Do we need regulation of bank capital?

Some evidence from the UK

IEA Current Controversies No. 40

by

Forrest Capie and Geoffrey Wood

March 2013

The Institute of Economic Affairs, 2 Lord North Street, London, SW1P 3LB; Tel 020 7799 8900; email iea@iea.org.uk

Forrest Capie, Cass Business School and Geoffrey Wood, Cass Business School and the University of Buckingham

Contents

Summary

- Banks need to keep both capital and liquidity against unforeseen events.
- During the early-mid nineteenth century, there were a number of banking crises. Banks responded by holding higher levels of capital.
- An analysis of bank capital shows that they adjusted their capital ratios according to the risks that they were taking and that they were well capitalised in comparison with the standards set by regulators under the Basel I and Basel II approaches.
- Indeed, when bank capital levels became very thin after the Second World War, banks were prevented by the Bank of England from raising more capital, despite their appeals to the Bank.
- During this long period of prudent management of the banking sector, there was no clear expectation that the state would have stepped in to save an insolvent bank in Britain.
- Capital regulation is relatively recent and led to banks trying to game the rules contributing to the complexity that was created in the banking system.
- An analysis of history demonstrates that capital regulation is not necessary if banks are not underwritten by the state.
- The principle of many of the reforms to banking law and regulation currently being proposed or implemented is correct. That principle, which should be at the heart of regulatory reform, is that banks should be wound up in an orderly way if they fail.
- The whole apparatus of bank capital regulation which has done so much to make the banking system more opaque should be abandoned. Attempts by the British government to require large banks to hold very high levels of capital are misguided.

Introduction

Banks have in essence changed little since they emerged from being mediaeval goldsmiths. Whatever other activities they have engaged in, their key business remains borrowing and lending; that is the business that makes them crucial to the economy. When engaged in that business they are crucial in two ways. They supply a part – in modern economies, by far the greater part – of the stock of money. And they transfer funds from lenders to borrowers – they act as financial intermediaries.

When engaged in borrowing and lending, they need to hold both capital and liquidity.

Capital comprises funds the bank actually owns. It can be provided by the bank's shareholders, or, depending on the corporate form, the partners in the bank or even by its sole owner. Such funds are needed because regardless of how well the bank is run, how well it treats its customers, and how well aware it is of its responsibilities to them, now and again it will lose money on a loan. Some or all of what it has lent will not be paid back. That, though, is no excuse (in either morality or law) for not repaying the people who have lent the bank money; so the bank needs some funds of its own to make up what is needed to repay depositors when that is necessary.

Liquidity is in some ways a rather trickier concept. It can first of all, and most easily, be thought of as cash the bank keeps in its own vaults. Some cash is needed because while payments into the bank match withdrawals most of the time, sometimes they fall short. Again, the bank is obliged to pay out what customers demand, so to avoid default and consequent closure they need some cash in hand.

How much they need depends, it is worth emphasising, on the attitude of the relevant central bank to supplying cash. The central bank may be willing to lend short term on non-penal terms on a wide range of securities as collateral. In such circumstances an individual bank need hold little liquidity. This is why the recent notion that there must be international regulation of the liquidity held by banks seems to be at least not completely thought through, and perhaps even foolish. But that discussion is for a future paper.

How much capital should banks hold to deal with the problem described above, and who should decide on the answer to that question?

Bank capital in Britain – the experience

Capital has always been important to British banks. Indeed, it was considered so important that the government and the Bank of England accepted the public interest argument that allowed the concealment of true profits and capital until as recently as 1970. Banks have always experienced a tension between having too much capital and too little. Strong capital positions are designed to give depositors confidence – indeed, in the 19th century, they were on occasions used as a competitive weapon to attract deposits. But the greater the capital the lower will be the return on capital and so there is a trade-off between depositor confidence and shareholder satisfaction. And of course the quality of the assets, i.e. the quality of what the bank has lent against or bought, is key to any calculation.

In the first half of the 19th century there were several hundred banks in England. Before 1826 these were all unlimited liability partnerships of no more than six partners. After 1826 joint-stock banking was permitted and banks gradually adopted that form.

In the periodic financial crises that appeared in the 19th century many banks failed, or suspended payment for a time[1], or merged, or were taken over.

There were no regulations as to what proportion of the balance sheet their capital or any other liability or asset might be. And across the middle two quarters of the century the regulations that were in place were gradually removed – such as the usury laws and compulsory unlimited liability. The first of these was removed in the 1830s and 1840s, the second in the late 1850s and early 1860s.

The banks therefore each had to find their own way to the appropriate balance sheet shape for each individual institution. Following the repeated financial crises in 1825, 1836, 1847, 1857 and 1866, the banks began cautiously, with very high capital/asset ratios and similarly high liquid assets ratios. But these gradually came down as trust and understanding developed. And after the 1870s, when it also became clear that the Bank of England had assumed the role of lender of last resort[2], there was an added reason for well-behaved banks to let their liquidity ratios fall slightly further.

The maintenance of ‘inner’ or ‘hidden’ reserves allowed banks to smooth their reported profits, reassuring depositors and shareholders by presenting a picture of financial soundness and prudent behaviour, and thereby contributing to financial stability. The practice of maintaining hidden reserves had been prevalent from the mid-nineteenth century - the Midland Bank, for example, had first established a hidden reserve in 1866[3].

[1] A modern counterpart to this is the right that building societies retained to require notice before withdrawal of funds. This right was rarely enforced, and many customers were probably not aware of its existence.
[2] For a discussion of this frequently misused term see Wood (2000).
[3] Sometimes too there was “callable capital” – an obligation on shareholders to provide more capital, in amounts related to their shareholding, should the bank need it.

By the beginning of the last quarter of the 19th century, the published capital ratios had settled at around 15 per cent with little variation across banks, and by the end of the century that figure had slipped to around 10-12 per cent. Discount Houses, specialised intermediaries which dealt with trade finance and also lent to banks, operated with much lower levels for the good reason that they had exceptionally detailed knowledge of their customers.

Bank capital in the 20th century before regulation

In the inflationary conditions of the First World War the ratios fell further, as much of the bank lending which led to the expansion of bank balance sheets was secured on government debt, then believed to be completely secure. Banks adjusted the capital that they held according to the risks that they were taking. In the years between the two world wars there continued to be remarkable stability in the banking sector, and, no doubt in consequence, the ratios slipped slightly further. In the 1920s and 1930s they had settled at around 7 per cent. The point needs stressing that English banks remained remarkably strong through these years and no doubt contributed to the stability in the economy and the avoidance of a great depression with the same characteristics – including bank failure – as the US depression.

In the Second World War the banks' capital ratios fell sharply. They were around 3 per cent. Their balance sheets expanded with government debt while private lending fell away. But as the ratios fell so too did the risk since the bulk of the balance sheet was made up of gilts. This continued to be the case in the long period of adjustment following the war. In fact the ratios reached their all-time lows in the 1950s when they were down to between 2 and 3 per cent. Raising capital after the war was not easy with the restrictions placed by the Capital Issues Committee (a committee which restricted access to capital markets by private sector borrowers so as to ensure there was always ready finance for the government).

This particular restriction on the banks began to be troublesome and bank chairmen spent a lot of time in the 1950s lobbying the Bank of England for support in allowing them to raise new capital. A note for the Chief Cashier made the problem clear: '…it will be seen that the capital structure of the Clearing Banks is far from sound…At present it is clear that in times of trouble they must either put footnotes in their balance sheets – which we deplore – or lean on us for financial aid which would be disastrous…The banks, [if] freed from restriction, should pursue energetically the implementation of a programme which, for good reasons, is long overdue' (quoted in Billings and Capie 2007 p.145).

It cannot be overemphasised, particularly in view of current discussions, that the banks themselves were keen to hold more capital and there was no desire to provide them with central bank or government financial assistance. It was government regulation that restricted the issuance of more capital.

As normality was restored and private lending came back to the position it had formerly occupied, and gilt holdings were correspondingly reduced, the capital/asset ratios did slowly come back to around 4 or 5 per cent in the 1960s. Note that there were still no regulations on capital of any kind in place; this raising of capital ratios was of the banks' own choice.

It should also be emphasised that all the figures quoted above are those presented to the public in the banks' balance sheets. It was well understood that the banks had further reserves arising from the use of prudent accounting standards. These hidden reserves did not just, as described above, allow a smooth picture of business to be presented. They also meant that the banks were, in fact, a good deal stronger than was presented.

When the true positions are calculated all of the figures given above can be raised by at least one percentage point so that at the lowest point of the 1950s capital ratios would be closer to 4 per cent. When risk weightings of the Basel type (discussed below) are applied the figures would become dramatically higher, reflecting the quality of the assets the banks held across most of this period. Thus the figures for the 1920s would show Basel-type risk-weighted ratios of around 14 per cent. Those of the Second World War would turn out to be the highest of all time being even higher than 14 per cent. And, in the 1960s, the ratios were of the order of 13 per cent.

As Lloyds' chairman commented in the 1950s: 'there is no rule of thumb method of deciding the size of the capital funds which a bank needs in order to carry on its business. The guiding principles are that the resources as a whole must be sufficient to provide absolute security for our depositors and the reserves sufficient to meet fluctuation in our trading from year to year…provision must… be made against the difficulties associated with the fluctuations in the market price of gilt-edged securities.'[4](1954)

Only one official report of the period considered capital explicitly and that was the *Prices and Incomes* report of 1967. It concluded: 'There do not appear to be any concerted views among the banks about the appropriate level at which these [reserves] should be maintained. The banks do however tend to consider their reserve requirements…in relation to total deposit obligations.' The banks in fact believed that they had achieved their desired capital position at the time of the report.

However, as inflation then took hold the banks were looking to raise their capital base further, since the money growth behind the inflation comprised mainly bank deposits, so bank balance sheets had grown commensurately. But that unfortunately coincided with the biggest stock market fall of all time to date in Britain from 1972 to 1974. The then-used main index fell from 533 in May 1972 to 160 in January 1974. Bank shares fared worse than most and some fell by as much as 70 per cent. It then of course became extremely difficult to raise new capital.

There was at that time no particular threat to the main retail banking sector[5] but there was a crisis in the secondary banking sector in the mid-1970s and that led to legislation in 1979. The Banking Act passed that year placed limits on individual exposures to ensure appropriate diversification. Exposures exceeding 25 per cent of capital required prior approval of the Bank of England. That marked the beginning of interference in bank operations. And soon after that, in the 1980s, the rules of Basel took over.

[4] Lloyd's Chairman's Report (1954 - Sir Oliver Franks was Chairman). Quoted in Billings and Capie, 2007.
[5] Even in the dire economic conditions of those years the main banks did not run through their voluntarily chosen equity capital holdings.

Basel rules – version one

Banks started to be regulated by laws in addition to normal company law at different times in different countries. The first hints of co-ordinated international regulation came in the wake of two 1974 bank failures – those of Bankhaus Herstatt and of Franklin National Bank. In large part due to the way in which regulatory authorities handled these failures[6], these relatively modestly sized banks caused considerable problems for other banks when they failed. This led to the formation of a standing committee of bank supervisors from the G10 countries; it has a permanent secretariat at the Bank for International Settlements in Basel – hence it is also known as the Basel Committee.

This committee started to concern itself with capital regulation in 1988. The 1988 Basel Accord established a set of "Capital Adequacy Standards" for internationally operating banks. This accord, known as Basel I, required banks to hold capital according to Basel risk asset rules. The ratio was:

Basel risk assets ratio = capital/weighted risk assets.

Capital was divided into tiers 1 and 2; tier 1 comprising mainly equity capital (which might be thought of as the first line of defence) and tier 2 representing supplementary capital. Tier 2 comprised instruments such as loan stock and subordinated debt.

It was recognised that risk assets were not homogeneous, but despite that recognition attention was paid only to credit risk – risk of default. Each asset held by a bank was assigned to one of five risk classes – put into one of five "risk buckets" in the inelegant terminology used. Each of these buckets had a different degree of risk weighting; the higher the risk, the higher the weighting, and the higher the capital required. The weighting was based on the generic nature of the borrower and no attention was paid to individual risks. Thus, for example, Rolls Royce would receive the same weighting as a newly started news agent and the government of Argentina would get the same risk weighting (zero) as that of the UK. This is perhaps surprising. Also worth remarking is the effect of the focus on credit risk alone. This meant that no funds had to be set aside to cover, for example, the effects of interest rate variations on the market value of long term debt.

It has been remarked (for example by Heffernan, 2005) that these and other anomalies, such as the fact that a lower risk weighting was given to an off-balance-sheet transaction with a business than to a straightforward loan to the same business, encouraged regulatory arbitrage - the use of some financial instruments to allow a reduction in capital without a corresponding reduction in risk. All that, of course, points to weaknesses in Basel I, and these and other problems led to Basel II. It is however useful to remark here – this is a point developed further below – that banks had no such incentives to get round the rules when, as described earlier, they themselves chose the capital ratio appropriate for their own business.

[6] Bankhaus Herstatt for example was closed without recognition of different time zones so, while the relevant authorities sensibly waited until markets had closed in Europe, the closure took place while markets were still open in New York.

Furthermore, even if businesses set their own capital levels and set them prudently, they would at the moment still have to have the amount of capital prescribed by the Basel rules. This 'regulatory capital' may be less than the bank's desired capital (sometimes called its 'economic capital') for some types of loan - for example, some government bonds- and too much for others - for example, mortgages with very low loan-to-value ratios. There was a tendency, therefore, for banks to decide on the prudent amount of capital they needed to hold given their balance sheet, and adjust the form of their lending, the assets they held or the extent of off-balance-sheet transactions to ensure that they had the right amount of economic capital but still exceeded the required regulatory capital. This encouraged complexity and much of the creation of the financial instruments that were implicated in the crash.

Some deficiencies with Basel I were recognised and, to some extent, tackled. In an amendment to Basel I, announced in 1996 and adopted by 1998, market risk – the risk of loss through changes in the market price of assets – was addressed. Banks were, subject to the approval of their regulator, allowed to use their own models to calculate market risk. These models were Value at Risk (abbreviated to VaR) models, and produced an estimate of the sensitivity of the value of a portfolio to market price movements to show how much a firm would lose for any movement in prices. In particular, they would show banks the probability that they could run through a given amount of capital.

If banks did not have an approved internal model they had to use the Basel "standardised approach". That was a "building blocks" approach. Four market risks were identified – interest rate, exchange rate, equity prices, and commodity prices – a capital charge was determined for each of these; and then these charges were added up. Note that because no relationship between the risks is allowed for, risk diversification, a classic principle of prudent banking (and indeed of prudent investment), is ignored and not rewarded in terms of the amount of regulatory capital that was required to be held.

It is not surprising that the deficiencies of Basel I led to further change. The changes they led to were contained in Basel II.

Basel rules – version two

A proposal was made for reform of capital regulation in 2001. This original "Basel II" proposal was subject to considerable adverse comment, and eventually a 'three pillar' approach was introduced. There were measures of credit risk, which allowed for previously neglected 'subtleties' such as recognising that some companies can be less risky than some countries, and the recognition of operational risk (risk arising from failure of a part of the bank's operations – computer failure stopping people getting their money is a good example). In addition to this more complex "risk pillar", there was a supervisory pillar and a market discipline pillar. These three pillars were supposed to support the structure of banking.

The risk pillar has already been described. What were the other two pillars? The supervisory pillar specified the responsibilities of national supervisors. These, in summary, were to ensure that banks measured their risks properly; that they encouraged review and updating of the way risks were measured; that they should encourage banks to hold above-minimum capital; and that they should encourage banks to restore capital to desired levels as soon as possible should banks experience losses. What is surprising about that list is not only that it all seems extremely obvious, but that it comprises functions that look like primary duties of bank management. The market discipline pillar encourages banks to disclose information on risk exposure, capital adequacy and methods of calculating capital requirements quarterly or semi-annually. This is all information which, if concealed or incorrect, could mislead markets.

As is well known, these three pillars did not support the structure of banking. There was a major banking crisis over a substantial part of the world. It would take too long to go deeply into why this happened, but every writer on the subject (a thorough survey is provided in Lastra and Wood, 2009, and a study of an individual episode can be found in "The Run on the Rock", the report by the Treasury Select Committee of the House of Commons on the failure of Northern Rock) identifies the same principle features: perverse incentives, complacent management and shareholders, inadequate evaluation of risk, and regulatory failure so gross as in some cases it deservedly could be described as incompetence[7]. What was the response to these failures?

Good examples are the Vickers Commission in the UK and the proposals of a team led by Paul Volcker in the USA. These proposals had much in common, despite the fact that they are often being contrasted.

[7] In view of this a reasonable question might be why so many regulators have survived in their posts, or even been promoted.

Vickers and Volcker

That they had much in common is not surprising, because they both were confronted with a similar muddled or, more kindly, two-part question. The two parts were essentially:

1. How could banks be prevented from failing, and, much more pertinently as will be made clear below?
2. How they could be closed in an orderly fashion if they got into difficulties?

Interestingly, the two parts of the question led to the same answer – separate investment banking (dealing in markets, essentially) from the traditional banking activities of borrowing and lending. They arrived at this answer by different routes, of course. There was an idea that investment banking, described by the naïve as casino banking, was more risky than normal banking. This of course was quite the opposite of what had been the case in the UK, where the notable failures were of normal commercial banks (RBS and HBOS) which threatened the stability of their investment banking operations.[8]

The Volcker Rule in its purest form prohibits deposit-taking banks from engaging in proprietary trading and investment in private equity or hedge funds. The Dodd-Frank Act (which embodied the Volcker Rule) enacted a slightly modified form of the rule which permitted limited investment in private equity or hedge funds (up to 3 per cent of Tier 1 Capital) and allowed trading for purposes of hedging, market making and liquidity management.

The main difference between the Volcker Rule and the proposal of the Vickers Commission relates to the location and height of the fence that divides the different banking activities. The Volcker Rule seeks to ban completely what is seen as the most risky sort of trading activity from being carried out in a deposit-taking bank, but allows most investment banking activity to remain. The Vickers ring-fence seeks to insulate the core activities of the deposit-taking bank from a wider range of risky or non-essential activities, but via a split rather than a ban.

The Independent Commission on Banking (Vickers) report argued that the advantage of a ring-fence over the Volcker Rule is that banning proprietary trading would have only a modest impact in the UK where this is a relatively limited activity. In contrast, ring-fencing should facilitate resolution of both the ring-fenced and non-ring-fenced entities.

Volcker has questioned elements of the UK ring-fence, raising concerns about the permeability of a ring-fence and the ability to maintain true independence in a crisis[9] – in which case the ring-fence

[8] It was also inconsistent with the results of proper study of US experience. See Benston (1990).

[9] Volcker has argued that it is unlikely that a bank would risk its reputation and name by letting one part of it fail unless it really had insufficient funds to pursue any other course. On these and other grounds the UK Commission on Banking Standards has proposed “electrifying” the ring fence. By that is meant including in legislation a provision allowing regulators to compel Volcker separation on a bank which persist in testing the robustness of the fence.

would introduce complexity and cost without even getting rid of proprietary trading risk. The Volcker Rule also has the advantage over Vickers of keeping together what some banks (especially in Europe) regard as complementary services aimed at both small and larger clients.

The real objective

What is notable, praiseworthy and too little emphasised about the recent Volcker and Vickers proposals is that they are a retrograde step. They are gradually moving us back to a pre-regulation era.

That may seem odd, for both involve telling banks what they cannot do. The first thing to note about that is that they are in effect moving away from prescriptive, Roman law type regulation – that you can do anything that is allowed – to a liberating, common law type approach – you can do anything that is not forbidden. That encourages innovation, rather than businesses having to wait for permission to do things, and innovation is essential for economic growth.

Now, at this point it may be said that innovation got us into the recent banking crisis. Is making that easier really a good idea? That leads to the second really notable aspect of the Vickers–Volcker proposals: the division of banks. That division should be thought of not as making banks failure-proof, but rather as making it possible for them to fail in a fashion as orderly as any other firm. For that reason the proposal of the Vickers Commission that British banks should hold very much higher levels of capital, perhaps higher than the EU will allow, seems to us to be at best beside the point. The support of the British government, the EU and the Bank for International Settlements for higher capital requirements to prevent failure is likewise misguided.

What we need can best be described as “better banking”. We need a system that treats customers as prospective long-term clients, and that operates so that the firms in the system seek to be stable. But at the same time it is essential that the system allows individual firms to fail. If it does not, there is no incentive for prudence beyond that provided by virtue, and, at least as importantly, a system which protects incumbents from failure is one where there will be no new entrants.
So, to summarise, we want “responsible” treatment of customers, prudent banks at the heart of the financial system, and the possibility of failure to help ensure these things and to encourage new entry and innovation.

Did we ever have such a system? Can we find it by going back to the past? It seems to have existed in Britain in the 19th century, but banks were in general smaller, the individuals who used them relatively much better off than those who use banking services nowadays and failure was allowed. More recently in Britain the system was stable, but it achieved that by being a highly restrictive cartel. Neither of these past situations seems a plausible choice now. A cartel would be undesirable, and probably illegal. And banks as small as the average size in the 19th century would be pretty small, and could not offer a wide range of services to businesses.

So what do we want?

What is of the essence is that banks can fail in an orderly fashion. The commonly asked question of which type of separation makes failure less likely is therefore the wrong question. The right question to ask is: which type of separation makes insolvency easier, if failure does happen? It is widely believed that Volcker separation makes failure less likely. That is inconsistent with the evidence, and in fact not what Volcker argues. He has observed that such a separated bank would be easier to liquidate – indeed, if only for the reason that there are no obvious triggers for the orderly closure of an investment bank.

Failure must be possible. It must be possible to close a bank down, with shareholders losing, if necessary, everything, and other losses borne in order of seniority, with management having to take its chances on the job market. The system must be designed, and seen to be designed, to allow this. The proposal for 'living wills', essentially a manual on how each bank works to guide a liquidator in his work, is also framed with the end of making orderly failure possible[10].

Talk of separation of different types of bank recognises the importance of failure. If that is recognised and implemented, the next and most obvious step is to recognise that banks, once they are forced by the possibility of failure to take responsibility for their actions, are best placed to judge their own capital requirements. Under this new system, if they do not do so sensibly, they will not be in the business of banking for long. The whole apparatus of bank capital regulation can then be abandoned. We would go back to the years of banks choosing their own capital ratios: the system which produced stability in British banking for more than a century. It will be a system where there is stability but not stasis. It was very clear from the early part of this paper that banks chose their own capital requirements rationally when they were allowed to do so (and the markets forced them to do so). Indeed, the distortion of their activities and the use of instruments designed to reduce capital was subsequent to the development of regulatory capital requirements and was, at least in part, caused by the imposition of such requirements.

[10] The suggestion that there be a special class of bond which can be called on and used in event of failure – sometimes called bail-in bonds – may also be viewed as contributing to this as they could be used to pay off the bank's creditors. They can also of course be thought of as a way of helping the bank to keep going after a shock.

References

Benston, George J. (1990) *The Separation of Commercial and Investment Banking: the Glass Steagall Act revisited and reconsidered*, Oxford University Press, London and Oxford.

Billings, Mark, and Forrest Capie (2007) 'Capital in British banking, 1920-1970', *Business History* vol. 49, No.2, March, pp. 139-162.

Heffernan, Shelagh (2005) *Modern Banking*, John Wiley and Sons Limited, Chichester.

House of Commons Treasury Committee, (2008) *The Run on the Rock*; Fifth report of Session 2007-08, Volume , HC – 56-1,Published by authority of the House of Commons: London The Stationary Office Limited

Lastra, Rosa, and Geoffrey Wood (2009), 'The Crisis of 2007 – 2009: Nature, Causes, and Reactions', *Journal of International Economic Law*, Vol. 13, No. 3, September 2010, pp. 531-550.

National Board of Prices and Incomes Report 34 on Bank charges. Cmnd 3293, HMSO 1967.

Wood, Geoffrey (2000) 'The Lender of Last Resort Reconsidered', *Journal of Financial Services Research*, 18(2), pp. 203 – 227.

Institute of Economic Affairs
2 Lord North Street
London
SW1P 3LB

www.iea.org.uk