Andrew G Haldane: On being the right size

Speech by Mr Andrew G Haldane, Executive Director, Financial Stability, Bank of England, at the Institute of Economic Affairs' 22nd Annual Series, The 2012 Beesley Lectures, at the Institute of Directors, London, 25 October 2012.

* * *

The views are not necessarily those of the Bank of England or the Financial Policy Committee. I would like to thank Andrew Gracie, Varun Paul, Kirsty Rodwell, Timothy Richards, Vicky Saporta, Rhiannon Sowerbutts and Belinda Tracey for their comments and contributions

1. Introduction

It is a pleasure to be delivering one of this year's Beesley Lectures. I want to discuss financial regulation. But my title is drawn from neither finance nor regulation. Instead it is drawn from evolutionary biology.

In 1928, evolutionary biologist J. B. S. Haldane wrote an important article whose title I have borrowed, *On Being the Right Size*.[1] The essential point was simple. The sheer size of an object, institution or animal determined their structure. In particular, as their size rose, their structure needed to strengthen more than proportionately if they were to remain robust and resilient.

This principle is sometimes enshrined in the so-called "square-cubed" law. A proportional rise in an object's size causes its area to rise by the square, and its volume by the cube, of that rise. At one level, this is simple mathematical geometry. Yet in the real world, it carries fundamental implications for evolutionary structure.

Take the animal kingdom. The square-cubed law explains why a flea, even if it were the size of a man, would not be capable of jumping to the moon. It explains why a hippopotamus cannot turn somersaults. And it explains why King Kong and Godzilla were physiological impossibilities – the weight transfer associated with a single step would have shattered their thigh bones.

When the world's biggest banking beasts took a step too far in 2008, they too folded under their own weight. Their physiological structure proved inadequate to make them robust and resilient. That is the essence of the "too-big-to-fail" problem. In the language of Haldane, international policymakers have concluded that many of the world's largest banks are not the right size given their existing physiological make-up.

Over the past few years, initiatives to solve the too-big-to-fail problem have come thick and fast. At root, each has aimed to strengthen the structure of the world's biggest banks. That is the good news. Claims that they have solved the too-big-to-fail problem appear to me, however, premature, probably over-optimistic. Worse, they risk sending a false sense of crisis comfort. That is the bad news.

To see why such a cautious conclusion is warranted, we begin by tracking the structural evolution of the financial system over the past few decades. We then consider the three most prominent policy initiatives aimed at tackling too-big-to-fail – systemic surcharges, resolution regimes and structural reform. In the final section we consider what supplementary policy options might be necessary to ensure banking is right-sized.

[1] Haldane (1928). Haldane is, regrettably, no relative.

2. Evolution of the financial system

The past fifty years have seen seismic shifts in the structure, size and composition of the global financial system. These changes gave birth to the too-big-to-fail problem. Chart 1 plots the ratio of banking sector assets-to-GDP, and its cross-country dispersion, for a set of 14 advanced countries over the past 140 years.[2]

For the better part of a century, between 1870 and 1970, financial deepening in these countries followed a modestly upward trend. Over this period, the average bank assets-to-GDP ratio rose from 16% to over 70%, or less than 6 percentage points per decade.

Since 1970, this trend has changed trajectory. The ratio of bank assets-to-GDP has more than doubled over the past 40 years, rising from around 70% to over 200%, or over 30 percentage points per decade. In other words, since 1970 financial deepening has occurred five times faster than in the preceding century. For some individual countries, the rise has been more dramatic still – in the UK, the ratio has risen five-fold.

In cross-country studies, financial deepening of this type has generally been found to have a positive effect on medium-term growth (Beck and Levine (2004)). Taken literally, this would suggest that the rise in banking scale over recent decades has provided a significant tailwind to medium-term growth in advanced countries. So it seemed in the pre-crisis period.

But that conventional wisdom has recently been called into question. IMF research has suggested that financial deepening can indeed be growth-positive – but within limits. There is a threshold at which private credit-to-GDP may begin to have a *negative* impact on GDP growth (Arcand, Berkes and Panizza (2012)). That threshold is found to lie at a private credit-to-GDP ratio of around 80–100%.

This finding is consistent with earlier cross-country evidence suggesting that, at credit-to-GDP ratios above unity, output volatility tends to increase (Easterly, Islam and Stiglitz (2000)).[3] This threshold lies significantly below current levels of financial depth in most advanced economies. In other words, taken at face value this evidence suggests that, at its current scale, banking could be acting as a headwind to medium-term growth.

Accompanying this dramatic rise in banking scale has been an equally-dramatic rise in banking concentration. Chart 2 plots the evolution of the asset share of the three largest banks in the US, UK, Germany and Switzerland over recent decades.

The US has undergone the most dramatic upwards shift, with the share of the top three banks rising from around 10% to 40% between 1990 and 2007. For other countries, there is a less dramatic rise in concentration but from a much higher starting point, with the top three banks accounting for between two-thirds and three-quarters of assets in the UK, Switzerland and Germany.

This dramatic rise in banking scale and concentration has been driven by financial liberalisation. But it has also spawned an acute problem for society due to escalating expectations of state support for the banking system. These expectations generate lower funding costs, in particular for the largest banks, which in turn encourages further expansion and concentration, worsening the too-big-to-fail dilemma. There is a self-perpetuating "doom loop" (Alessandri and Haldane (2010)).

The size of the resulting "implicit subsidy" from the state to the big banks has been the subject of recent study – and no little controversy (Noss and Sowerbutts (2012)). There are a number of possible methods for estimating the subsidy. Perhaps the simplest is found by comparing the "standalone" and "support" ratings assigned to debt issued by large banks.

---

[2] The data are from Jorda, Schularick and Taylor (2011).

[3] This has strikingly close parallels with recent econometric evidence suggesting that growth is harmed at debt-to-GDP ratios above unity, for both the public sector (Reinhart and Rogoff (2010)) and private sectors (Cechetti et al (2011)).

The difference between these ratings gives an estimate, used by the market when pricing debt, of the probability of state support.

Chart 3 plots the difference between these ratings for the 29 institutions deemed by the Financial Stability Board (FSB) last year to be the world's most systemically-important. In the pre-crisis period, this difference averaged 1.3 notches. This sounds material, if not breathtaking. It suggests that, while non-trivial, too-big-to-fail may not have been a first-order driver of the rising scale and concentration in banking.

Yet even small notches of support can translate into big implicit subsidies if balance sheets are large. So it is for the world's largest banks. Over the period 2002 to 2007, the implied annual subsidy to the world's largest banks averaged $70 billion per year using a ratings-based measure (Chart 4). That is roughly 50% of the average post-tax profits of these banks over the period.

As the crisis struck, this implicit promise became explicit. Financial support was extended to the banking system in the form of capital injections, guarantees and liquidity insurance. On some estimates, this support rose to three-quarters of annual GDP in some countries. In response to these interventions, there has predictably been a further ratcheting-up in ratings-implied degrees of state support to banks.

By 2009, the ratings difference had more than doubled to above three notches, with the implied monetary subsidy over $700 billion per year. That was well in excess of average annual pre-crisis profits of these firms. Even if an over-estimate, the scale of the implied subsidy signalled something dramatic was at play. Too-big-to-fail had become hard-wired into the structure and pricing of the financial system.

3. Systemic surcharges

One way of interpreting these implicit subsidies is as the market's best guess of how much a policymaker would be willing to pay each year to avoid the failure of the world's biggest banks. They proxy the expected social costs of big bank failure. In the jargon, they capture a systemic externality.

This notion of a systemic externality has underpinned recent academic and policy efforts to solve the too-big-to-fail problem. Brunnermeier et al (2009) use this framework to motivate levying a tax – a "Pigouvian tax" – on institutions posing systemic risk externalities. This tax would be set at levels which offset the effects of the bank's actions on wider society. A number of academics have since proposed measures along broadly Pigouvian lines (Archarya et al (2010)).

Rather remarkably, policy reforms in practice have followed closely in the spirit of these proposals. In 2010, the FSB announced its intention to introduce a "systemic surcharge" of additional capital on the world's largest banks. In July 2011, the Basel Committee published a methodology for measuring systemic importance based on indicators of bank size, connectivity and complexity, with additional capital of up to 2.5% depending on this score.

In November 2011, the FSB endorsed this methodology and announced the 29 systemically-important entities to which it would apply. This framework will be finalised this year, before being phased-in from 2016. Legislation is already in place, or is being drawn up, to implement the systemic surcharge in the United States and Europe.

These proposals are clearly a practical step in the right direction. By boosting levels of capital in the system, the probability of big bank failure will be reduced. You would have got good odds back in 2007 that something as seemingly elliptical as a Pigouvian tax on systemic risk would have found its way onto the regulatory statute books. Now we have it. "We are all Pigouvians now", even if most of us cannot spell it.

The practical question is how far systemic surcharges take us in tackling the systemic risk externality. In other words, at current levels by how much have systemic surcharges reduced

expected losses for the financial system? That is an empirical question. To assess it, consider the impact of capital surcharges on the expected losses facing the 29 institutions identified by the FSB.

Using each of these banks' balance sheets, we generate a measure of default probability using the Merton (1974) contingent claims model. We assume that the base level of equity for these banks is 7% (the new Basel III minimum) and that, in the event of default, they suffer losses on their assets of around 30%.[4] To keep things simple, banks' assets are assumed to be (log-)normally distributed, in line with Merton (1974), and default occurs only when a bank's capital is fully exhausted.

Assume initially that default risks across banks are independent. This is a highly conservative assumption, as in practice bank default probabilities are highly correlated at times of stress. Indeed, bank correlation coefficients in crisis often head towards one. For that reason, this thought-experiment provides a lower bound on expected losses across the financial system.

Chart 5 shows expected losses across the 29 banks at different levels of the systemic surcharge. In the absence of any surcharge, expected losses across the system are just less than $200 billion per year. Were every large bank instead to face the maximum capital surcharge of 2.5%, then expected system-wide losses would fall by around 60% relative to their base level. And to remove 90% of the systemic externality – expected losses of around $5 billion – a surcharge of over 7% would be needed.

A more plausible experiment would be to assume a non-zero correlation among bank defaults. For example, the failure of a large bank which caused it to fire-sale assets could impose externalities on other large banks holding these same assets (Wagner (2009)). To place an upper bound on expected losses in the face of these fire-sale externalities, assume instead a correlation coefficient of one.

Chart 5 illustrates the impact on expected system-wide losses of a high degree of default correlation among the big banks. The expected system-wide loss increases to around $750 billion per year – similar in size to the implicit subsidy at its peak. A 2.5% surcharge now only reduces expected system-wide losses to around $350 billion per year. To lower expected system-wide losses to be around $5 billion per year would require a surcharge of around 15% – six times its current upper limit.

If anything, these thought-experiments probably produce conservative estimates of system-wide losses and the necessary systemic surcharge. For example, bank asset returns are in practice much fatter-tailed than the log-normal distribution. And in practice, banks are likely to default well before their capital is fully exhausted. Relaxing either assumption would push up estimates of expected losses and the surcharge necessary to curtail these losses (Schanz et al (2012)).

Nonetheless, if expected system-wide losses are a reasonable proxy for the system-wide externalities large banks pose, this analysis delivers a rather gloomy prognosis. At current levels of the surcharge, a large chunk of the systemic externality would remain untouched. If too-big-to-fail is the problem, then systemic surcharges seem to offer only a partial solution.

Resolution regimes

Capital surcharges lower systemic externalities by lowering default probabilities for the world's largest banks. An alternative way of lowering those externalities would be to reduce the collateral damage associated with their failure. This has been a key motivation for a second strand of the reform debate – the design of effective resolution regimes.

[4] James (1991). We also assume bank asset volatility of 4% per year, in line with previous empirical studies.

Few examples better illustrate the costs of getting this wrong than the spiralling queues outside branches of Northern Rock in September 2007. Despite being a medium-sized retail bank, Northern Rock's failure caused systemic disruption and put taxpayers' money at risk. In response, the UK put in place in 2009 a special resolution regime for banks, providing the Bank of England with tools for winding-down a failed bank.

As the crisis illustrated, financial failure which causes systemic disruption is not confined to banks. And to avoid taxpayer bail-out, losses may need to be imposed on a wide class of bank creditors, including holders of debt as well as equity – for example, by "bail-in". Over recent years, these resolution lessons have been enshrined in banking legislation.

For example, in the United States Title II of the Dodd-Frank Act was passed in July 2010. It creates a new regime for the liquidation of financial companies, banks and non-banks, which pose a systemic financial stability risk. It enables losses to be imposed on creditors in resolution, while also prohibiting state bail-outs.

Internationally, in November 2011 the G20 endorsed the FSB's *Key Attributes of Effective Resolution Regimes,* developed by an international working group chaired by my colleague Paul Tucker. Efforts are now underway to align national resolution regimes with these principles. As part of that, in Europe a draft Directive on bank recovery and resolution was published in June 2012.

These initiatives are an important practical step in the right direction, lowering the societal costs of bank failure. As with systemic surcharges, it is striking how much progress has been made in so short a space of time on so complex an issue. The practical question is how far this takes us towards removing the too-big-to-fail externality.

During a bank resolution, one way of ensuring continuity of banking services is by transferring assets and/or liabilities of a failing firm to a third party. But the only entity with sufficient financial and managerial resource to absorb a large asset or liability portfolio, without suffering chronic indigestion, is another big bank. So it was during the crisis – for example, Bear Stearns was swallowed by JP Morgan Chase, Merrill Lynch by Bank of America and Washington Mutual by Citigroup.

This makes for a rather uncomfortable evolutionary trajectory, with rising levels of banking concentration and ever-larger too-big-to-fail banks. Levels of banking concentration have risen in many countries since 2007, precisely because of such shot-gun marriages by over-sized partners. In other words, resolving big banks may have helped yesterday's too-big-to-fail problem, but at the expense of worsening tomorrow's.

One way of avoiding this problem is to re-capitalise a bank by bailing-in its creditors, rather than transferring its assets. But resolution rules of this type are not problem-free either. Like all policy rules, they face what economists call a time-consistency problem. Whether a rule is followed in practice depends on the balance of costs and benefits at the time crisis strikes, not at the time the rule is written. That is why policy might in practice lack consistency over time – hence time-inconsistency.

Consider that trade-off when a big, complex bank hits the rocks. On the one side is a simple, but certain, option – state bail-out. On the other is a complex, and less certain, option – resolution. Policymakers face a trade-off between placing losses on a narrow set of tax-payers today (bail-in) or spreading that risk across a wider set of tax-payers today and in the future (bail-out). If governments are risk-averse and wish to smooth the pain across tax-payers and across time, then bail-out may look attractive on the day.

Financial history certainly suggests so. The history of big bank failure is a history of the state blinking before private creditors (Haldane (2011)). Recent crisis experience has written another chapter in this history. Next time may be different. For example, the public backlash against future bail-outs could reinforce government's resolve to impose losses on creditors. And recent US legislation in principle locks the tax-payer cashbox and throws away the key. Looking forward, the issue is whether this ex-ante rule is ex-post credible.

As Chart 6 illustrates, implied levels of support for the US's biggest banks are much higher than before the crisis. More telling still, the passage of Dodd-Frank appears to have had little impact on levels of implied state support. It is early days for this new resolution regime and credibility may take time to be earned. Nonetheless, at present the market believes the time-consistency problem for big banks is as acute as ever.

Even if it might appear the expedient option on the day of crisis, it is questionable whether bail-out is the optimal response over the medium-term. Chart 7 looks at the response of bank and sovereign CDS spreads around the time of bank bail-outs in a selection of crisis countries. While bail-outs lowered bank CDS spreads, as might be expected, bail-out came at the expense of a rise in sovereign CDS spreads.

It is not difficult to see why. The financial crisis has caused huge damage to the balance sheets of governments in advanced economies. For the G20 countries, the IMF forecast that the debt-to-GDP ratio will rise by almost 40 percentage points between 2007 and 2016, to almost 120%. At these levels, public sector debt may be a significant drag on medium-term growth (Rogoff and Reinhart (2010)).

For economies with large banking systems and without a credible resolution regime, this leaves policymakers caught between a rock and hard place. When the call comes to ride to the banking rescue, government may be unable to afford not to. But nor, at least over the medium term, can they afford to. This is just the dilemma facing advanced countries today.

Structural reform

One way of lessening that dilemma, and at the same making resolution and bail-in more credible, is to act on the scale and structure of banking directly. Perhaps not surprisingly, recent regulatory reforms have sought to do just that. They have taken seriously the maxim that, if a bank is too big to fail, then it is too big. The result has been detailed proposals for structural reform in a number of advanced countries.

In the United States, the "Volcker rule" has been introduced. This prohibits US-operating banks from undertaking proprietary trading and restricts private equity activity. The rule was tabled in October 2011 and became law in July 2012. Banks have two years to comply.

In the United Kingdom, the proposals of the "Vickers Commission" include placing a ring-fence around retail banking activities, supported by higher levels of capital. The final version of these proposals was tabled in September 2011 and legislation to enact them is planned by 2015. Banks will have until 2019 to comply.

Most recently in Europe, the "Liikanen plan" was announced in October 2012. It proposes that the investment banking activities of universal banks be placed in a separate entity from the remainder of the banking group. There are at present no plans to legislate these proposals.

As with the other reform strands, it is remarkable how quickly radical structural reform proposals are finding their way onto the statute book. And although different in detail, these proposals share a common motivation: separation of certain investment and commercial banking activities. In theory, such a separation delivers financial stability benefits of two distinct types (Boot and Ratnovski (2012)).

First, separation reduces the risk of cross-contamination. Riskier investment banking activities, when they go wrong, can pollute and dilute the financial resources of the retail bank. This potentially inflicts losses (or fear of losses) on depositors. It may also constrain banks' ability to make loans to the real economy when it might most need them. This is a crisis-time benefit of separation.

Second, separation can secure an improved pre-crisis allocation of financial resources from a societal perspective. High private return investment banking activities may crowd-out the human and financial resources devoted to high social return commercial banking activities.

Investment banking activities might also piggy-back on the cheaper cost of deposit funding. In effect, universal banking allows privately optimal, but socially sub-optimal, cross-subsidisation. This is a normal-times benefit of separation.

Both of these costs were evident ahead of, and during, the recent crisis. Ahead of crisis, resources gravitated to the investment banking side of the fence. Between 2000 and 2007, UK banks' trading books rose six times as fast as their banking books. Human capital made the same journey, helped by investment banking salaries rising four times as fast as commercial bank salaries since 1980.

In the teeth of crisis, risk cross-contamination became a potent factor. Basic banking services in universal banks were often subject to severe disruption from trading book losses, which exceeded by many multiples the capital allocated to them. That is why national deposit insurance schemes were extended and in some cases became temporarily unlimited. It is also why repeated attempts have had to be made to resuscitate weak credit growth over the past few years.

So how far will existing structural proposals take us in harnessing these benefits? Volcker, Vickers and Liikanen seek legal, financial and operational separation of activities. So in principle each ought to prevent cross-contamination at crisis time. Whether they do so in practice depends on loopholes in, or omissions from, the ring-fence. And each of the existing proposals has open questions on this front.

For example, the Volcker rule separates only a fairly limited range of potentially risky investment bank activities, in the form of proprietary trading. The Vickers proposals mandate only a limited range of basic banking activities to lie within the ring-fence, namely deposit-taking and overdrafts. And the Liikanen plans allow a wide range of derivative activity to lie outside of the investment banking ring-fence.

It could be argued that these loopholes are modest. But as the history of the Glass-Steagall Act demonstrates, today's loophole can become tomorrow's bolthole, today's ring-fence tomorrow's string vest. At a minimum, this suggests the need for full and faithful implementation of the spirit as well as the letter of the Volcker, Vickers and Liikanen plans if risk cross-contamination is to be avoided

A larger question-mark still hangs over whether these proposals will lead to a sea-change in the allocation of resource to retail and investment banking. The cultures of investment and retail banking are quite distinct. Retail banking relies on forming long-term relationships, while investment banking is inherently shorter-term and transactional. Housing these sub-cultures under one roof makes achieving the necessary separation of cultures and capital a significant operational headache.

At a minimum, such a separation of culture and capital is likely to require entirely separate governance, risk and balance sheet management on either side of the ring fence. Without that, human and financial resource allocation either side of the ring-fence will become blurred. For example, without separate debt issuance for retail and investment banking, the cost of debt for a big bank will be a blended mix. The implicit subsidy in funding costs would then remain and with it one of the main distortions associated with too-big-to-fail.

Only time will tell whether cultural separation can be achieved under the existing structural reform proposals. In the go-go years, will these reforms be sufficient to prevent the grass always appearing greener on the riskier side of the (ring-)fence? This is the acid test of the structural reform agenda.

Where next?

Progress has been made over the past few years towards eliminating too-big-to-fail, with further progress on implementation planned. But today's task is even more daunting than before the crisis. The big banks are even bigger. The system itself is more concentrated. And

despite reform efforts, the market's best guess today about tomorrow's implicit subsidy is far-larger than before the crisis struck, at over $300 billion per year (Chart 3). The market believes that illicit state promise is even more likely to be kept.

The wrong conclusion to draw would be that existing reforms have failed or are unnecessary – quite the contrary. Rather it is that these reform initiatives, while necessary, may be insufficient to eliminate the too-big-to-fail externality. If so, what are the alternatives? Several have been mooted.

Re-sizing the capital surcharge is one possibility. This would further reduce default probabilities among the biggest banks, thereby lowering the expected system-wide losses associated with big bank failure. Taking the earlier illustrative example, to reduce materially expected system-wide losses for the world's largest banks would require a capital surcharge several times larger than its current upper limit. Interestingly, this would take bank capital ratios to levels not dissimilar to recent quantitative estimates of their optimal value (Miles et al (2011), Hellwig et al (2011)).

Placing limits on bank size is a second option. By reducing balance sheet exposures, this measure would reduce directly system-wide losses in the event of big bank failure. The Dodd-Frank Act includes an explicit limit on the maximum deposit market share of US banks, capping it at 10%. But this does not prevent banks rising to a scale, relative to GDP, at which they could imperil state solvency. For that reason, limiting bank size relative to GDP has recently been proposed by a number of commentators and policymakers (Fisher (2011), Hoenig (2012), Johnson and Kwak (2010)), Tarullo (2012)).

Full structural separation of investment and commercial banking, a modern-day Glass-Steagall Act, has continued to attract support. The main benefit this would bring, relative to structural ring-fencing, is that it would eliminate loopholes from the ring-fence and better ensure that the distinct cultures of retail and investment banking were not cross-contaminated. That would lessen the risk of basic banking activities being starved of human or financial capital, both ahead of and during crisis. Full separation may also be operationally simpler to implement than the existing structural proposals.

Finally, *enhanced banking competition* would potentially help to reduce some of the problems of too-big-to-fail by reducing the degree of banking concentration. Greater exit from banking, through enhanced resolution regimes, can help. But a bigger problem still is bank entry: the UK went 100 years without a new retail bank being set up. One potential barrier to banking entry is the difficulty of switching deposit accounts and loan contracts. A shared banking platform, containing customer account details, would dramatically reduce the frictions in search and switch for deposit and loan products for customers (Leadsom (2012)). It could lower materially barriers for new banking entrants.

A powerful counter-argument to all of these more radical proposals is that they could erode the economies of scale and scope associated with large banks. These economies clearly do exist in banking, as they do in other industries. For example, fixed costs are large in finance and spreading them widely ought to deliver productivity improvements.

The interesting question is at what point these economies of scale are exhausted. Indeed, informational and managerial *dis*economies of scale are likely at some scale, whatever the business line. In his classic theorem, Ronald Coase tells us that firms will seek a privately-optimal size which balances the benefits of economies of scale against these diseconomies (Coase 1937)).

How does all of this apply in banking? The empirical evidence on economies of scale and scope in banking is surprisingly patchy. Early studies, using data from the 1980s, failed to find scale economies much beyond bank asset sizes of around $100 million (Pulley and Braunstein (1992)). Empirical studies in the 1990s nudged up the optimal bank scale to around $10 billion (Amel et al (2004), Mester (2005)).

Most recently, a small number of studies using data from the 2000s have pointed to scale-economies at much higher asset thresholds. For example, Wheelock and Wilson (2012) find scale economies for banks with assets up to $1 trillion and Feng and Serilitis (2009) for bank with assets up to around $1.5 trillion. Using data on banks with assets in excess of $100 billion, Mester and Hughes (2011) not only find scale-economies, but argue that these may increase with bank size.

At face value, these findings pose a real challenge to policy options which re-size banks. Or do they? Bank of England research has re-looked at the evidence on economies at different banking scales (Davies and Tracey (2012)). As Chart 8 shows, in standard models there is evidence of scale economies for banks with assets above $100 billion. Indeed, these economies tend to rise with banking scale.

But this finding is based on estimates of banks' funding costs which take no account of the implicit subsidy associated with too-big-to-fail. Removing this subsidy raises banks' funding costs, lowers estimates of bank value-added and thereby reduces measured economies of scale. As Chart 9 shows, once an allowance is made for the implicit subsidy, the picture changes dramatically. There is no longer evidence of economies of scale at bank sizes above $100 billion. If anything, there is now evidence of diseconomies which rise with bank size, consistent with big banks becoming "too big to manage".

This evidence reconciles Coase's Theorem with the too-big-to-fail phenomenon. In line with Coase, banks have chosen the size which maximises their private value. But implicit subsidies may have artificially boosted the privately-optimal bank size. Subtracting this subsidy, removing the state crutch, would suggest a dramatically lower socially-optimal banking scale. Like King Kong and Godzilla, these giants would arguably then be physiological impossibilities.

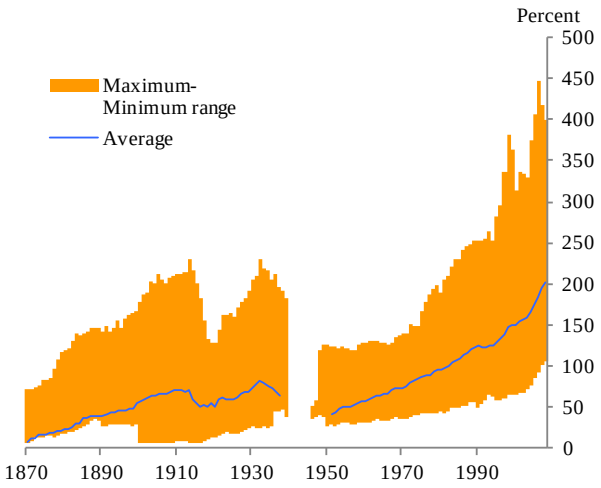
What about economies of scope? A recent study by Boyd and Heitz (2011) conducts a simple, but compelling, thought-experiment. They compare the lowest-available estimate of the social cost of the crisis with the highest-available estimate of the private benefit of scale and scope economies in banking. The social costs of too-big-to-fail exceed the private benefits of scale economies by an order of magnitude.

Conclusion

In his 1928 article, J B S Haldane observed that when "you drop a mouse down a thousand-yard mine shaft it walks away, a rat is killed, a man is broken, a horse splashes". When big banks disappeared down the mineshaft in 2008, their splashes generated a tsunami. To prevent that, their physiology needs to change. Existing change initiatives are right in direction, but may be insufficient in degree. There may be a distance to travel before banking is the right size.

Chart 1

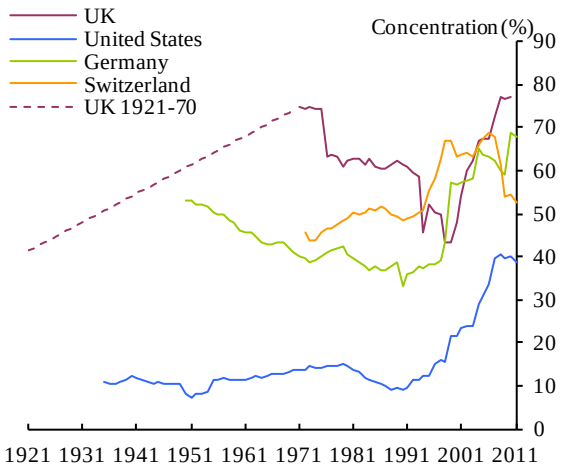
Bank assets/GDP in selected countries



Source: Jorda, Schularick and Taylor (2011)

Chart 2

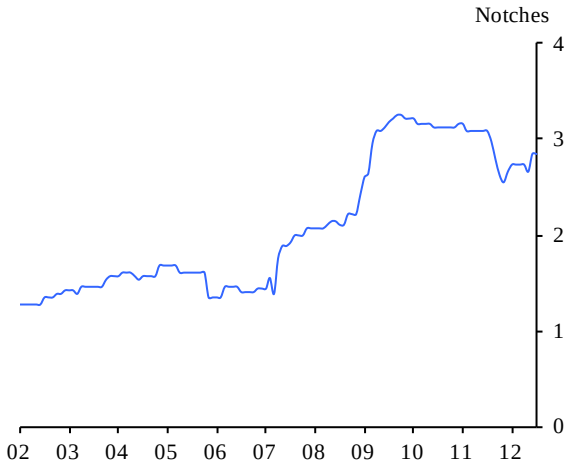
Banking concentration in selected countries



Source: Bank of England calculations, Capie, F. and Webber, A. (1985), FDIC, Bundesbank, Swiss National Bank, Bank of England calculations

Chart 3

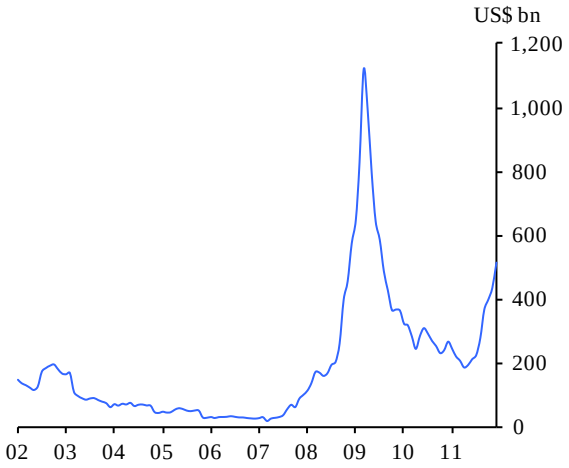
Ratings uplift for systemic institutions



Source: Moody's, Bank of England calculations

Chart 4

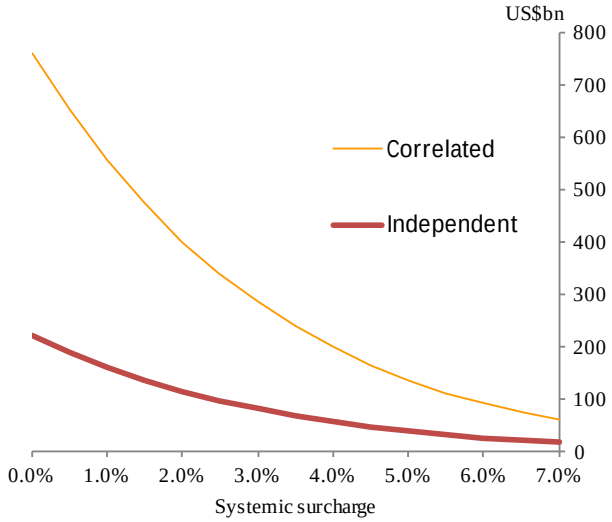
Implicit subsidy for systemic institutions



Source: Bank of England calculations

Chart 5

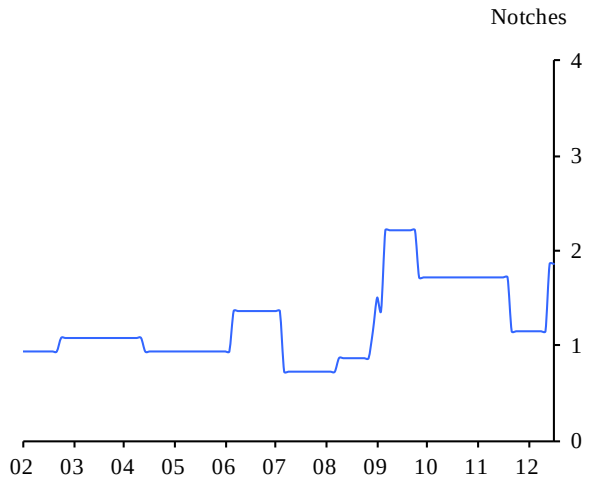
Expected system-wide loss



Source: Bank of England calculations

Chart 6

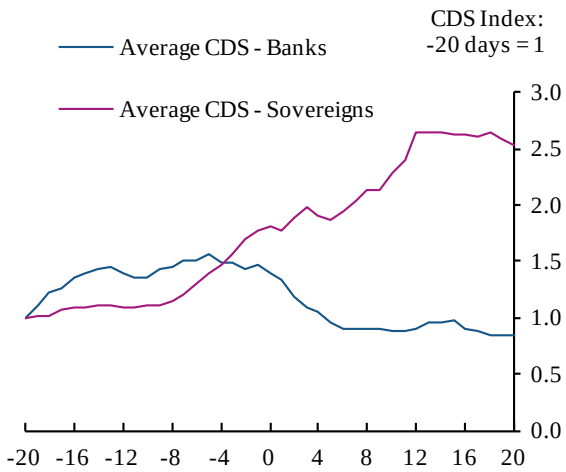
Bank ratings uplift in the US



Source: Moody's, Bank of England calculations

Chart 7

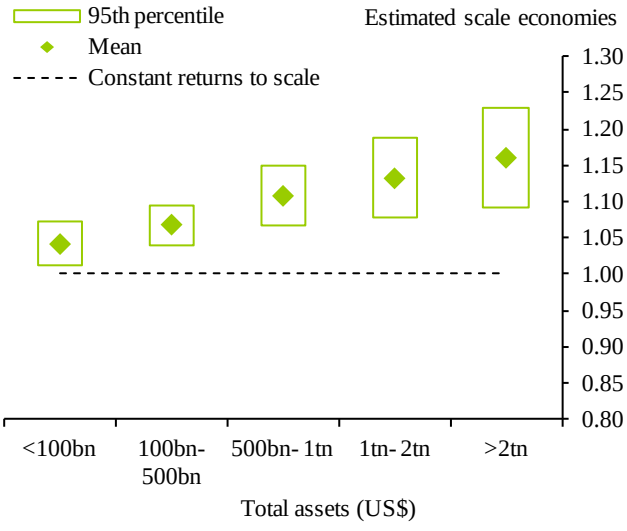
Sovereign and Bank CDS during bailout announcements



Source: Markit, BoE calculations

Chart 8

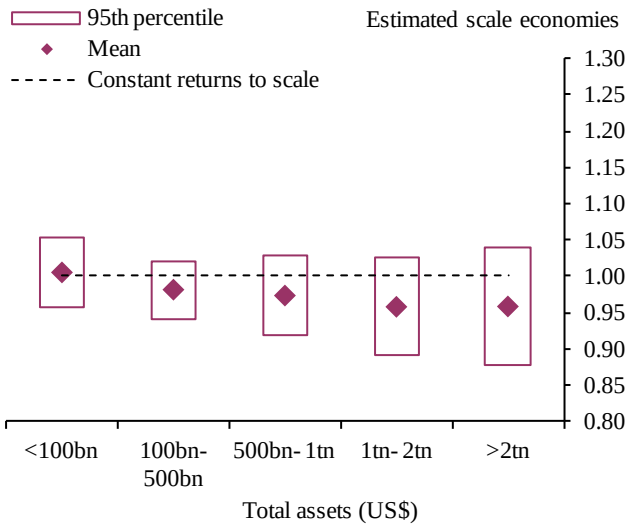
Economies of scale (assuming no implicit subsidy)



Source: Capital IQ, Bank of England calculations

Chart 9

Economies of scale (implicit subsidy-adjusted)



Source: Capital IQ, Bank of England calculations

References

Acharya, V, Pedersen, L Philippon, T and Richardson, M (2010), *"A tax on systemic risk",* NYU Stern Working Paper.

Admati, A., DeMarzo, P., Hellwig, M. and Pfleiderer, P (2011), *"Fallacies Irrelevant Facts and Myths in the Discussion of Capital Regulation: Why Bank Equity is not Expensive"* (March 23, 2011). Rock Center for Corporate Governance at Stanford University Working Paper No. 86.

Alessandri, P and Haldane, A (2010), *"Banking on the state",* Bank of England.

Amel, D and Barnes, C and Panetta, F and Salleo, C (2004), *"Consolidation and efficiency in the financial sector: A review of the international evidence,"* Journal of Banking and Finance, Elsevier, vol. 28(10), pages 2493-2519, October.

Boot, A and Ratnovski L (2012), *"Banking and Trading",* IMF Working Paper WP/12/238.

Boyd, J and Heitz (2011), *"The Social Costs and Benefits of Too-Big-To-Fail Banks: A bounding exercise",* Working Paper.

Brunnermeier, M, Crockett, A, Goodhart, C, Persaud, A and Shin, H (2009), "*The Fundamental Principles of Financial Regulation,"* London, Centre for Economic Policy Research.

Cecchetti, S and Mohanty, M and Zampolli, F (2011). *"The real effects of debt",* BIS Working Papers 352, Bank for International Settlements.

Coase, R.H. (1937), *"The Nature of the Firm,"* Economica, New Series, Vol. 4, No. 16. (Nov 1937), pp. 386–405.

Davies, R and Tracey, B (2012), "*Too big to be efficient? The impact of too big to fail factors on scale economies for banks*", Mimeo

Easterly, W., Islam, R., and Stiglitz, J. (2000), "*Shaken and Stirred, Explaining Growth Volatility,"* Annual Bank Conference on Development Economics. World Bank, Washington D.C..

Feng, G., and A. Serilitis (2009), *"Efficiency, technical change, and returns to scale in large US banks: panel data evidence from an output distance function satisfying theoretical regularity."* Journal of Banking and Finance, 34(1), 127–138.

Fisher, R (2011), *"Taming the Too-Big-To-Fails: Will Dodd-Frank be the ticket or is Lap-band surgery required?"* Remarks before Columbia University's Politics and Business Club.

Hoenig, T (2012), *"Back to basics: A better alternative to Basel Capital Rules",* Speech to The American Banker Regulatory Symposium, September 14. http://www.fdic.gov/news/news/speeches/chairman/spsep1412_2.html

Hughes, J and Mester, L (2011), *"Who said large banks don't experience scale economies? Evidence from a risk-return-driven cost function,"* Working Papers 11-27, Federal Reserve Bank of Philadelphia.

James, C (1991), "*The Losses Realized in Bank Failures*", Journal of Finance, 46, issue 4, p. 1223–42.

Johnson, S and Kwak, J (2010), *"13 Bankers: The Wall Street Takeover and the Next Financial Meltdown",* Pantheon

Jorda, O and Schularick, M and Taylor, A (2011), *"Financial Crises, Credit Booms, and External Imbalances: 140 Years of Lessons,"* IMF Economic Review, Palgrave Macmillan, vol. 59(2), pages 340–378, June.

Leadsom, A (2012), *"How an old hand would change Barclays"*, Financial Times Opinion Editorial, July 4. http://www.ft.com/cms/s/0/fb094718-c5f7-11e1-b57e-00144feabdc0.html#axzz2A6YBHLDV

Levine, R (2004), *"Finance and Growth: Theory and Evidence,"* NBER Working Papers 10766, National Bureau of Economic Research, Inc.

Merton, Robert C. (1974), *"On the Pricing of Corporate Debt: The Risk Structure of Interest Rates",* Journal of Finance, Vol. 29, No. 2, (May 1974), pp. 449–470.

Mester, L.J. (2008), *"Optimal industrial structure in banking,"* in Handbook of Financial Intermediation, Arnoud Boot and Anjan Thakor, eds. Amsterdam: North-Holland, 133–162.

Miles D, Yang J and Marcheggiano, G (2012), *"Optimal Bank Capital"*, Economic Journal.

Noss, J and Sowerbutts, R (2012), *"The Implicit Subsidy of banks"* Bank of England FS Papers Series, FS Paper No 15.

Panizza, U and Arcand, J-L and Berkes, E (2012), *"Too Much Finance?,"* IMF Working Papers 12/161, International Monetary Fund.

Philippon, T and Reshef, A, (2009), *"Wages and Human Capital in the U.S. Financial Industry: 1909–2006,"* NBER Working Papers 14644, National Bureau of Economic Research, Inc.

Pulley, L. B. and Y. M. Braunstein (1992), *"A composite cost function for multiproduct firms with an application to economies of scope in banking"*, Review of Economics and Statistics, 74(2), 221–30.

Reinhart, C and Rogoff, K, (2010), "*Growth in a Time of Debt,"* American Economic Review, American Economic Association, vol. 100(2), pages 573–78, May.

Segoviano, M and Goodhart, C (2009), *"Banking Stability Measures,"* IMF Working Paper 09/04 (Washington: International Monetary Fund).

Schanz, J, Aikman, D, Collazos, P, Farag, M, Gregory, D and Kapadia, S (2011), *"The long-term economic impact of higher capital levels",* Bank of England, Mimeo

Tarullo, D (2012), *"Financial Stability Regulation"*, Speech At the Distinguished Jurist Lecture, University of Pennsylvania Law School, Philadelphia, Pennsylvania.

Wagner, W (2009), *"In the Quest of Systemic Externalities: A Review of the Literature*", mimeo

Wheelock, D.C., and Wilson P.W. (2012), *"Do Large* Banks *have Lower Costs? New Estimates of Returns to Scale for U.S. Banks."* Journal of Money, Credit, and Banking, 44(1), 171–199.