor and the manager are different because of the superior knowledge possessed by the latter. This is the economic mechanism by which the portfolio allocation of the manager will differ from what the investor would choose, if they decided to invest directly.

Precisely, we investigate the situation where the investor has to decide optimally whether to directly invest in the market or to delegate their portfolio decisions to the better informed manager. In the former case, the investor does not benefit from the manager’s information or skill, while in the latter situation they do, but in exchange for a compensating management fee.

We will assume that both the manager and the investor are risk averse, try to maximize the expected utility of their terminal wealth (at date T) and are endowed with a logarithmic utility function. This means that their relative risk aversion is a constant (equal to one) irrespective of their wealth’s level. This choice is extensively encountered in the literature, is motivated by simplicity as it allows for quasi-explicit solutions, and leads to results that are easy to interpret.

The manager’s program
We limit our study to linear compensation schemes, where the manager’s fees are proportional to the fund’s terminal asset value. This is the case for the vast majority of mutual funds, and a small but non-negligible fraction of hedge funds [Ackermann et al. (1999)]. We ignore the complexities entailed by a two-fold compensation structure that includes an additional incentive fee since it is exceptional in mutual funds. The adopted linear scheme eases the interpretation of the results and allows us to focus on the impact of superior information possessed by the manager, irrespective of any particular effects brought about by a convex compensation structure.

The fee due by the investor is paid at the maturity date T of the delegation contract (the investment horizon.) Consequently, assuming they have no outside capital, the manager’s wealth at time T is equal to the management fee they receive:

\[ \text{Fee}(T) = fV_M(T), \]

where \( f, (0 < f < 1) \), is the constant, contractual fee rate to be determined endogenously and \( V_M(T) \) is the fund’s gross value at terminal date T.

Solving the manager’s optimization problem allows us to express the (random, as of initial date) optimal gross value of the fund, \( V_M(T) \), as a positive, linear function of \( V_M(0) \), the investor’s initial wealth devoted to the fund, and a non-linear function of the interest rate, the asset risk premia, and the sources of risk affecting the economy. It should be noted that \( V_M(T) \) does not depend on the contractual fee rate “f.” This is because this coefficient is multiplicative and the manager’s relative risk aversion is constant under log utility. That the endogenous fee parameter “f” is independent of the optimal strategy is what will make its solution relatively easy to get.

Along the way, we derive the dynamics of the corresponding optimal portfolio selected by the manager, i.e., the way the weights of the different assets composing the portfolio evolve through time.

Examples of superior information
The issue of informational heterogeneity among market participants is solved by using a sophisticated mathematical
Superior information and compensation fees of active mutual funds

The manager’s utility is infinite. The perturbation can, for example, be interpreted as the refinement of the information about market flows associated with a specific event.

In the third example, the manager’s additional information consists of the knowledge of the terminal value of some asset price(s) perturbed by an independent noise constant over time. This means that the insider has an initial advantage over direct investors (say a private information), but that this advantage does not increase over time. This example has been studied by Pikovsky and Karatzas (1996) and Amendinger et al. (1998). The quality of the superior information detained by the insider is an increasing function of a particular, crucial parameter, which we call “s.” This parameter thus measures the “magnitude” of the additional information (alternatively, the additional skill) the insider benefits from. For “s” positive but strictly smaller than one, the manager’s utility is finite, which is realistic. This model can be interpreted as follows. The additional expected utility (welfare) generated by the manager increases with the parameter “s.” Thus, “s” commands the manager’s capacity to generate abnormal returns (alpha):

- If “s” were set to zero, the manager would have the same skill in gathering and processing information as that of the direct investor.
- If “s” were set to one, the manager would have perfect forecasting power, and thus could reap arbitrage gains and achieve additional infinite utility.
- In the plausible situation where “s” is positive but strictly smaller than one, the manager has superior information (skills) but no perfect foresight. This will be the situation we assume below.

The last two examples are very similar, as they assume the knowledge of the terminal value of an asset (or portfolio) perturbed by some noise. Both are compatible with an “enhanced” price discovery process that may result from aggressive research policy or private information seeking and can lead to almost certain abnormal returns. This can be, for example, the case of some “small cap” stocks neglected by the bulk of financial analysts. Generally, both specifications are compatible with strategies such as event-driven arbitrage, risk arbitrage, stock split arbitrage, IPO (initial public offerings) arbitrage, etc.4

4 The term “arbitrage” appended by practitioners to these strategies is of course improper from an academic viewpoint. True (or pure) arbitrage is riskless in theory and is a free lunch.
Although the literature is mixed on whether funds can consistently generate abnormally high risk-adjusted returns, i.e., alpha, many studies have documented that some funds have created alpha by implementing these strategies (see, for example, Alexander et al. (2007) and Cremers and Petajisto (2009) for comprehensive analyses). According to another strand of research, (Lakonishok and Vermaelen (1990), and Ikenberry et al. (1995)), abnormal returns could be obtained from a trading rule by which one buys a stock on the day following the announcement of a self-tender offer or an open market stock repurchase.

Finally, one may ask where these three illustrations stand with respect to Fama’s (Fama (1970)) efficient market hypothesis (EMH). Empirical evidence suggests that capital markets are efficient in their weak and semi-strong forms most of (but not all) the time. All our cases are obvious violations of the strong form of EMH. The first example clearly violates even its weak form. The second case may be compatible with the semi-strong form, provided some relevant parameters are properly constrained. The last example is compatible with the semi-strong form of EMH, as it takes an insider to generate abnormal returns.

The investor’s reservation fee
Investors have free access to the complete financial market and can allocate their wealth directly across all available assets. Alternatively, they can delegate their wealth to the manager, whose behavior was described above, who will spare them the (implicit) cost of direct investment and make them benefit from their superior information in exchange for compensation fees.

We proceed as follows. We consider one particular investor and first solve their maximization program in the case of direct investment in the market. This classical problem has been studied in particular by Karatzas et al. (1991) and Cox and Huang (1989). We derive their optimal terminal wealth, \( W(T) \), thus their optimal expected utility, and the corresponding optimal portfolio process.

We then redo the whole optimization process under the assumption that the investor delegates their portfolio decisions to the manager. Note that their terminal wealth is now equal to the managed fund’s terminal value \( V_m(T) \) minus the manager’s compensation fees \( fV_m(T) \), i.e., \( (1 - f)V_m(T) \), where \( f \) is, as we recall, to be determined endogenously. The portfolio delegation contract is attractive to the customer if the expected utility of the wealth net of fees induced by the manager’s optimal trading strategy, \( (1 - f)V_m(T) \), is not smaller than the expected utility of the wealth \( W(T) \) derived from direct investment.

We consider the third case of superior information discussed above, with “s” positive but strictly smaller than one. We assume a typical fund manager enjoying a de facto monopoly and able to tap a very large pool of competitive investors. Consequently, the typical investor will accept to pay fees to the manager until the point where they are indifferent between investing directly in the market and delegating their portfolio decisions. The restriction to log utility allows for a direct computation of the investor’s “reservation” \( f^* \) parameter beyond which they will prefer direct investment. Moreover, and since the investor does not have access to the manager’s inside information, we assume that, for reputational reasons, the manager will not claim more than the reservation fees corresponding to their own alpha. In that sense, the manager is concerned by long-term customer relationship.

We are then in a position to express quantitatively the exact condition under which the investor benefits from the delegation of their portfolio to the informed (or skillful) manager and the exact value of \( f^* \), their reservation fee parameter.

The “alpha” generated by the investor directly investing in the market is nil by construction. The manager generates some “alpha” due to a specific extra information they gathered, or due to a more aggressive research effort or superior analytical and computational skills. The parameter \( f^* \) is an increasing function of this relative ability. In the next section, we use the relative capacity that follows from the third illustration of the section on examples of superior information.

Simulations
We assess the relevance of our main result regarding the value of the reservation fee rate \( f^* \) through Monte Carlo simulations. Our purpose is to show that the \( f^* \) we obtain lies within a reasonable range for plausible values of the parameters. We choose the following baseline values: \( V_m(0) = 100 \), the average return on

---

5 Dynamic aspects that would involve cash flows from and to the fund on the part of investors according to past performance are beyond the scope of this study. Our model can accommodate the case of oligopolistic managers by assuming that investors do not know ex-ante their relative skills and, therefore, choose them randomly and pay fees up to their reservation levels.
risky assets is 8%, the riskfree rate of interest is 3%, (i.e., the average risk premium is 5%), the average volatility of the risky assets is 15%, and the number of simulations is 100,000. In addition, the investment horizon T is either 1, 2, 3, or 5 years.

The manager’s additional information increases with the parameter “s” by definition. It also increases with the investment horizon T because, as uncertainty is larger for longer horizons, its relative reduction for a given level of “s” is also larger. Consequently, the optimal proportional fees also depend positively on “s” and T.

To provide some background, Goetzmann et al. (2003), for example, report that for a portfolio volatility of 15%, the annualized required excess return (alpha) needs to be 3% to 4% to justify a performance fee of 15% to 20%. For a volatility of 25%, the alpha required to justify the same performance fee ranges from 3.5% to 7.5%. Including the regular fees, the total percentage of wealth claimed by the hedge fund manager can be between 30% and 40%, as stated in the introduction. Investing in a hedge fund, therefore, would only appear rational if it provided a large, positive risk-adjusted return in compensation.

Simulation results are reported in Table 1 and displayed graphically in Figures 2 and 4 below. The manager’s alpha and the reservation fee rate f* are almost identical for small values of “s” (less than or equal to 15% for T up to 3 years and less than or equal to 10% for T equal to 5 years), which means that the manager reaps nearly all the benefit from a small or moderate information gap. As “s” increases, however, and, more importantly, the investment horizon T gets longer, the investor shares an increasing proportion of the manager’s informational advantage. Some large figures in Table 1 (which are illustrative only and not to be taken too literally) are not realistic for the mutual fund industry but can match the performance of the most successful hedge funds. It is thus apparent that picking the right fund manager is potentially extremely rewarding.

The corresponding annualized gross Sharpe ratios obtained by the manager before fees are reported in Table 2.6 For our given set of parameters, the annualized Sharpe ratio the direct investor can achieve is equal to 0.33 (= (0.08 – 0.03)/0.15) only. Our results vindicate, even for “s” equal to 10 or 15%, the potential gain associated with skillful managers.

We finally simulate the investor's expected utility (welfare) across different values of the information parameter “s” and various levels of the fee parameter f*. Numbers in bold appearing in the diagonal of Table 3 correspond to the expected utility of a delegated portfolio after paying the manager the reservation fee parameter f*. By definition, these values are all equal (to 4.64). For a given fee, the investor’s welfare is obviously an increasing function of “s”, and for a given “s”, it is evidently a decreasing function of f*.

Figure 1 generalizes Table 3 and displays the continuous relationship between the level of expected utility and the value of

---

### Table 1: Reservation proportional fees f*

<table>
<thead>
<tr>
<th></th>
<th>T = 1.0</th>
<th></th>
<th>T = 2.0</th>
<th></th>
<th>T = 3.0</th>
<th></th>
<th>T = 5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>alpha</td>
<td>f*</td>
<td>alpha</td>
<td>f*</td>
<td>alpha</td>
<td>f*</td>
<td>alpha</td>
</tr>
<tr>
<td>5%</td>
<td>0.14%</td>
<td>0.14%</td>
<td>0.57%</td>
<td>0.57%</td>
<td>1.28%</td>
<td>1.27%</td>
<td>3.53%</td>
</tr>
<tr>
<td>10%</td>
<td>0.67%</td>
<td>0.67%</td>
<td>2.64%</td>
<td>2.60%</td>
<td>5.86%</td>
<td>5.69%</td>
<td>15.88%</td>
</tr>
<tr>
<td>15%</td>
<td>1.76%</td>
<td>1.75%</td>
<td>6.83%</td>
<td>6.60%</td>
<td>14.90%</td>
<td>13.85%</td>
<td>39.05%</td>
</tr>
<tr>
<td>20%</td>
<td>3.68%</td>
<td>3.61%</td>
<td>13.84%</td>
<td>12.92%</td>
<td>29.38%</td>
<td>25.45%</td>
<td>73.27%</td>
</tr>
<tr>
<td>25%</td>
<td>6.74%</td>
<td>6.52%</td>
<td>24.30%</td>
<td>21.57%</td>
<td>49.70%</td>
<td>39.17%</td>
<td>116.60%</td>
</tr>
<tr>
<td>30%</td>
<td>11.35%</td>
<td>10.73%</td>
<td>38.64%</td>
<td>32.05%</td>
<td>75.51%</td>
<td>53.00%</td>
<td>165.35%</td>
</tr>
<tr>
<td>35%</td>
<td>17.93%</td>
<td>16.41%</td>
<td>56.90%</td>
<td>43.39%</td>
<td>105.60%</td>
<td>65.22%</td>
<td>215.43%</td>
</tr>
<tr>
<td>40%</td>
<td>26.89%</td>
<td>23.58%</td>
<td>78.65%</td>
<td>54.46%</td>
<td>138.26%</td>
<td>74.91%</td>
<td>263.41%</td>
</tr>
</tbody>
</table>

The table presents the simulated values of the manager’s alpha and the corresponding reservation fee parameter f* for different time horizons T and different values of the parameter “s” that increases the quality of the superior information.

---

6 The Sharpe Ratio is equal to (E(μ)/t)-r/σsqrt(t) where t is the number of investment years.
the information parameter “s”, for various levels of the optimal fees f*. The corresponding values of “s” leading to the level of welfare achieved by a direct investment (i.e., 4.64) are obtained at the intersection with the horizontal line that corresponds to the “No delegation” case. The relationship is generally convex, and the convexity is strong for “s” larger than 40%. This is because the weight of the risky assets in the optimal portfolio explodes for large values of the information parameter.

**Empirical evidence**

**Sample description**

Our data is downloaded from the CRSP Survivor-Bias-Free U.S. Mutual Fund Database, which provides open-ended mutual fund data from December 1961 for funds of all investment objectives, principally equity funds, taxable and municipal bond funds, international funds, and money market funds.

From CRSP we gather the following data related to the funds’ composition and trading activity: specialized primarily in equities or in debts, open to investors or not, retail fund or not, and institutional fund or not. We also collect for each fund the net monthly returns, computed as changes in NAV (Net Asset Value) including reinvested dividends from one period to the next, the amount of managed assets and the management fees calculated as a percentage of its assets.

We consider only funds for which the track record, as measured by net returns, and management fees are available. We also exclude funds for which the marginal compensation rate has been negative. This leaves us with 23,916 funds. We first report elements relative to the empirical distribution of these sampled funds in Figure 3 below. Fifty-five percent of the funds are equity funds, 36% are debt funds, and 9% invest in both equities and debt instruments.

Table 4a provides descriptive statistics for management fees by fund type. The mean and median marginal compensation rates are both 0.50%. They are slightly higher for equity funds (0.61% and 0.69%, respectively). Our results are in line with those of Coles et al. (2000), who find a mean (median) marginal compensation of 0.55% (0.50%). Deli (2002) finds comparable (though slightly higher) results with a mean (median) rate of 0.68% (0.65%). The funds’ durations are displayed in Table 4b. The mean (median) duration for all fund types is 12 (10) years. Fixed

![Figure 1: Expected utility of wealth](image-url)
Superior information and compensation fees of active mutual funds

Income funds have longer lives than equity and mixed funds, with a mean (median) of 13 (12) years.

Most of the funds in our sample are open-end funds as shown in Table 5a. Almost 76% are open to investment with little discrepancy across different fund types.

Table 5b reports that 26% of the funds studied in our sample are declared to be institutional funds. Table 5c shows that 43% of the 23,916 funds claim to be retail funds. For both institutional and retail, equity funds are the majority, with relatively few invested in mixed instruments.

Figure 2: Alpha
The figure plots the simulated values of the manager’s alpha for different values of the parameter “s” (from 0% to 40%) and the time horizon T (1 year, 2 years, 3 years and 5 years).

Figure 3: Descriptive statistics on all funds
Data analysis
From the 23,916 funds, we exclude those for which the management fees are negative (this may happen when the fund makes reimbursements) and those with an insufficient track record (less than one year of data). Our selected sample here then includes 23,869 funds.

For each fund, we use monthly net returns as stated above and compute its mean annualized net return over its duration. For a given duration, depending on the objective of the fund, we calculate the mean return of an equivalent investment in a closely related benchmark. For equity funds, the selected benchmark is the S&P 500. For fixed income funds, the benchmark is the 10-year treasury.

Figure 4: Reservation fees
The figure plots the simulated values of the reservation fee parameter $f^*$ for different values of the parameter “$s$” (from 0% to 40%) and the time horizon $T$ (1 year, 2 years, 3 years, and 5 years).

Figure 5: Percentage of funds doing better than our calculated benchmark
Superior information and compensation fees of active mutual funds

year T-note. And for mixed or alternative funds, the benchmark is taken as the one that performed best among the S&P 500 and the T-note, which is admittedly very harsh.

Following this approach, we show in Figure 5 above, that 57% over performed their assigned benchmark in the long run. For these funds, the explanation provided by our theoretical framework for the existence of portfolio delegation is plausible. The percentage of equity funds doing better than their benchmark is larger than that of the whole sample. The percentage is also higher for mixed (investing in both equity and fixed-income assets) funds, with roughly 74% of these outperforming their benchmark. On the other hand, we observe that the percentage of fixed income funds performing better than their benchmark is much lower than that of the whole sample (36.5%).

We also find that among funds identified as retail, 55% do better than their assigned benchmark, which is slightly lower than the percentage observed over the whole sample (57%). We also find

---

As we do not know the targeted mix of equities and debts of these funds, we cannot compute their benchmark as a weighted average.
that the percentage of institutional funds performing better than their benchmark is even higher (60%). Finally, we find that the percentage of outperforming open-end funds is slightly lower than that of the complete sample at 56%.

We then calculate for each type of fund the min, max, mean and median outperformance of the fund manager after and before management fees. Results are presented in Tables 6a and 6b, respectively. On average, the performance over the benchmark is quite substantial (300 basis points) and the management fee is roughly 50 basis points (the difference between the corresponding means or medians in Tables 6a and 6b). It should be kept in mind, however, that our illustrative sample was not corrected for the survivor bias, which tends to overestimate the funds’ actual performance.

Estimation of the “s” parameter

From the data gathered in Table 6c, we are able to calibrate empirical values for the “s” parameter, which represents in our theoretical setup the higher information possessed by the manager. If we retain the median “s” (18.0%), and the median alpha (2.74%, see Table 6b), we readily see from the comparison with Table 1, that this is compatible with a theoretical median investment horizon of roughly one year for the “average fund,” which seems realistic enough. It is rather remarkable that this still holds true when we segregate equity and fixed income funds. For the sub-sample of equity funds, the median “s” and alpha are 19.3% and 3.33%, respectively, and for the fixed income funds, the analogous figures are 14.4% and 1.61%. In both cases, this implies an investment horizon of almost one year.

Conclusion

The “filtration enlargement” technique allowed us to characterize the parameter of the compensation contract in terms of information differential. We thereby provided a new justification for the existence of portfolio delegation and a novel explanation for the manager’s alpha. Assuming that the manager and the investor have log utility, we have obtained quasi-explicit (unreported) solutions and characterized the investor’s reservation fee rate for a given information differential.

Simulations showed that our specification of the manager’s additional information (the knowledge of a terminal value perturbed by a constant noise) reasonably matches the industry practice in terms of management fee levels uncovered by our empirical study.

A possible extension of this work would be to consider utility
functions more general than the log to derive the investor's reservation fee parameter. However, solving the problem would be much more involved. Indeed, a fixed point problem arises in this context and, furthermore, it is impossible to disentangle the standard terms from terms due to the insider's change of probability in the optimal portfolio strategy. Another natural extension would be to assume a convex compensation scheme, a more realistic provision for most hedge funds. The (non-analytical) solution would be difficult to derive and simulations tricky to perform. The main thrust of our simpler approach, however, would not be hurt by these extensions.

References


Jeulin, J., and M. Yor, 1995, Gossissement de filtration: exemples et applications, Springer-Verlag
Mansuy, R., and M. Yor, 2005, Random times and (enlargement of) filtrations in a Brownian setting, Springer-Verlag, Berlin

Hedge fund contracts are often characterized by a particular compensation scheme that includes a ‘high-water mark’ provision: the manager receives, in addition to a given proportion of the fund’s net asset value (NAV), a compensation proportional to the NAV conditional on the latter being larger than some predetermined value.
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.

© 2013 EYGM Limited.
All Rights Reserved.
EYG No. EK0210

In line with EY’s commitment to minimize its impact on the environment, this document has been printed on paper with a high recycled content.

This material has been prepared for general informational purposes only and is not intended to be relied upon as accounting, tax, or other professional advice. Please refer to your advisors for specific advice.

ey.com

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

Accredited by the American Economic Association

ISSN 2049-8640