Andrew G Haldane: The $100 billion question

Comments by Mr Andrew G Haldane, Executive Director, Financial Stability, Bank of England, at the Institute of Regulation & Risk, Hong Kong, 30 March 2010.

* * *

I am grateful to Dele Adeleye, David Aikman, Marnoch Aston, Richard Davies, Colm Friel, Vaiva Katinaite, Sam Knott, Priya Kothari, Salina Ladha, Colin Miles, Rhiannon Sowerbutts and Aron Toth for comments and contributions.

The car industry is a pollutant. Exhaust fumes are a noxious by-product. Motoring benefits those producing and consuming car travel services – the private benefits of motoring. But it also endangers innocent bystanders within the wider community – the social costs of exhaust pollution.

Public policy has increasingly recognised the risks from car pollution. Historically, they have been tackled through a combination of taxation and, at times, prohibition. During this century, restrictions have been placed on poisonous emissions from cars – in others words, prohibition. This is recognition of the social costs of exhaust pollution. Initially, car producers were in uproar.

The banking industry is also a pollutant. Systemic risk is a noxious by-product. Banking benefits those producing and consuming financial services – the private benefits for bank employees, depositors, borrowers and investors. But it also risks endangering innocent bystanders within the wider economy – the social costs to the general public from banking crises.

Public policy has long-recognised the costs of systemic risk. They have been tackled through a combination of regulation and, at times, prohibition. Recently, a debate has begun on direct restrictions on some banking activities – in other words, prohibition. This is recognition of the social costs of systemic risk. Bankers are in uproar.

This paper examines the costs of banking pollution and the role of regulation and restrictions in tackling it. In light of the crisis, this is the $100 billion question. The last time such a debate was had in earnest followed the Great Depression. Evidence from then, from past crises and from other industries helps define the contours of today's debate. This debate is still in its infancy. While it would be premature to be reaching policy conclusions, it is not too early to begin sifting the evidence. What does it suggest?

Counting the systemic cost

One important dimension of the debate concerns the social costs of systemic risk. Determining the scale of these social costs provides a measure of the task ahead. It helps calibrate the intervention necessary to tackle systemic risk, whether through regulation or restrictions. So how big a pollutant is banking?

There is a large literature measuring the costs of past financial crises.[1] This is typically done by evaluating either the fiscal or the foregone output costs of crisis. On either measure, the costs of past financial crises appear to be large and long-lived, often in excess of 10% of pre-crisis GDP. What about the present crisis?

The narrowest fiscal interpretation of the cost of crisis would be given by the wealth transfer from the government to the banks as a result of the bailout. Plainly, there is a large degree of

[1] For example, Reinhart and Rogoff (2009).

uncertainty about the eventual loss governments may face. But in the US, this is currently estimated to be around $100 billion, or less than 1% of US GDP. For US taxpayers, these losses are (almost exactly) a $100 billion question. In the UK, the direct cost may be less than £20 billion, or little more than 1% of GDP.

Assuming a systemic crisis occurs every 20 years, recouping these costs from banks would not place an unbearable strain on their finances. The tax charge on US banks would be less than $5 billion per year, on UK banks less than £1 billion per year.[2] Total pre-tax profits earned by US and UK banks in 2009 alone were around $60 billion and £23 billion respectively.

But these direct fiscal costs are almost certainly an underestimate of the damage to the wider economy which has resulted from the crisis – the true social costs of crisis. World output in 2009 is expected to have been around 6.5% lower than its counterfactual path in the absence of crisis. In the UK, the equivalent output loss is around 10%. In money terms, that translates into output losses of $4 trillion and £140 billion respectively.

Moreover, some of these GDP losses are expected to persist. Evidence from past crises suggests that crisis-induced output losses are permanent, or at least persistent, in their impact on the level of output if not its growth rate.[3] If GDP losses are permanent, the present value cost of crisis will exceed significantly today's cost.

By way of illustration, Table 1 looks at the present value of output losses for the world and the UK assuming different fractions of the 2009 loss are permanent – 100%, 50% and 25%. It also assumes, somewhat arbitrarily, that future GDP is discounted at a rate of 5% per year and that trend GDP growth is 3%.[4] Present value losses are shown as a fraction of output in 2009.

As Table 1 shows, these losses are multiples of the static costs, lying anywhere between one and five times annual GDP. Put in money terms, that is an output loss equivalent to between $60 trillion and $200 trillion for the world economy and between £1.8 trillion and £7.4 trillion for the UK. As Nobel-prize winning physicist Richard Feynman observed, to call these numbers "astronomical" would be to do astronomy a disservice: there are only hundreds of billions of stars in the galaxy. "Economical" might be a better description.

It is clear that banks would not have deep enough pockets to foot this bill. Assuming that a crisis occurs every 20 years, the systemic levy needed to recoup these crisis costs would be in excess of $1.5 trillion per year. The total market capitalisation of the largest global banks is currently only around $1.2 trillion. Fully internalising the output costs of financial crises would risk putting banks on the same trajectory as the dinosaurs, with the levy playing the role of the meteorite.

It could plausibly be argued that these output costs are a significant over-statement of the damage inflicted on the wider economy by the banks. Others are certainly not blameless for the crisis. For every reckless lender there is likely to be a feckless borrower. If a systemic tax is to be levied, a more precise measure may be needed of banks' distinctive contribution to systemic risk.

One such measure is provided by the (often implicit) fiscal subsidy provided to banks by the state to safeguard stability. Those implicit subsidies are easier to describe than measure. But one particularly simple proxy is provided by the rating agencies, a number of whom provide

---

[2] The levy on US banks announced by the US government in January takes the $100 billion loss and recoups it over 10 years rather than 20.

[3] IMF (2009).

[4] The results are plainly sensitive to the choice of discount rate and trend growth rate. Other things equal, the higher the discount rate and the lower the trend growth rate, the smaller the losses.

both “support” and “standalone” credit ratings for the banks. The difference in these ratings encompasses the agencies’ judgement of the expected government support to banks.

Table 2 looks at this average ratings difference for a sample of banks and building societies in the UK, and among a sample of global banks, between 2007 and 2009. Two features are striking. First, standalone ratings are materially below support ratings, by between 1.5 and 4 notches over the sample for UK and global banks. In other words, rating agencies explicitly factor in material government support to banks.

Second, this ratings difference has increased over the sample, averaging over one notch in 2007 but over three notches by 2009. In other words, actions by government during the crisis have increased the value of government support to the banks. This should come as no surprise, given the scale of intervention. Indeed, there is evidence of an up-only escalator of state support to banks dating back over the past century.[5]

Table 3 takes the same data and divides the sample of UK banks and building societies into “large” and “small” institutions. Unsurprisingly, the average rating difference is consistently higher for large than for small banks. The average ratings difference for large banks is up to 5 notches, for small banks up to 3 notches. This is pretty tangible evidence of a second recurring phenomenon in the financial system – the “too big to fail” problem.

It is possible to go one step further and translate these average ratings differences into a monetary measure of the implied fiscal subsidy to banks. This is done by mapping from ratings to the yields paid on banks’ bonds;[6] and by then scaling the yield difference by the value of each banks’ ratings-sensitive liabilities.[7] The resulting money amount is an estimate of the reduction in banks’ funding costs which arises from the perceived government subsidy.

Table 4 shows the estimated value of that subsidy for the same sample of UK and global banks, again between 2007 and 2009. For UK banks, the average annual subsidy for the top five banks over these years was over £50 billion – roughly equal to UK banks’ annual profits prior to the crisis. At the height of the crisis, the subsidy was larger still. For the sample of global banks, the average annual subsidy for the top five banks was just less than $60 billion per year. These are not small sums.

Table 4 also splits UK banks and building societies into “Big 5”, “medium” and “small” buckets. As might be expected, the large banks account for over 90% of the total implied subsidy. On these metrics, the too-big-to-fail problem results in a real and on-going cost to the taxpayer and a real and on-going windfall for the banks. If it were ever possible to mint a coin big enough, these would be the two sides of it.

These results are no more than illustrative – for example, they make no allowance for subsidies arising on retail deposits. Nonetheless, studies using different methods have found similarly-sized subsidies. For example, Baker and McArthur ask whether there is a difference in funding costs for US banks either side of the $100 billion asset threshold – another $100 billion question.[8] They find a significant wedge in costs, which has widened during the crisis. They calculate an annual subsidy for the 18 largest US banks of over $34 billion per year. Applying the same method in the UK would give an annual subsidy for the five largest banks of around £30 billion.

This evidence can provide only a rough guide to systemic scale and cost. But the qualitative picture it paints is clear and consistent. First, measures of the costs of crisis, or the implicit

---

5 Haldane (2009a).

6 Using the end-year yield on the financial corporates bond index across the ratings spectrum.

7 For example, banks’ retail deposits are excluded but unsecured wholesale borrowing is included.

8 Baker and McArthur (2009).

subsidy from the state, suggest banking pollution is a real and large social problem. Second, those entities perceived to be "too big to fail" appear to account for the lion's share of this risk pollution. The public policy question, then, is how best to tackle these twin evils.

Taxation and prohibition

To date, the public policy response has largely focussed on the role of prudential regulation in tackling these problems. Higher buffers of capital and liquid assets are being discussed to address the first problem. And add-ons to these capital and liquidity buffers for institutions posing the greatest systemic risk are being discussed to address the second.[9] In essence, this is a *taxation* solution to the systemic risk pollution problem.[10]

There is a second approach. On 21 January 2010, US President Barack Obama proposed placing formal restrictions on the business activities and scale of US banks. Others have made complementary proposals for structural reform of banking.[11] Typically, these involve separation of bank activities, either across business lines or geographies. In essence, this is the *prohibition* solution to the systemic pollution problem.

This sets the scene for a great debate. It is not a new one. The taxation versus prohibition question crops up repeatedly in public choice economics. For centuries it has been central to the international trade debate on the use of quotas versus subsidies. During this century, it has become central to the debate on appropriate policies to curtail carbon emissions.[12]

In making these choices, economists have often drawn on Martin Weitzman's classic public goods framework from the early 1970s.[13] Under this framework, the optimal amount of pollution control is found by equating the marginal social benefits of pollution-control and the marginal private costs of this control. With no uncertainty about either costs or benefits, a policymaker would be indifferent between taxation and restrictions when striking this cost/benefit balance.

In the real world, there is considerable uncertainty about both costs and benefits. Weitzman's framework tells us how to choose between pollution-control instruments in this setting. If the marginal social benefits foregone of the wrong choice are large, relative to the private costs incurred, then quantitative restrictions are optimal. Why? Because fixing quantities to achieve pollution control, while letting prices vary, does not have large private costs. When the marginal social benefit curve is steeper than the marginal private cost curve, restrictions dominate.

The results flip when the marginal cost/benefit trade-offs are reversed. If the private costs of the wrong choice are high, relative to the social benefits foregone, fixing these costs through taxation is likely to deliver the better welfare outcome. When the marginal social benefit curve is flatter than the marginal private cost curve, taxation dominates. So the choice of taxation versus prohibition in controlling pollution is ultimately an empirical issue.

To illustrate the framework, consider the path of financial regulation in the US over the past century. The US announcements in January are in many respects redolent of US financial reforms enacted during the late 1920s and early 1930s. Then, restrictions were imposed on both bank size and scope, in the form of the McFadden (1927) and Glass-Steagall (1933) Acts. The history of both, viewed through Weitzman's lens, is illuminating for today's debate.

---

9 Basel Committee on Banking Supervision (2009).

10 For example, Brunnermeier *et al* (2009), NYU Stern School of Business (2009).

11 For example, Kay (2009), Kotlikoff (2010).

12 Stern (2006).

13 Weitzman (1974).

The McFadden Act (1927) in the US gave nationally-chartered banks broadly the same branching rights as state banks within the state. But it also confirmed the effective prohibition on national banks opening new branches across state lines that had previously been implicit in the US National Banking Act (1864). It covered a wide range of banking functions, including deposit-taking and brokerage.

The motivation behind the Act appears to have been in part political, reflecting lobbying by small unit banks under threat from larger competitors. But it also had an economic dimension, as a check on the dangers of "excessive concentration of financial power".[14] The same too-big-to-fail arguments are of course heard today, though the concerns then were competition rather than crisis-related ones. Weitzman's marginal social benefit curve was perceived to be steep, made so by state-level competition concerns.

McFadden appeared to be fairly effective in limiting the size of US banks from the 1930s right through to the mid-1970s. Over this period, the average asset size of US banks in relation to nominal GDP was roughly flat (Chart 1). As recently as the early 1980s, it was still at around its level at the time of the Great Depression.

The 1980s marked a watershed, with interstate branching restrictions progressively lifted. States began to open their borders to out-of-state Bank Holding Companies (BHCs). The 1982 Garn-St Germain Act allowed any BHC to acquire failed banks and thrifts, regardless of the state law. Finally, the Riegle-Neal Act of 1994, which took effect in 1997, largely lifted restrictions on interstate branching for both domestic BHCs and foreign banks.

The rationale for this change of heart was a mirror-image of the 1920s. Large banks convinced politicians of the high private costs of restrictions, which inhibited the efficiency of their offering to the public. In Weitzman's framework, private costs trumped social benefits. The effects of the removal of interstate restrictions were dramatic. The average size of US banks, relative to GDP, has risen roughly threefold over the past 20 years (Chart 1). Too-big-to-fail was reborn in a new guise.

The US Banking Act (1933) was co-sponsored by Senator Carter Glass and Representative Henry Steagall – hence "Glass-Steagall". It prevented commercial banks from conducting most types of securities business, including principal trading, underwriting and securities lending. It also banned investment banks from taking deposits. The key functions of commercial and investment banking were effectively prised apart.

The Act was motivated by stability concerns in the light of the Great Depression. The stock market boom of the 1920s had been fuelled by cheap credit from the banks. The stock market crash of 1929 brought that, and a great many US banks, to a shuddering halt. Among many banks, net losses on securities were as great as losses on loans. These losses transmitted to the real economy through a collapse in lending, whose stock halved between 1929 and 1933.

Against this economic backdrop, and amid heated banker-bashing, it is easy to see how the social benefits of segregation were perceived as far outweighing the private costs at the time. Kennedy (1973) describes how "Stock dealings which had made bankers rich and respected in the era of affluence now glared as scarlet sins in the age of depression. Disillusionment with speculators and securities merchants carried over from investment bankers to commercial bankers; the two were often the same, and an embittered public did not care to make fine distinctions". Glass and Steagall made just such a distinction. They underpinned it with legislation, signed by President Roosevelt in June 1933.

As with McFadden, Glass-Steagall appears to have been effective from the 1930s right up until the latter part of the 1980s. Measures of concentration in the US banking system

[14] Chapman and Westerfield (1942).

remained broadly flat between the 1930s and the late 1980s (Chart 2). But competitive pressures were building from the late 1970s onwards. Strains on US commercial banks intensified from alternative lending vehicles (such as mutual funds and commercial paper markets) and from overseas banks. The private costs of restrictions were rising.

Legislators responded. After 1988, securities affiliates within BHCs were permitted, though were still subject to strict limits. In 1999, the Gramm-Leach-Bliley Act revoked the restrictions of Glass-Steagall, allowing co-mingling of investment and commercial banking. This came as a specific response to the perceived high private costs of restrictions relative to the perceived social benefits – again, in a reversal of the Weitzman calculus from the early 1930s.

As with size, the effects of liberalisation on banking concentration were immediate and dramatic. The share of the top three largest US banks in total assets rose fourfold, from 10% to 40% between 1990 and 2007 (Chart 2). A similar trend is discernible internationally: the share of the top five largest global banks in the assets of the largest 1000 banks has risen from around 8% in 1998 to double that in 2009.

This degree of concentration, combined with the large size of the banking industry relative to GDP, has produced a pattern which is not mirrored in other industries. The largest banking firms are far larger, and have grown far faster, than the largest firms in other industries (Chart 3). With the repeal of the McFadden and Glass-Steagall Acts, the too-big-to-fail problem has not just returned but flourished.

In the light of the Great Recession, and the large apparent costs of too-big-to-fail, does Weitzman's cost-benefit calculus suggest there is a case for winding back the clock to the reforms of the Great Depression? Determining that requires an assessment of the benefits and costs of restrictions.

The benefits of prohibition

The potential benefits of restricting activity in any complex adaptive system, whether financial or non-financial, can roughly be grouped under three headings: modularity, robustness and incentives. Each has a potentially important bearing on systemic resilience and hence on the social benefits of restrictions.

(a) Modularity

In 1973, Nobel-prizing winning economist Robert Merton showed that the value of a portfolio of options is at least as great as the value of an option on the portfolio.[15] On the face of it, this seems to fly in the face of modern portfolio theory, of which Merton himself was of course one of the key architects. Whatever happened to the benefits of portfolio diversification?

The answer can be found in an unlikely source – Al'Qaeda. Although the precise organisational form of Al'Qaeda is not known with certainty, two structural characteristics are clear. First, it operates not as a centralised, integrated organisation but rather as a highly decentralised and loose network of small terrorist cells. Second, as events have shown, Al'Qaeda has exhibited considerable systemic resilience in the face of repeated and on-going attempts to bring about its collapse.

These two characteristics are closely connected. A series of decentralised cells, loosely bonded, make infiltration of the entire Al'Qaeda network extremely unlikely. If any one cell is incapacitated, the likelihood of this undermining the operations of other cells is severely reduced. That, of course, is precisely why Al'Qaeda has chosen this organisational form.

[15] Merton (1973).

Al'Qaeda is a prime example of modularity and its effects in strengthening systemic resilience.

There are many examples from other industries where modularity in organisational structure has been deployed to enhance systemic resilience. Computer manufacture is one. During the late 1960s, computers were highly integrated systems. Gradually, they evolved into the quintessential modular system of today, with distinct modules (CPU, hard disk, keyboard) which were replaceable if they failed without endangering the functioning of the system as a whole. This improved resilience and reliability.

In the computing industry, modularity appears to have had an influence on industry structure. Since the 1970s, the computer hardware industry has moved from a highly concentrated structure to a much more fragmented one. In 1969, IBM had a market share of over 70%. By this century, the market share of the largest hardware firm was around a third of that. Modularity has meant the computer industry has become less prone to "too-big-to-fail" problems.

Other examples of modularity in organisational structures include:

- The management of forest fires, which typically involves the introduction of firebreaks to control the spread of fire;[16]
- The management of utility services, such as water, gas and electricity, where the network often has built-in latencies and restrictions to avoid overload and contagion;
- The management of infectious diseases which these days often involves placing restrictions on travel, either within a country (as in the case of foot-and-mouth disease in the UK) or outside of it (as in the case of H5N1);[17]
- The control of computer viruses across the world wide web, which is typically achieved by constructing firewalls which restrict access to local domains;
- Attempts on the world domino toppling record, which involve arranging the dominos in discrete blocks to minimise the risk of premature cascades.

These are all examples where modular structures have been introduced to strengthen system resilience. In all of these cases, policy intervention was required to affect this change in structure. The case for doing so was particularly strong when the risk of viral spread was acute. In some cases, intervention followed specific instances of systemic collapse.

The North American electricity outage in August 2003 affected 55 million people in the US and Canada. It had numerous adverse knock-on effects, including to the sewage system, telephone and transport network and fuel supplies. A number of people are believed to have died as a consequence. This event led to a rethinking of the configuration of the North American electricity grid, with built-in latencies and stricter controls on power circulation.

In the mid-1980s, an attempt on the world domino-toppling record – at that time, 8000 dominos – had to be abandoned when the pen from one of the TV film crew caused the majority of the dominos to cascade prematurely. Twenty years later a sparrow disturbed an attempt on the world domino-toppling record. Although the sparrow toppled 23,000 dominos, 750 built-in gaps averted systemic disaster and a new world record of over 4 million dominos was still set. No-one died, except the poor sparrow which (poetically if controversially) was shot by bow and arrow.

So to banking. It has many of the same basic ingredients as other network industries, in particular the potential for viral spread and periodic systemic collapse. For financial firms

[16] Carlson and Doyle (1999).

[17] Kelling *et al* (2003).

holding asset portfolios, however, there is an additional dimension. This can be seen in the relationship between *diversification* on the one hand and *diversity* on the other.[18] The two have quite different implications for resilience.

In principle, size and scope increase the diversification benefits. Larger portfolios ought to make banks less prone to idiosyncratic risk to their asset portfolio. In the limit, banks can completely eradicate idiosyncratic risk by holding the market portfolio. The "only" risk they would face is aggregate or systematic risk.

But if all banks are fully diversified and hold the market portfolio, that means they are all, in effect, holding the same portfolio. All are subject to the same systematic risk factors. In other words, the system as a whole lacks diversity. Other things equal, it is then prone to generalised, systemic collapse. Homogeneity breeds fragility. In Merton's framework, the option to default selectively through modular holdings, rather than comprehensively through the market portfolio, has value to investors.

The precise balance between diversification and diversity depends on banks' balance sheet configuration. What does this suggest? Charts 4 and 5 plot the income variability of a set of 24 global banks against their asset size and a measure of the diversity of their business model.[19] There is no strong relationship between either size or diversity and income volatility. If anything the relationship is positively sloped, with size and diversity increasing income variability, not smoothing it.

Charts 6 and 7 look at banks' experience during the crisis. Size and diversity are plotted against banks' write-downs (per unit of assets). Again, if anything, these relationships are positively sloped, with larger, more diversified banks suffering proportionally greater losses. This is consistent with evidence from econometric studies of banking conglomerates which has found that larger banks, if anything, exhibit greater risk due to higher volatility assets and activities.[20]

This evidence is no more than illustrative. But it suggests that, in the arm wrestle between diversification and diversity, the latter appears to have held the upper hand. Bigger and broader banking does not obviously appear to have been better, at least in a risk sense. In banking, as on many things, Merton may have had it right.

(b) Robustness

The Merton result holds in a world in which investors form judgements based on knowledge of the distribution of risk. But in complex dynamic systems, the distribution of risk may be lumpy and non-linear, subject to tipping points and discontinuities.[21] Faced with this, the distribution of outcomes for the financial system as a whole may well be incalculable. The financial system may operate in an environment of uncertainty, in the Knightian sense, as distinct from risk.

There is a literature on how best to regulate systems in the face of such Knightian uncertainty.[22] It suggests some guideposts for regulation of financial systems. First, keep it simple. Complex control of a complex system is a recipe for confusion at best, catastrophe at

---

[18] Beale *et al* (2009).

[19] A Herfindahl-Hirschman index of revenue concentration is constructed to measure diversification, with a measure of zero meaning that the HHI = 1, i.e. revenue is concentrated solely on one activity. Revenue concentration is calculated across three buckets for the last pre crisis year (2006) – Retail and commercial banking; Corporate and investment banking; Asset and wealth management.

[20] De Nicolo (2000).

[21] Haldane (2009b).

[22] See, for example, Aikman *et al* (2010).

worst. Complex control adds, not subtracts, from the Knightian uncertainty problem. The US constitution is four pages long. The recently-tabled Dodd Bill on US financial sector reform is 1,336 pages long. Which do you imagine will have the more lasting impact on behaviour.

Second, faced with uncertainty, the best approach is often to choose a strategy which avoids the extreme tails of the distribution. Technically, economists call this a "minimax" strategy – minimising the likelihood of the worst outcome. Paranoia can sometimes be an optimal strategy. This is a principle which engineers took to heart a generation ago. It is especially evident in the aeronautical industry where air and space disasters acted as beacons for minimax redesign of aircraft and spaceships.

Third, simple, loss-minimising strategies are often best achieved through what economists call "mechanism design" and what non-economists call "structural reform". In essence, this means acting on the underlying organisational form of the system, rather than through the participants operating within it. In the words of economist John Kay, it is about regulating structure not behaviour.[23]

Taken together, these three features define a "robust" regulatory regime – robust to uncertainties from within and outside the system. Using these robustness criteria, it is possible to assess whether restrictions might be preferable to taxation in tackling banking pollution. To illustrate this, contrast the regulatory experience of Glass-Steagall (a restrictions approach) and Basel II (a taxation approach).

Glass-Steagall was simple in its objectives and execution. The Act itself was only 17 pages long. Its aims were shaped by an extreme tail event (the Great Depression) and were explicitly minimax (to avoid a repetition). It sought to achieve this by acting directly on the structure of the financial system, quarantining commercial bank and brokering activities through red-line regulation. In other words, Glass-Steagall satisfied all three robustness criteria. And so it proved, lasting well over half a century without a significant systemic event in the US.

The contrast with Basel II is striking. This was anything but simple, comprising many thousands of pages and taking 15 years to deliver. It was calibrated largely to data drawn from the Great Moderation, a period characterised by an absence of tail events – more minimin than minimax. Basel II was underpinned by a complex menu of capital risk weights. This was fine-line, not red-line, regulation. In short, Basel II satisfied few of the robustness criteria. And so it proved, overwhelmed by the recent crisis scarcely after it had been introduced.

(c) Incentives

Tail risk within some systems is determined by God – in economist-speak, it is exogenous. Natural disasters, like earthquakes and floods, are examples of such tail risk. Although exogenous, even these events have been shown to occur more frequently than a normal distribution would imply.[24] God's distribution has fat tails.

Tail risk within financial systems is not determined by God but by man; it is not exogenous but endogenous. This has important implications for regulatory control. Finance theory tells us that risk brings return. So there are natural incentives within the financial system to generate tail risk and to avoid regulatory control. In the run-up to this crisis, examples of such risk-hunting and regulatory arbitrage were legion. They included escalating leverage, increased trading portfolios and the design of tail-heavy financial instruments.[25]

---

[23] Kay (2009).

[24] Korup and Clague (2009).

[25] Haldane (2009a) discusses some of these strategies in greater detail and the payoffs they generate.

The endogeneity of tail risk in banking poses a dilemma for regulation. Putting uncertainties to one side, assume the policymaker could calibrate perfectly tail risk in the system today and the capital necessary to insure against it. In an echo of the 1979 Madness song, banks would then have incentives to position themselves "One Step Beyond" the regulatory buffer to harvest the higher returns that come from assuming tail risk. They do so safe in the knowledge that the state will assume some of this risk if it materialises. Tail risk would expand to exhaust available resources. Countless crises have testified to this dynamic.

This dynamic means it is hazardous to believe there is a magic number for regulatory ratios sufficient to insure against tail risk in all states of the world. Because tail risk is created not endowed, calibrating a capital ratio for all seasons is likely to be, quite literally, pointless – whatever today's optimal regulatory point, risk incentives mean that tomorrow's is sure to be different.

In response, some economists have proposed corner solutions to the systemic risk problem – in effect, radical structural redesign. Starting with Irving Fisher in the 1930s, some have proposed narrow banks with a 100% liquid asset ratio to protect the liquidity services banks provide.[26] Others have proposed mutual fund banks with a 100% equity ratio to safeguard banks' solvency.[27] These limiting solutions are proof to risk incentives. The one guaranteed safe hiding place for the risk-fearing policymaker is the corner.

One criticism of these proposals is that they might raise materially the cost of capital to banks and hence to the real economy. For example, 100% capital ratios could cause the economy-wide cost of capital to sky-rocket given the premium charged for equity over debt finance. This same argument is frequently heard in debates about more modest rises in banks' capital ratios. But there are good counter-arguments that need also to be weighed.

By lowering risk, higher levels of equity ought to lower banks' cost of debt finance. Indeed, in a frictionless world Modigliani and Miller famously showed that this effect would fully offset the higher cost of equity, thereby leaving the total cost of capital for banks unchanged.[28] In other words, the cost of capital for a bank may be unaffected by its capital structure, at least when distortions in the economy are small. Even when they large, some offset in debt costs is likely.

It is possible to go one step further and argue that higher bank capital ratios could potentially *lower* banks' cost of capital. The size of the premium demanded by holders of equity is a long-standing puzzle in finance – the equity premium puzzle.[29] Robert Barro has suggested this puzzle can be explained by fears of extreme tail events.[30] And what historically has been the single biggest cause of those tail events? Banking crises. Boosting banks' capital would lessen the incidence of crises. If this lowered the equity premium, as Barro suggests, the cost of capital in the economy could actually fall.

The costs of prohibition

Turning to the other side of the equation, what does existing evidence tell us about the costs to banks of restrictions, whether on the scale or scope of their activities? In Weitzman's framework, how significant are the private costs of restrictions? Fortunately, there is a reasonably rich empirical literature on economies of scale and scope in banking.

---

[26] Kay (op.cit.).

[27] Kotlikoff (2010).

[28] Modigliani and Miller (1958). See also Miles (2009).

[29] Mehra and Prescott (1985).

[30] Barro (2006).

(a) Economies of scale

On economies of *scale*, the literature tends either to look at the cross-sectional efficiency of banks of different sizes, or the time-series efficiency of banks either side of a merger. As it turns out, both roads reach the same destination. Economies of scale appear to operate among banks with assets less, perhaps much less, than $100 billion. But above that threshold there is evidence, if anything, of *dis*economies of scale. The Weitzman marginal private cost curve is U-shaped.[31]

Experience in the US following the McFadden Act suggests size in banking can bring benefits. Over 9,000 US banks failed during the Great Depression, the majority of which were unit banks. Friedman and Schwarz (1963) blame the absence of bank branching for the high failure rate among US banks. The costs of limited branching were also felt well after the Great Depression in higher-cost provision of banking services, in particular for larger companies using lending syndicates to finance large-scale investment.[32]

US experience after McFadden chimes with cross-country evidence drawn, in particular, from developing countries. For example, using a dataset of 107 countries, Barth *et al* (2004) assess the effects of restrictions on the efficiency and stability of the financial system. They find evidence that restrictions are damaging to both, in particular barriers to foreign bank entry.

Do those arguments resonate within advanced country banking systems today? Two comprehensive studies in the mid-1990s found that economies of scale in banking are exhausted at relatively modest levels of assets, perhaps between $5–10 billion.[33] A more recent 2004 survey of studies in both the US and Europe finds evidence of a similar asset threshold.[34] Even once allowance is made for subsequent balance sheet inflation, this evidence implies that economies of scale in banking may cease at double-digit dollar billions of assets.

Evidence from banking mergers offers little more encouragement. There is no strong evidence of increased bank efficiency after a merger or acquisition.[35] And there is little to suggest cross-activity mergers create economic value.[36] That rather chimes with recent crisis experience. Of the bank mergers and acquisitions which have taken place recently, the majority have resulted in the merged firm under-performing the market in the subsequent period. Of course, all econometric studies have their limitations so these results do not close the case. Nonetheless, the uniformity of the evidence is striking.

(b) Economies of scope

Turning from economies of scale to economies of *scope*, the picture painted is little different. Evidence from US bank holding companies suggests that diversification gains from multiple business lines may be more than counter-balanced by heightened exposures to volatile income-generating activities, such as trading.[37] This mirrors the evidence from Charts 4 and 5 and from the Great Depression. Internationally, a recent study of over 800 banks in

---

[31] Santomero and Eckles (2000).

[32] For example, Calomiris and Hubbard (1995).

[33] Saunders (1996), Berger and Mester (1997).

[34] Amel *et al* (2004).

[35] For example, Berger and Humphrey (1997) based on a survey of over 100 studies.

[36] For example, De Long (2001).

[37] Stiroh and Rumble (2006).

43 countries found a conglomerate "discount" in their equity prices.[38] In other words, the market assigned a lower value to the conglomerate than the sum of its parts, echoing Merton's 1973 insight. This is evidence of diseconomies of scope in banking.

On the face of it, these findings are a puzzle. The most likely cause was articulated by Austin Robinson back in the 1930s – "Man's mind and man's memory is essentially a limited factor…Every increase in size beyond a point must involve a lengthening of the chain of authority…at some point the increasing costs of co-ordination must exceed the declining economies".[39] Oliver Williamson's "span of control" theory of organisations made rigorous this intuition thirty years later.

The essence of these arguments is that limits on the optimal size and scope of firms may be as much neurological as technological. Numbers of synapses may matter more than numbers of servers. The history of military units provides a good illustration. In Roman times, the optimal size of a military unit was 100 – hence the Roman centurion. This was the maximum number of men a general felt able to know well enough to lead and control. The constraint was neurological.

Two millennia have passed. Extraordinary advances have been made in military telecommunications technology. And the optimal size of the military unit in the US army today? Just under 100 people.[40] The number of relationships humans are felt able to maintain is believed to lie below 150 – so-called Dunbar's Law.[41] For most of us, it is single-digits. That number has been roughly the same since the dawn of time, despite the extra-ordinary recent advance of technology and social networks. As Nicholas Christakis has observed, Facebook "friends" are not really your friends.

With hindsight, this crisis has provided many examples of failures rooted in an exaggerated sense of knowledge and control. Risks and counterparty relationships outstripped banks' ability to manage them. Servers outpaced synapses. Large banks grew to comprise several thousand distinct legal entities. When Lehman Brothers failed, it had almost one million open derivatives contracts – the financial equivalent of Facebook friends. Whatever the technology budget, it is questionable whether any man's mind or memory could cope with such complexity.

To sum up, the maximum efficient scale of banking could be relatively modest. Perhaps it lies below $100 billion. Experience suggests there is at least a possibility of diseconomies of scale lying in wait beyond that point. Conglomerate banking, while good on paper, appears to be more mixed in practice. If these are not inconvenient truths, they are at least sobering conjectures. They also sit awkwardly with the current configuration of banking.

In 2008, 145 banks globally had assets above $100 billion, most of them universal banks combining multiple business activities. Together, these institutions account for 85% of the assets of the world's top 1000 banks ranked by Tier 1 capital. If these institutions could be resolved easily, so that the systemic consequences of their failure were limited, efficiency considerations could perhaps be set to one side. Or, put in Weitzman's terms, the social benefits of cutting banks down to size would be low.

But crisis experience has demonstrated that the apparatus does not currently exist to resolve safely these institutions. There are no examples during this crisis of financial institutions

---

[38] Laeven and Levine (2007); see also Schmid and Walter (2009) for recent US evidence.

[39] Robinson (1934).

[40] Christakis and Fowler (2009).

[41] Dunbar (1993).

beyond $100 billion being resolved without serious systemic spillovers.[42] Instead, those in trouble have been bailed-out. The same 145 institutions account for over 90% of the support offered by governments during the course of the crisis.

In the light of the crisis, and in the language of Weitzman, the marginal social benefits of restrictions could be greater than the marginal private costs. The maximum efficient scale of banking may lie below the maximum resolvable scale. A large part of the effort of the international community over the past few years has been directed at increasing the maximum resolvable scale of banks – for example, through improved resolution regimes and living wills.[43] If successful, that effort would shift the balance of the Weitzman cost/benefit calculus in the direction of bigger banks; it could help achieve the modularity, robustness and better aligned incentives which restrictions otherwise deliver.

But if this effort is unsuccessful, past evidence and present experience pose a big question about existing banking structures. Against that backdrop, it is understandable that restrictions on scale and activity are part of today's debate about solutions to the systemic pollution problem. $100 billion may not just be the question; it may also be part of the answer.

Conclusion

We are at the start of a great debate on the future structure of finance, not the end. Some fear that momentum for radical financial reform will be lost. But financial crises leave a scar. This time's sovereign scar should act as a lasting reminder of the criticality of reform. Today's crisis has stretched some state's sinews to the limit. Both literally and metaphorically, global finance cannot afford another.

The history of banking is that risk expands to exhaust available resources. Tail risk is bigger in banking because it is created, not endowed. For that reason, it is possible that no amount of capital or liquidity may ever be quite enough. Profit incentives may place risk one step beyond regulation. That means banking reform may need to look beyond regulation to the underlying structure of finance if we are not to risk another sparrow toppling the dominos.

Today's financial structure is dense and complex, like a tropical rainforest. Like the rainforests, when it works well it is a source of richness. Yet it is, as events have shown, at the same time fragile. Simpler financial eco-systems offer the promise of greater robustness, at some cost in richness. In the light of a costly financial crisis, both eco-systems should be explored in seeking answers to the $100 billion question.

---

[42] Washington Mutual, with assets of around $300bn, was resolved by FDIC, but was perceived by many to have caused systemic spillovers.

[43] For example, Tucker (2010).

References

Aikman, D, Barrett, P, Kapadia, S, King, M, Proudman, J, Taylor, T, de Weymarn, I, and T Yates (2010), *Uncertainty in Macroeconomic Policy Making: Art or Science?*, available at http://www.bankofengland.co.uk/publications/speeches/2010/speech432.pdf

Amel, D, Barnes, C, Panetta, F and C Salleo (2004), "Consolidation and Efficiency in the Financial Sector: A Review of the International Evidence", *Journal of Banking & Finance, vol. 28(10).*

Baker, D and T McArthur (2009), "The Value of the 'Too Big to Fail' Big bank Subsidy", *Centre for Economic and Policy Research.*

Barro, R J (2006), "Rare Disasters and Asset Markets in the Twentieth Century", *Quarterly Journal of Economics, vol. 121(3), p.823–866*.

Barth, J R, Caprio Jr., and R Levine (2004), "Bank Supervision and Regulation: What Works Best?", *Journal of Financial Intermediation, vol. 13(2), p.205–248*.

Basel Committee on Banking Supervision (2009), *Consultative Document: Strengthening the Resilience of the Banking Sector*.

Beale, N, Rand, D, Arinaminpathy, N and R M May (2009), *Conflicts Between Individual and Systemic Risk in Banking and Other Systems*, forthcoming.

Berger, A and D Humphrey (1997)*,* "Efficiency of financial institutions: international survey and directions for future research", *European Journal of Operational Research, vol. 98, p.175–212*.

Berger, A N and L J Mester (1997), "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?", *Journal of Banking & Finance, vol. 21(7), p.895–947*.

Brunnermeier, M, Crockett, A, Goodhart, C, Persaud, A and H Shin (2009), "The Fundamental Principles of Financial Regulation", *ICMB-CEPR Geneva Report on the World Economy 11*.

Calomiris, C and G Hubbard (1995), "Tax Policy, Internal Finance, and Investment: Evidence from the Undistributed Profits Tax of 1936–1937", *Journal of Business, vol. 68, p.443–482*.

Carlson, J M and J Doyle (1999), "Highly Optimized Tolerance: A Mechanism for Power Laws in Designed Systems", *Physical Review E, vol. 60(2), p.1412–1427.*

Chapman, J M and R B Westerfield (1942), *Branch Banking*, Harper & Brothers.

Christakis, N A and J H Fowler (2009), *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*, Little Brown and Company.

DeLong, G L (2001), "Stockholder Gains From Focusing Versus Diversifying Bank Mergers", *Journal of Financial Economics, vol. 59, p. 221–252*.

De Nicolo, G (2000), "Size, Charter Value and Risk in Banking: An International Perspective", *International Finance Discussion Papers No.689, Board of Governors of the Federal Reserve System.*

Dunbar, R I M (1993), "Coevolution of Neocortical Size, Group Size and Language in Humans", *Behavioral and Brain Sciences, vol. 16(4), p.681–694.*

Freidman, M and A J Schwartz (1963), *A Monetary History of the United States, 1867–1960,* Princeton University Press.

Haldane, A G (2009a), *Banking on the State,* available at http://www.bankofengland.co.uk/publications/speeches/2009/speech409.pdf

Haldane, A G (2009b), *Rethinking the Financial Network,* available at http://www.bankofengland.co.uk/publications/speeches/2009/speech386.pdf

IMF (2009), *World Economic Outlook: Crisis and Recovery,* available at http://www.imf.org/external/pubs/ft/weo/2009/01/pdf/text.pdf

Kay, J (2009), *Narrow Banking: The Reform of Banking Regulation,* Centre for the Study of Financial Innovation.

Kelling, M J, Woolhouse, M E J, May, R M and B T Grenfell (2003), "Modelling Vaccination Strategies Against Foot-and-mouth Disease", *Nature, vol. 421, p.136–142.*

Kennedy, S E (1973), *The Banking Crisis of 1933,* University Press of Kentucky.

Korup, O and J J Clague (2009), "Natural Hazards, Extreme Events, and Mountain Topography", *Quaternary Science Reviews, vol. 28(11–12), p.977–990*.

Kotlikoff, L J (2010), *Jimmy Stewart is Dead: Ending the World's Ongoing Financial Plague with Limited Purpose Banking*, John Wiley and Sons.

Laeven, L and R Levine (2007), "Is There are Diversification Discount in Financial Conglomerates?", *Journal of Financial Economics, vol. 85(2), p.331–367*.

Mehra, R and E C Prescott (1985), "The Equity Premium: A Puzzle", *Journal of Monetary Economics, vol. 15, p.145–161*.

Merton, R C (1973), "Theory of Rational Option Pricing", *Bell Journal of Economics and Management Science, vol. 4(1), p.141–183.*

Miles, D (2009), *The Future Financial Landscape,* available at http://www.bankofengland.co.uk/publications/speeches/2009/speech418.pdf

Modigliani, F and M Miller (1958), "The Cost of Capital, Corporation Finance and the Theory of Investment", *American Economic Review, vol. 48(3)*.

NYU Stern School of Business (2009), *Restoring Financial Stability: How to Repair a Failed System*, John Wiley & Sons.

Reinhart, C M and K Rogoff (2009), *This Time is Different: Eight Centuries of Financial Folly*, Princeton University Press.

Robinson, A (1934), "The Problem of Management and the Size of Firms", *The Economic Journal, vol. 44(174), p/242–257*.

Santomero, A M and D L Eckles (2000), "The Determinants of Success in the New Financial Services Environment: Now That Firms Can Do Everything, What Should They Do and Why Should Regulators Care?", *Federal Reserve Bank of New York Economic Policy Review (October), vol. 6(4), p.11–23*.

Saunders, A (1996), *Financial Institutions Management: A Modern Perspective*, Irwin Professional Publishing.

Schmid, M M and I Walter (2009), "Do Financial Conglomerates Create or Destroy Economic Value?", *Journal of Financial Intermediation, vol. 18(2), p.193–216*.

Stern, N (2006), *Stern Review on the Economics of Climate Change*, Cabinet Office-HM Treasury, available at http://www.hm-treasury.gov.uk/sternreview_index.htm

Stiroh, K and A Rumble (2006), "The Dark Side of Diversification: the Case of US Financial Holding Companies", *Journal of Banking and Finance*, *Vol.80, p.2131–2161.*

Tucker, P (2010), *Resolution of Large and Complex Financial Institutions: The Big Issues,* available at http://www.bankofengland.co.uk/publications/speeches/2010/speech431.pdf

Weitzman, M L (1974), "Prices vs. Quantities", *Review of Economic Studies, vol. 41, p.477–91*.

Annex

Table 1

Present value of output losses (% of 2009 GDP)

	Fraction of initial output loss which is permanent		
	25%	**50%**	**100%**
UK	130	260	520
World	90	170	350

Source: Bank calculations

Table 2

Average ratings difference for a sample of banks and building societies

	2007	**2008**	**2009**	**Average (2007–09)**
UK	1.56	1.94	4.00	**2.50**
Global	1.68	2.36	2.89	**2.31**
Average	1.63	2.21	3.24	**2.36**

1. All figures are year-end
2. The UK sample contains 16 banks and building societies in 2007 and 2008 and 13 in 2009. The global sample contains a sample of 26 banks across a range of sizes and countries for 2007 and 28 banks in 2008 and 2009.

Source: Moody's and Bank calculations.

Table 3

Average ratings difference for UK banks and building societies[(a)]

Category	**Mean**	**Max difference in sample**	**Min difference in sample**
2007			
Large banks	2.67	12	1
Small banks	0.14	1	0
2008			
Large banks	2.78	10	1
Small banks	0.86	2	0
2009			
Large banks	4.67	7	3
Small banks	3.43	6	0
Average (2007–2009)			
Large banks	**3.37**	**10**	**2**
Small banks	**1.48**	**3**	**0**

(a) The "Large" category includes HSBC, Barclays, RBS, Lloyds TSB, Alliance & Leicester and Bradford & Bingley (up to 2008), and Nationwide. The "Small" category includes building societies: Chelsea, Coventry, Leeds, Principality, Skipton, West Bromwich and Yorkshire. The ratings are year-end.

Source: Moody's and Bank calculations.

Table 4

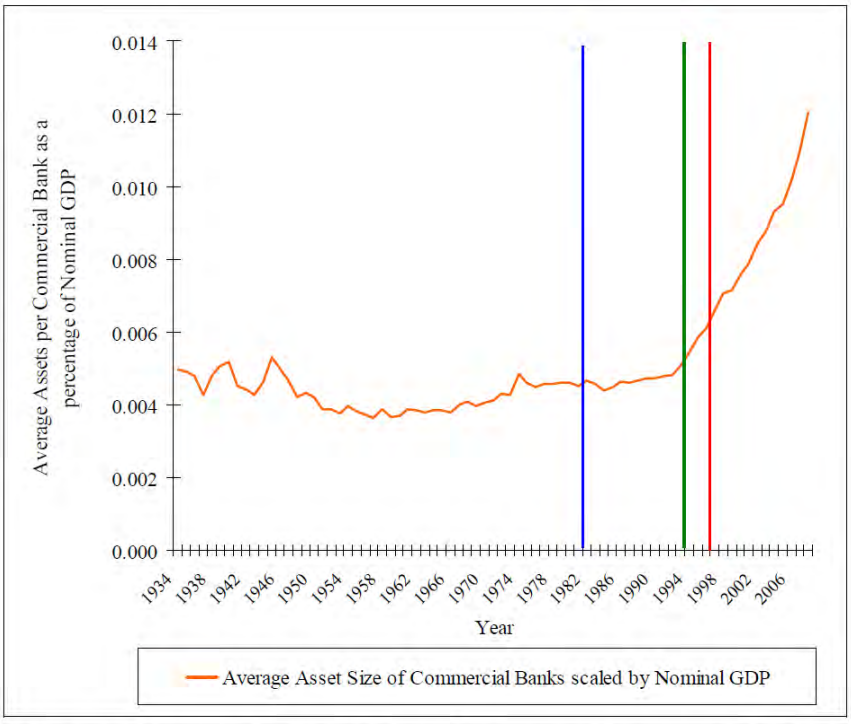
Estimated subsidy for UK banks and building societies (£bn) and global banks ($bn)

		2007		2008		2009		Average (2007–09)	
		Subsidy	Subsidy / Total liabilities	Subsidy	Subsidy / Total liabilities	Subsidy	Subsidy / Total liabilities	Subsidy	Subsidy / Total liabilities
UK	Sample Total	11		59		107		**59**	
Big 5	Total	9		52		103		**55**	
	Average	2	0	10	1	26	2	**13**	**1**
Medium	Total	1		7		3		**4**	
	Average	0	0	3	3	3	2	**2**	**2**
Small	Total	0		1		1		**1**	
	Average	0	0	0	1	0	1	**0**	**1**
Global	Sample Total	37		220		250		**169**	
Big 5	Total	18		83		71		**57**	
	Average	4	0	17	1	14	1	**12**	**1**

See footnotes for tables 2 and 3 for details on sample.
Source: Moody's, Bank of America Merrill Lynch, Bankscope published by Bureau van Dijk Electronic Publishing and Bank calculations

Chart 1

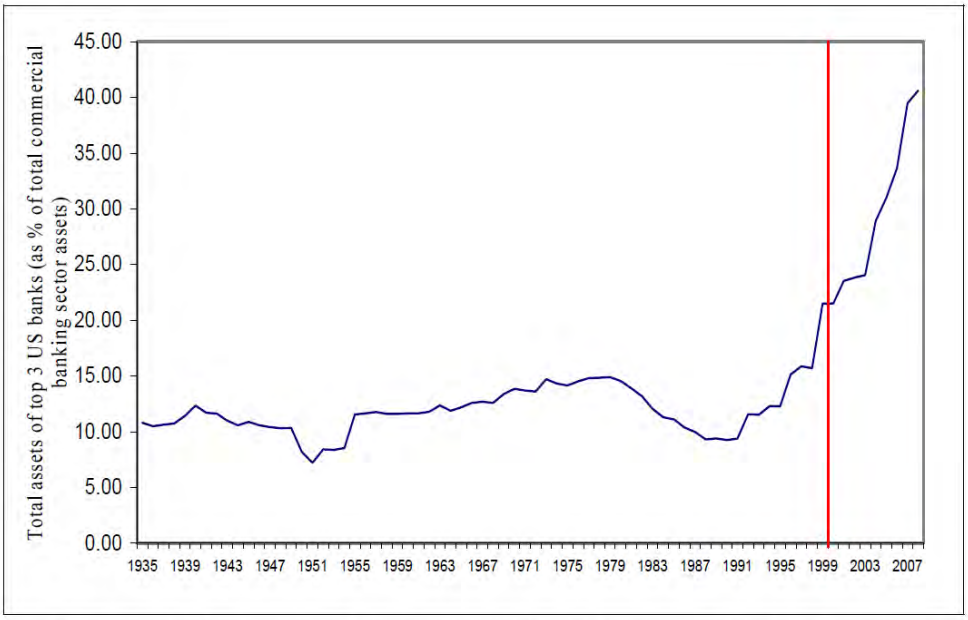
Average assets relative to GDP of US commercial banks[(a)]



(a) Blue vertical line represents the 1982 Garn-St Germain Act, green vertical line represents the 1994 Riegle-Neal Act, red vertical line represents the Riegle-Neal Act coming into effect in 1997.
Source: FDIC and www.measuringworth.org

Chart 2

Concentration of the US banking system[(b)]

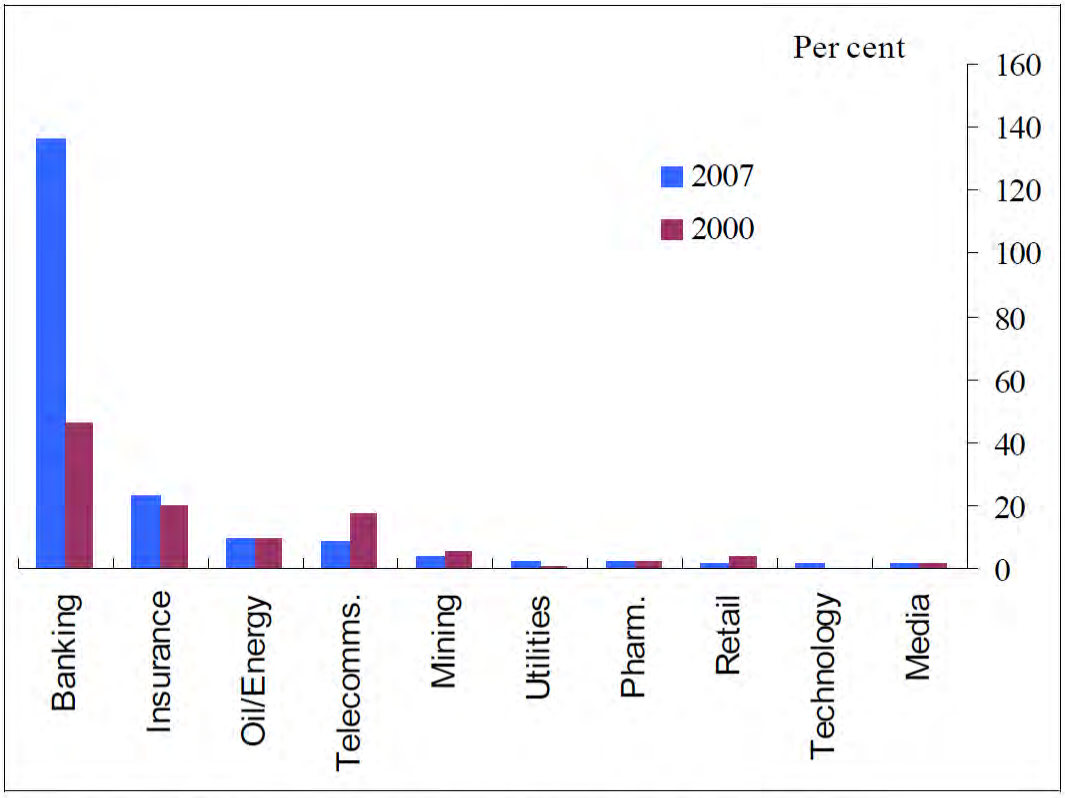


(a) Red line represents the Gramm-Leach-Bliley Act (1999) which revoked restrictions of Glass-Steagall
(b) Top 3 banks by total assets as a % of total banking sector assets
(c) Data includes only the insured depository subsidiaries of banks to ensure consistency over time - for example, non-deposit subsidiaries are not included.

Source: FDIC

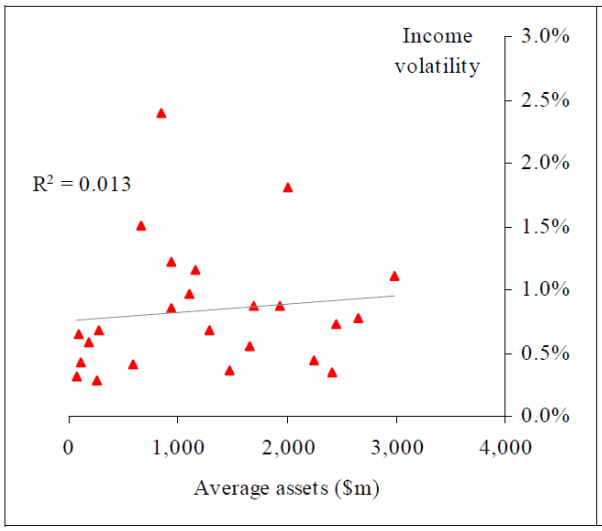
Chart 3

Largest UK company's assets in each sector relative to GDP



Source: Bureau van Dijk Electronic Publishing, International Monetary Fund and Bank calculations.

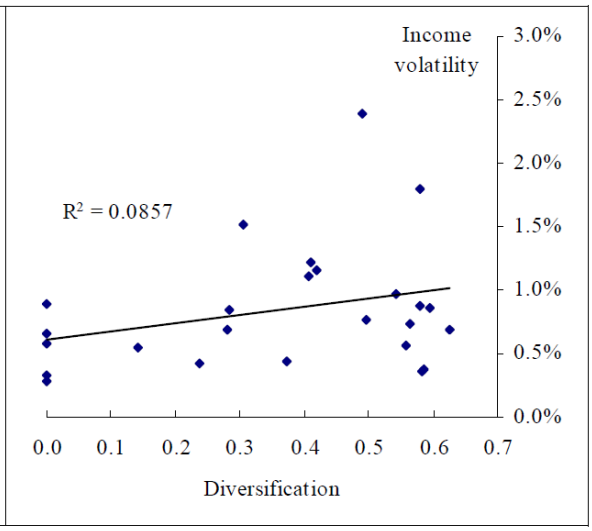
Chart 4: Bank size and volatility



Notes: Average assets are calculated for 24 banks between 2006 and 2008. Income volatility is measured as the standard deviation of operating income (per asset) over the period 1997-2008.

Source: Bankscope, published accounts and Bank calculations.

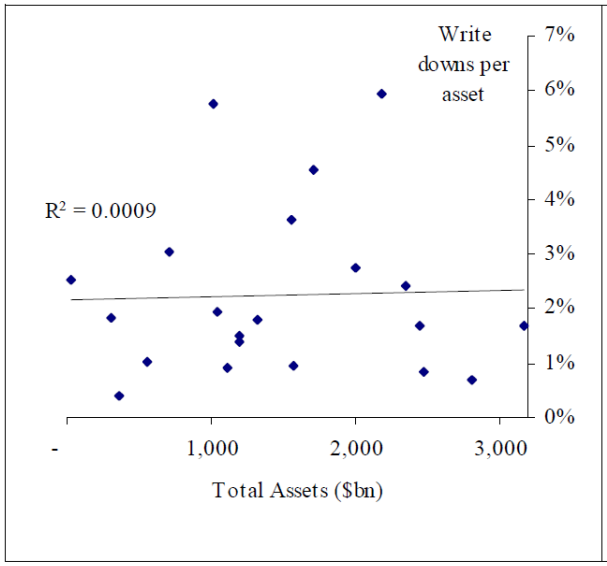
Chart 5: Bank diversification and volatility



Notes: Pre-crisis diversification and income volatility for a sample of 25 banks. Diversification index based on revenue concentration, as described in the main text.

Source: Bankscope, published accounts and Bank calculations.

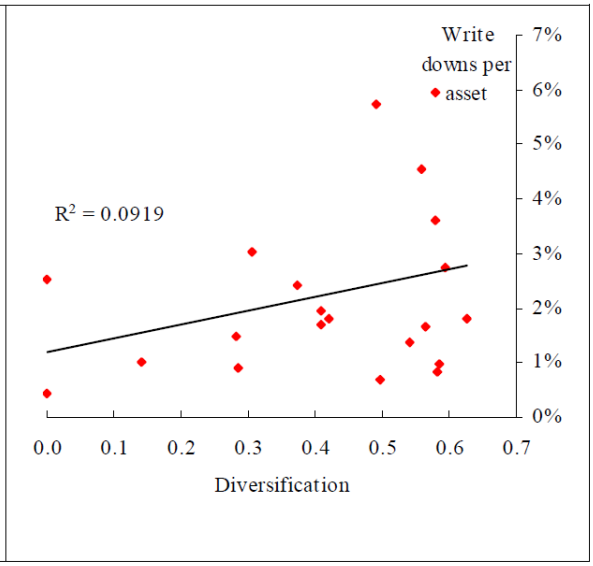
Chart 6: Bank size and write downs



Notes: Total assets for a sample of 21 banks for 2007. Cumulative write downs over the course of the crisis are shown (from 2007 Q4 to 2009 Q3).

Source: Bankscope, published accounts and Bank calculations.

Chart 7: Bank diversification and write downs



Notes: Sample of 21 banks. Cumulative write downs over the course of the crisis are shown (from 2007 Q4 to 2009 Q3).

Source: Bankscope, published accounts and Bank calculations.