Systemic risks and the ‘too-big-to-fail’ problem

Alan D. Morrison*

Abstract The financial crisis of 2007–9 resulted in state intervention in financial markets around the world, and the state became a major shareholder in many banks. While state bailouts were politically sensitive, policy-makers had little alternative but to supply funds to financial institutions that were viewed as ‘too big to fail’, or TBTF. In this paper, I review the history of, and the rationale for, the TBTF policy. I argue that, from a policy perspective, the most important costs of the TBTF problem are incurred ex ante, in the form of distorted incentives that arise as a consequence of distortions to the capital markets, and to the choice of banks’ scale and scope. I argue that it is impossible credibly to withdraw the TBTF policy, and, hence, that it should be managed so as to minimize the costs of these distortions. In this context, I discuss the role of policy in institutional design, in the restriction of bank scope, and in designing appropriate capital regulations.

Key words: too big to fail, bailout, bank scope, bank scale

JEL classification: G21, G28

I. Introduction

The financial crisis of 2007–9 saw significant state intervention in financial markets all over the world. In the United Kingdom, the Treasury stepped in to support fragile and failing banks. As a result of its recapitalization of the Royal Bank of Scotland (RBS), the State acquired 83 per cent of the rights to RBS’s cash flows, and 68 per cent of the voting rights; the recapitalization of the Lloyds Banking Group left the government with 41 per cent of the Group’s voting and cash flow rights. The Treasury lent money to London Scottish Bank, Dunfermline Building Society, and the Icelandic banks operating in London so that they could repay customer deposits, even though all of these banks were insolvent. A Special Liquidity Scheme was introduced in April 2008, which enables banks to swap their illiquid assets for marketable Treasury securities.1

In all, the National Audit Office (2009) identified £850 billion of support to the UK banking sector, provided as a mixture of guarantees, shares, and loans to banks. It estimated the total UK taxpayer exposure to banking sector losses at £955 billion. Its most recent

* Saïd Business School and Merton College, Oxford, e-mail: alan.morrison@sbs.ox.ac.uk

1 See National Audit Office (2010) for details of these and other government support schemes.
doi: 10.1093/oxrep/grr020

© The Author 2012. Published by Oxford University Press.
For permissions please e-mail: journals.permissions@oup.com
estimate of the scale of the exposure is £512 billion. As of 1 December 2010, the taxpayer’s paper loss on RBS and Lloyds shares was £12.5 billion (National Audit Office, 2010).

The UK’s experience was not unique: taxpayers in many developed countries bailed out large banks. For example, a central pillar of the American response to the crisis was the Troubled Asset Relief Program, or TARP. TARP comprises 12 programmes, each of which supports financial institutions, asset prices, the US automotive sector, or financially impaired homeowners (Office of the Special Inspector General for the Troubled Asset Relief Program (OSIGTARP), 2011b, pp. 36–7). The financial institution programmes support weak financial firms. Two of the financial institutional support programmes assist undercapitalized institutions: these comprise the Capital Purchase Program (CPP), which purchases preference shares in banks, the Capital Assistance Program (CAP), which complements the CPP by providing capital to undercapitalized bank holding companies, and the Systemically Significant Failing Institutions (SSFI) Program, whose investments are intended to prevent the failure of AIG. The Targeted Investment Program (TIP) was used to support Citigroup and Bank of America Corporation. Finally, the Asset Guarantee Program (AGP) provided loss insurance on impaired mortgage and similar assets that were regarded as critical to financial market confidence.

One of the largest recipients of government aid in the United States was Citigroup. OSIGTARP (2011a) analyses the scale of Citigroup’s assistance, and the procedures by which it was dispensed. The government provided some Citigroup asset guarantees and invested $20 billion in Citigroup in exchange for Citigroup stock. J.P. Morgan’s March 2008 takeover of Bear Stearns was facilitated by the Federal Reserve’s provision of, first, a $30 billion discount window loan to J.P. Morgan, and, later, by a 10-year $29 billion loan to a newly established special purpose vehicle that acquired a portfolio of Bear Stearns assets (Financial Crisis Inquiry Commission, 2010, p. 21). As of 3 October 2010,$474.8 billion of obligations had been made by TARP programmes, including the AIG recapitalization of 14 January 2011; $469 billion obligations were created, of which $410.1 billion were spent on 26 January 2011 (OSIGTARP, 2011b, pp. 40, 41). The estimated cost of the TARP programmes to the US taxpayer is currently $25 billion (Congressional Budget Office, 2010).

The financial crisis was widely perceived to be the result of irresponsible lending practices and cavalier risk-taking by banks. The decision to invest taxpayer funds to stop the same banks from failing was unpopular. The first TARP legislation was actually rejected by the House of Representatives on 29 September 2008 and, after TARP was accepted by Congress, the Bush administration had to fight to persuade Congress to release the last $350 billion of earmarked funds.6

---

2 Specifically, they support banks, bank holding companies, and also other financial institutions that are deemed critical to the financial system.

3 The SSFI programme could in theory be extended to other non-bank firms whose health is critical to the wider economy.

4 According to OSIGTARP (2011a), Citigroup accepted the government’s terms with some reluctance, believing them to be excessively harsh.

5 In this context, obligations are definite commitments of funds that create a legal liability (OSIGTARP, 2011b, p. 41).

Notwithstanding the opprobrium that bankers attracted, policy-makers felt that they had little alternative but to extend support to large financial institutions. Ben Bernanke, Chairman of the Federal Reserve System, identified the central problem when he acknowledged in a speech at the Council on Foreign Relations (Bernanke, 2009) that: ‘in the present crisis, the too-big-to-fail issue has emerged as an enormous problem.’

The ‘too-big-to-fail’ (TBTF) problem was indeed a serious concern during the financial crisis. Mitigating its effects is one of the most critical challenges facing policy-makers and legislators today. In this essay I examine the genesis of the problem. I then outline its consequences, before discussing some possible policy responses. I conclude by summarizing the key points of my argument.

II. The TBTF problem

The TBTF problem is now a widely acknowledged issue in banking regulation, but the problem was formally articulated relatively recently, in 1984. The phrase was used in a comment by Congressman McKinney when C. Todd Conover, the US Comptroller of the Currency, testified before Congress on the $4.5 billion government rescue of the money centre bank, Continental Illinois. The rescue package protected all of Continental’s creditors, even though only 10 per cent of its $30 billion deposits were insured (Financial Crisis Inquiry Commission, 2010). Congressman McKinney remarked that: ‘[w]e have a new kind of bank. It is called too big to fail, and it is a wonderful bank.’

In the wake of Mr Conover’s testimony, a further 11 banks were identified in the Wall Street Journal as TBTF. In fact, it is quite possible, although not inevitable, that the shareholders in TBTF banks may be allowed to lose money. The phrase applies more precisely to institutions whose uninsured creditors and depositors are protected by the regulatory authorities and, hence, by the taxpayer (US Treasury, 1991).

When a TBTF bank receives taxpayer support, it is because the regulatory authorities believe that its failure could impose severe negative externalities upon other economic actors. Hence, to use a currently popular expression, a very large bank may be a ‘systemically important financial institution’, or SIFI. The two expressions are not synonymous; while many SIFIs are systemically important by virtue of their size, size is neither a necessary nor sufficient condition for a firm to be a SIFI (Thomson, 2009). A large bank’s systemic importance could derive from a number of sources. For example, the failure of a large bank might undermine confidence in other large institutions with apparently similar investment portfolios and, hence, trigger runs on those institutions (see, for example, Chari and Jagannathan (1988) and Chen (1999)). Similarly, a financial institution that is linked to many other institutions via lending and borrowing relationships, or because it supplies vital payment services to them, may cause those firms to fail if it fails itself (for example, see Rochet and Tirole (1996), Freixas and Parigi (1998), Freixas et al. (1999), and Allen and Gale (2001)).

Two papers examine regulatory reputation as a rationale for forbearance, although neither is directly concerned with the TBTF problem. In Boot and Thakor (1993), regulators are excessively lenient in dealing with failing banks, because they wish to protect their
reputations. In contrast, Morrison and White (2010) argue that such reputational concern can be socially justified. The reason is that the failure of one bank may cause depositors to lose confidence in other banks supervised by the same regulator, and so cause a socially damaging run. When this is the case, the supervisor’s concern for its reputation may be socially justifiable.

It seems that bank borrowers frequently struggle to find alternative sources of funds when their main lender fails. This is true even when the bank failure does not reflect bad lending decisions or macroeconomic weakness: Ashcraft (2005) finds persistent reductions in local GDP after the closure of Texan banks as a result of holding company failure precipitated by problems elsewhere in the United States. Hence, bank failure has economic costs even when it does not result in the failure of other financial firms.

The arguments above are precisely in line with those used by Ben Bernanke to justify the Fed’s rescue of Bear Stearns to the Joint Economic Committee. Normally, the market sorts out which companies survive and which fail, and that is as it should be. However, the issues raised here extended well beyond the fate of one company. Our financial system is extremely complex and interconnected, and Bear Stearns participated extensively in a range of critical markets. With financial conditions fragile, the sudden failure of Bear Stearns likely would have led to a chaotic unwinding of positions in those markets and could have severely shaken confidence. The company’s failure could also have cast doubt on the financial positions of some of Bear Stearns thousands of counterparties and perhaps of companies with similar businesses. Given the current exceptional pressures on the global economy and the financial system, the damage caused by a default by Bear Stearns would have been severe and extremely difficult to contain. Moreover, the adverse effects would not have been confined to the financial system but would have been felt broadly in the real economy through its effects on asset values and credit availability.

Notwithstanding these remarks, it is very difficult to know for sure that a bank’s failure would have a seriously deleterious effect upon the rest of the financial sector. OSIGTARP (2011a) argues that the decision to rescue Citigroup was ‘startlingly ad hoc’, based ‘as much on gut instinct and fear of the unknown as on objective criteria’.

Of course, during a financial crisis it is almost inevitable that some decisions will be made on the hoof, as new information emerges. And the immediate costs of an incorrect closure decision are probably higher than those of an unnecessary bailout. In a recent speech, the Kansas City Federal Reserve Bank Chairman Thomas Hoenig (2010) remarked, ‘what Federal Reserve chair, FDIC chair, or secretary of Treasury, would risk an economic collapse when, by making some creditors whole, the panic might be stopped?’ The Financial Crisis Inquiry Commission (2011, p. 433) points to the same economic and political calculus, noting that ‘if you bail out AIG and you’re wrong, you will have wasted taxpayer money and provoked public outrage. If you don’t bail out AIG and you’re wrong, the whole financial system collapses.’

In short, there are economic reasons for believing that the costs of allowing certain financial institutions to fail are too high to bear. Moreover, the political calculus points towards bailouts.

---

It is probably inevitable that some institutions will remain TBTF. If this is the case, it is important that we understand the implications of an implicit or explicit TBTF policy.

III. Implications of a TBTF policy

(i) Cost of capital

Bailing out a bank is very costly. In addition to the direct administrative expenses of a bailout, governments need to raise fresh bank capital, and they borrow or raise taxes to do so. As a result, funds are diverted from alternative uses, potentially at a significant opportunity cost (hence, claims that US taxpayers will make a $12.3 billion profit on the government’s Citigroup investment are misleading: the profit net of opportunity costs is surely much lower). But the direct costs of a bailout are weighed against the alternatives when they are incurred, and bailouts are chosen over alternatives that are in expectation more expensive. In short, bailouts are probably necessary and cost-effective when they are performed. The real cost of a TBTF policy comes from its effect upon the day-to-day business decisions that bankers make.

The incentive effects of a TBTF policy are intuitively clear. When bankers do not expect to bear the costs of bank failure, they are likely to assume higher risks. OSIGTARP (2011b, p. 7) characterizes the risk assessment process in a TBTF institution as one of ‘heads I win, tails the Government bails me out’.

In practice, though, casual empiricism suggests that few bankers think in precisely these terms. It seems more likely that they were guided to take excessive risks by the invisible hand. As noted above, while shareholders in TBTF institutions lose heavily upon their failure, uninsured depositors and bond holders do not. As a result, they have few incentives to monitor the actions of the banks to which they lend. Moreover, even if they were completely aware of the risks that TBTF banks were taking, their creditors would not reflect those risks in the price of debt, because they are not exposed to them.

In short, market discipline is attenuated in TBTF institutions, because their cost of debt finance is relatively risk-insensitive. Even a manager who is unaware of the existence of a TBTF policy will therefore take advantage of it, by assuming high levels of cheap debt, and then assuming risky positions that benefit his shareholders, who have convex pay-offs, and therefore welcome risk.

Bankers’ own comments about the cost of capital suggest that this story explains some of their actions. There is a widespread belief in the banking sector that equity is a ‘more expensive’ source of funds than debt. We have known for about 50 years that it is not:

---

9 For example, Kane (1990) identifies significant costs of generous bailouts of Savings and Loans institutions in the United States. He places the cost of inefficient funding of the Resolution Trust Corporation (RTC) at $2 billion per year; he argues that this figure is dwarfed by the costs of inefficient salvage, arising because the RTC was unable effectively to appraise, manage, and sell impaired assets.


11 For example, see ‘Pandit and King clash over Basel III’, Francesco Guerrera and Daniel Pimlott, Financial Times, 25 October 2010, in which the chief executive of Citigroup claims that equity is more costly than debt.

12 Modigliani and Miller (1958, 1963); see Admati et al. (2010) for a recent discussion of the cost of capital in the context of the banking sector.
cost of capital depends upon the use to which it is put, and not upon the type of security that is sold to raise it. Market distortions, such as implicit government bailouts, ineffective corporate governance, and the tax rebate on corporate debt, mean that, in some circumstances, a judicious choice of capital structure can impose the costs of poor performance or of failure upon third parties. Such capital structures certainly appear cheaper to bank shareholders, but their total social opportunity cost is no lower than that of alternative capital structures. This is an argument for addressing the root distortions, and not for helping bankers to take advantage of them though an appropriate choice of capital requirements.

There is more rigorous academic evidence in support of the hypothesis that the TBTF policy artificially lowers the cost of bank funds, and so distorts resource allocation. For example, after the Wall Street Journal identified 11 banks that were TBTF after the 1984 bailout of Continental Illinois, their share prices rose 1.3 per cent (O’Hara and Shaw, 1990). After the bailout, bank bond spreads ceased to reflect bank business risk (Avery et al., 1988), and the relationship between bond spreads and ratings for the 11 TBTF banks flattened after the TBTF policy was revealed, in line with a reduction in risk-sensitivity (Morgan and Stiroh, 2005). Moreover, Boyd and Gertler (1993) show that, after 1984, larger banks assumed more risks than smaller ones, which were not underwritten by the TBTF policy. However, Laeven and Levine (2009) demonstrate that the extent to which bank risk-taking is exacerbated by regulation depends upon their governance. For example, they demonstrate, using an international sample of privately held banks, that deposit insurance is associated with risk-taking only when banks have a sufficiently large equity holder to act upon the associated incentives. Their work suggests that incentives for banks to expand might depend upon their governance structure, as well as upon whether large banks are viewed locally as TBTF.

(ii) Scale expansion

If TBTF status yields benefits then one would expect banks that achieve it to appreciate in value. Kane (2000) finds evidence in support of this hypothesis. He examines bank ‘mega mergers’, that involve very large banks, and finds that, in contrast to other mergers, the stock of acquiring firms rises when such a merger is announced. Penas and Unal (2004) find that bond prices react positively to bank mergers between 1991 and 1997, and attribute this value increase in part to the acquisition of TBTF status (as well as some diversification gains and the realization of some synergies). Indeed, it appears that precisely because a TBTF premium exists, banks are prepared to pay more for mergers that generate it: using merger data from 1991–2004, Brewer and Jagtiani (2009) show that banking organizations were willing to pay more for mergers that would push the post-merger organization over a $100 billion asset threshold, which, they argue, was sufficiently large for a bank to be viewed as TBTF. They estimate that at least $14 billion of added premia were paid for such deals; both the stock and the bond markets reacted positively to these deals.

The evidence of the last paragraph suggests strongly that, as a consequence of the TBTF policy, banks are expanding for reasons other than the realization of scale efficiencies. Berger and Mester (1997) argue that, at least in the United States, the technologically optimal size of a bank is about $25 billion. But many banks are far bigger than this. Demirgüç-Kunt and Huizinga (2010) identify 30 listed banks worldwide whose liabilities exceeded half of their country’s GDP, and report that 12 have liabilities in excess of $1 trillion.

Banks have expanded very rapidly in the last two decades. DeYoung (2010, Table 31.3) presents data on the scale of the 10 largest commercial banks in the United States for 1988,
1997, and 2007. The average size of the largest three banks increased by a factor of 2.3 between 1988 and 1997, and by a factor of 3.65 over the succeeding 10 years. All of the 10 largest banks had (nominal) assets of more than $100 billion in 2007, while only Citicorp exceeded this figure in 1988. If anything, the European experience has been even more extreme. Goddard et al. (2010, Tables 32.1–32.6) present data on the size of the largest European banks. Of the largest US banks, only Citigroup has more assets than any of the five largest European banks. The banking sectors of European countries expanded very rapidly over the two decades prior to the 2008 crisis. For example, the largest bank in Spain in 1986 was Banco Central, with assets of $28 billion; in 2006, it was Banco Santander with assets of $1.1 trillion. While the rate of change was slower in other major European economies, it was nevertheless extremely rapid.

Of course, recent rapid expansion in the banking sector could reflect new economies of scale. Recent technological advances may have pushed the efficient scale for a bank beyond Berger and Mester’s (1997) $25 billion estimate. For example, DeYoung (2005) argues that internet banking generates much higher economies of scale than traditional forms of banking. Nevertheless, the evidence on the price effects of bank mergers suggests that, in part, banks are being guided by a rational price mechanism, distorted in turn by a TBTF policy, to assume a socially excessive size. An opposing perspective is presented by Demirgüç-Kunt and Huizinga (2010), who use evidence from an international sample of banks to demonstrate that, when banks are very large relative to their national economy, their funding costs are more sensitive to risk proxies, despite TBTF subsidies. Demirgüç-Kunt and Huizinga interpret these data as evidence that some banks are too-big-to-save, rather than TBTF. Of course, this begs the question of why such banks emerge: their returns on assets appear small so, in line with evidence presented above, they do not appear to expand in pursuit of economies of scale. Demirgüç-Kunt and Huizinga suggest that the motivation for expansion may be entirely managerial, and hence that it is evidence of poor corporate governance.

Lastly, bank scale expansion leads inevitably to increased banking sector concentration. This may serve to increase the probability that creditors assign to bailout in failure states: Acharya and Yorulmazer (2007) argue that regulators are more likely to bail out banks when many fail together, because they cannot then rely upon survivors to acquire smaller failed banks. This problem, which they dub ‘too-many-to-fail’, is more likely to arise when there are fewer banks.

(iii) Scope expansion

Banks have expanded the scope as well as the scale of their operations in the last decade and a half. The 1933 Glass–Steagall Act outlawed universal banking in the United States: that is, it made it illegal for a deposit-taking institution to participate in the securities business. The Act was inspired by the belief that commercial banks were using their investment banking affiliates to underwrite security issues by their low-quality borrowers, and so to ensure that their loans were repaid. It now seems that there was no merit in this argument.\footnote{See in particular Kroznerr and Rajan (1994) and Puri (1996). The passage of the Act and the academic literature that examines its consequences are surveyed by Morrison (2010).}

With the benefit of hindsight, it is possible to identify an unanticipated positive consequence of the Glass–Steagall Act. Americans did not have deposit insurance when
the Glass–Steagall Act was passed, and the Act was not designed to manage the incentive effects of government support of the banking sector. Nevertheless, this appears to have been an unintended consequence of the Act: after its passage, it was impossible for the securities arm of a commercial bank to sustain losses so large that they threatened the viability of the commercial bank, and so forced a taxpayer-funded bailout. The 1999 Gramm–Leach–Bliley Act repealed the Glass–Steagall Act, so that universal banks could emerge in the United States. It is possible that the emergence of universal banking in the United States extended the state safety net from commercial banking to other types of financial firm. On the other hand, it is also possible that the effect of deposit insurance creep is outweighed by the diversification benefits of universal banking, which render failure, and hence bailout, less likely.

The evidence on this point is rather inconclusive but, prior to the crash, it appeared weakly to suggest that universal banking did not exacerbate the incentive effects of a TBTF policy. Universal banking has been the norm in Europe for over 100 years (Morrison, 2010), and there was little evidence prior to the crash that it extended the reach of the deposit insurance safety net. Within the US, there is no evidence that the riskiness of commercial banks was affected by the 1987 introduction of so-called Section 20 subsidiaries of commercial banks with the ability to perform some limited securities underwriting (Cornett et al., 2002). Simulation of portfolios of combined commercial bank loans and insurance company investments reveals diversification effects (Allen and Jagtiani, 2000). However, such simulations do not necessarily predict what would actually happen were the simulated firms actually to merge: Freixas et al. (2007) argue that a universal bank has increased incentives to assume risk, so as to overcome diversification effects and profit from state subsidies. Moreover, the evidence for scope economies in banking is rather weak. Although Barth et al. (2000) and Berger et al. (2000) make a case for economies of scope in universal banks, Allen and Rai (1996) and Vander Vennet (1999) find only limited evidence of scope economies in European banks, and Hughes and Mester (1992) find evidence of diseconomies of scope at larger banks. Nevertheless, Cyberto-Ottone and Murgia (2000) find that scope expansion can raise shareholder wealth. If there is little evidence for scope economies in banking then Cyberto-Ottone and Murgia’s findings suggest that scope expansion may simply take advantage of the TBTF policy.

The 2007–9 financial crisis provides stronger support for the hypothesis that universal banks increase the scale of the TBTF effect. Government support was extended to a large number of non-banking firms. Part of the problem here appears to have been the sheer complexity of the firms in question. For example, Michael Foot, formerly head of banking supervision at the Bank of England, told the 2009 House of Lords inquiry into Banking Supervision and Regulation that some banks are now ‘too-complex-to-manage’, and remarked that ‘I used to look at Citibank and I wondered how any group of human beings could actually run that entity’ (House of Lords, 2009, paras 231 and 43, respectively).

A non-banking example of extreme complexity in the financial sector was AIG, which sat at the heart of a complex network of over-the-counter derivatives, including credit default swaps. AIG was regulated in the United States by the Office of Thrift Supervision, which appears to have been insufficiently technically adept to supervise the firm (Financial Crisis Inquiry Commission, 2011, ch. 18). When AIG’s problems with short-term funding and collateral demands became severe in September 2008, no one was able to understand what the consequences of its failure would be. Irrespective of its ultimate resolution, the consequences of this uncertainty could have been sufficient to trigger a general market
panic. Faced with this danger, and unable in the time available accurately to assess AIG’s counterparty risk exposures, the Federal Reserve arguably had little choice but to support AIG. It lent $85 billion to AIG; ultimately, taxpayer funds committed to AIG reached $182 billion.

Market participants may have suspected before the financial crisis that non-deposit-taking financial firms would be beneficiaries of the TBTF policy. Mergers supported during the crisis by the US federal government and by American bank regulators, such as the acquisition of Merrill Lynch by Bank of America and of Bear Stearns by JPMorganChase, confirmed this suspicion. The consequence in the future is likely to be a greater distortion of incentives in securities markets firms. An early indication of this fact was the reaction of stock prices to the announcement of TARP fund injection into 10 large banking organizations on 14 October 2008. Share-price reactions were significantly positive, even for banks not rescued; the stock market reaction to the extension of the discount window to investment banks during the financial crisis was also positive (Brewer and Jagtiani, 2009).

(iv) Fiscal consequences

The TBTF policy is a response to perceived systemic risks. But, as the policy becomes more entrenched, and as it encompasses an ever larger banking sector, it may itself generate systemic fragility, because bailouts of institutions that are systemically important may undermine the credit worthiness of the government. If this is the case, bank regulators are caught between a rock and a hard place, forced to choose between the collapse of the financial system, and of the state’s finances.

Some recent evidence supports this contention. Brown and Dinç (2009) examine government bailout policies with respect to banks in 21 emerging market countries in the late 1990s. They find that banks are less likely to experience a government takeover if their peers are also fragile; Brown and Dinç interpret this as evidence of a ‘too many to fail’ problem, which arises because the state is unable simultaneously to support many banks.

Acharya et al. (2010) examine the effect of bank bailouts upon systemic risk. Sovereigns that announced bank bailouts during the financial crisis saw substantial increases in market perceptions of their default risk, as measured by the prices of credit default swaps (CDSs). Moreover, subsequent movements in CDS prices for sovereigns and large banks were subsequently very correlated. Acharya et al. also present preliminary evidence that, in addition to the transmission of risk from banking sector to sovereign borrower, increased sovereign default risk is transmitted to the banking sector via banks’ holdings of sovereign debt.

In related work, Demirgüç-Kunt and Huizinga (2010) examine bank valuation from 1991 to 2008, and CDS spreads between 2001 and 2008; the final year of the former time series covers 717 listed banks in 34 countries, and the final year of the latter contains 59 banks in 20 countries. They identify a positive relationship between bank size and its market-to-book ratio: that is, the market places a higher value upon assets held by large banks, suggesting that a TBTF policy obtains across the broad sample of countries that they examine. In addition, they use the liabilities-to-GDP ratio of a bank as an indicator of its systemic importance.

---

14 A credit default swap is a trade under which one party commits to make good any principal losses experienced by another party on a (possibly notional) position in a credit-risky position. CDS prices are therefore a good barometer of market perceptions of creditworthiness.
Demirgüç-Kunt and Huizinga show that banks that are systemically large by this measure respond more negatively to a deterioration in the public deficit.

Demirgüç-Kunt and Huizinga interpret their results as indicating that large banks became too big to save during the financial crisis, and they argue that these firms could raise their value by shrinking their balance sheets, or by divesting divisions. But this need not be the only story consistent with their data. The share prices of systemically large banks are positively related to bank risk in their dataset, and the CDS prices of these banks are negatively related to bank risk. In short, systemically large banks retain some of the characteristics associated in other studies with institutions that are TBTF. If this is the case, they remain at or near the head of the queue for state assistance during a financial crisis. It follows that, while the existence of extremely large banks has lowered the value of state assistance, it has had much less effect on the incentive to access that assistance.

Demirgüç-Kunt and Huizinga (2010) report data on the systemic importance of the banking sector, measured as the ratio of quoted bank liabilities to GDP. This ratio reached 9 at the end of 2007 in Iceland; at the end of 2008 it was 6.3 and 5.5 in Switzerland and the UK, respectively. The largest banking sector liabilities to GDP ratios at the end of 2008 are reported in Table 1.

In summary, while it is probably inevitable that modern governments operate a TBTF policy, such a policy generates significant costs. The most important of these relate not to the execution of the policy, but to the consequences that it has for ex ante banker behaviour. A TBTF policy reduces the risk sensitivity of bank financing, and so encourages risk-taking. It creates strong incentives for scale and scope expansion in banking; this expansion could in turn undermine the government’s fiscal position, and so undermine its ability to make any sort of effective response to a financial crisis. The financial crisis of 2007–9 demonstrated the importance of these effects, and many countries will experience the fiscal consequences of

<table>
<thead>
<tr>
<th>Country</th>
<th>Bank liabilities/GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switzerland</td>
<td>6.293</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>5.498</td>
</tr>
<tr>
<td>Belgium</td>
<td>2.916</td>
</tr>
<tr>
<td>France</td>
<td>2.737</td>
</tr>
<tr>
<td>Netherlands</td>
<td>2.469</td>
</tr>
<tr>
<td>Ireland</td>
<td>2.393</td>
</tr>
<tr>
<td>Denmark</td>
<td>2.330</td>
</tr>
<tr>
<td>Singapore</td>
<td>2.266</td>
</tr>
<tr>
<td>Australia</td>
<td>2.132</td>
</tr>
<tr>
<td>Sweden</td>
<td>1.982</td>
</tr>
<tr>
<td>Canada</td>
<td>1.799</td>
</tr>
<tr>
<td>Spain</td>
<td>1.749</td>
</tr>
<tr>
<td>Japan</td>
<td>1.657</td>
</tr>
<tr>
<td>South Africa</td>
<td>1.625</td>
</tr>
<tr>
<td>Greece</td>
<td>1.482</td>
</tr>
<tr>
<td>Italy</td>
<td>1.432</td>
</tr>
<tr>
<td>Israel</td>
<td>1.377</td>
</tr>
<tr>
<td>Germany</td>
<td>1.350</td>
</tr>
<tr>
<td>Greece</td>
<td>1.301</td>
</tr>
<tr>
<td>Austria</td>
<td>1.251</td>
</tr>
</tbody>
</table>

Source: Demirgüç-Kunt and Huizinga (2010, Table 3).
their TBTF policy for many years. One of the themes of recent policy discussions has therefore been the search for ways to mitigate the most severe consequences of the TBTF policy. The following section examines some of these.

IV. Managing the TBTF policy

In some situations, it is probably inevitable that large fragile banks will receive government support and, hence, it is probably inevitable that *ex ante* banker incentives will be distorted to some extent. Nevertheless, banking policy should aim as far as possible to contain the adverse consequences of the TBTF policy. I argued in section III that many of these consequences follow inevitably from the effect that a TBTF policy has upon large banks’ cost of capital. This effect can be reduced by ensuring that the bank’s uninsured claimants have only a low likelihood of receiving financial support after a failure. In this section, I examine three ways of accomplishing this: by designing better institutions and, in particular, better insolvency procedures; by reducing bank scope below the size at which they are TBTF; and by increasing bank capital requirements.

(i) Institutional design

The effect of a TBTF policy upon the cost of capital, and, hence, upon banker incentives, depends upon the regulator’s ability to intervene in a financially distressed large bank without compensating shareholders and uninsured investors in bank stocks and bonds. This ability depends in turn upon the institutional and legal environment within which the regulator operates.

The United States introduced regulation in 1991 that was intended to assuage the effects of the TBTF policy that was revealed by the Continental Illinois bailout. The Federal Deposit Insurance Corporation Improvement Act (FDICIA) of that year requires regulators to act quickly against distressed firms, and limits their discretion to support distressed banks (Wall, 1993). Stern and Feldman (2004) argue that the FDICIA did not resolve the TBTF problem, noting that regulators still have the ability and the incentives to bail out failed institutions and, indeed, the financial crisis indicates that the FDICIA did not fully resolve the TBTF problem. However, as I remark above, it is unrealistic to imagine that the FDICIA or any other piece of legislation could ever entirely resolve the TBTF problem. Prior to the crisis there is some evidence that the FDICIA was at least partially effective. For example, while Flannery and Sorescu (1996) note that a TBTF policy undermines market discipline, they also identify a more pronounced relationship between bond yield spreads and bank risk in the wake of the FDICIA and other reforms; Angbazo and Saunders (1997) estimate a lower systemic risk for large banks after the passage of FDICIA; and, using post-FDICIA data, Ennis and Malek (2005) do not find that very large banks perform differently to smaller banks in the post-FDICIA period.

The FDICIA gave US regulators the ability to intervene early in fragile banks. A positive side-effect was probably that bankers could not perfectly anticipate regulatory actions: this ‘constructive ambiguity’ probably helped to ameliorate the incentive costs of the TBTF policy (Freixas, 1999).

While the FDICIA did not, and probably could not, entirely erase the problems of a TBTF policy, it did demonstrate that a competent regulator, armed with the power of early
intervention, is able to go some way towards convincing uninsured bank investors that they are not immune to the consequences of the risks that banks take and, hence, can serve to restore some of the right risk-taking incentives to banks. But the FDICIA did not apply to non-banks; arguably, the absence of something like the FDICIA was a contributory factor in the failure of the US government-supported agencies Fannie Mae and Freddie Mac (‘Fannie and Freddie’).

Prior to the crisis, approximately half of total residential mortgage debt in the United States was securitized by, or held in the portfolios of, Fannie and Freddie. In 2007, the two agencies owned and guaranteed $5.2 trillion of mortgages, with an equity capital ratio of less than 2 per cent (Financial Crisis Inquiry Commission, 2011, p. 309). Fannie and Freddie were public companies owned by shareholders, but their size raised concerns about the macroeconomic consequences of their failure (Frame and Wall, 2002). Moreover, Fannie and Freddie had credit lines at the Treasury; while market players could not be sure of a government bailout, it seemed very likely prior to the financial crisis that the agencies were both TBTF. The Financial Crisis Inquiry Commission (2011, p. 316) notes the widely held market perception prior to the financial crisis was that they ‘continued to possess an asset that is hard to quantify: the implicit full faith and credit of the US government’.

Given their systemic importance and their apparent TBTF status, it was particularly important that, like the banks covered by the FDICIA, Fannie and Freddie were closely and expertly regulated. However, it appears that they were not. Their regulator, the Office of Federal Housing Enterprise Oversight (OFHEO) had none of the legal powers available to bank supervisors to enforce judgements, set capital requirements, deal with funding, and manage receivership (Financial Crisis Inquiry Commission, 2011, p. 40). According to the Financial Crisis Inquiry Commission (2011, p. 323), the agencies were able to deploy political power to restrict regulation and oversight, and the OFHEO ‘lacked the capacity’ adequately to supervise them. The agencies were able to use poor valuation methodologies without effective challenge or oversight and, given their TBTF status, they had the right incentives to do so. John Kerr, the Federal Housing Finance Association examiner, described Fannie as ‘the worst-run financial institution’ he had encountered as a bank regulator (Financial Crisis Inquiry Commission, 2011, p. 322).

Fannie and Freddie’s failure is well documented. Sharp drops in the US property market after 2007, large holdings of low-quality borrowings, and a liquidity shortage in the summer of 2008 left both agencies in severe financial distress; it appeared that both were insolvent, or very close to insolvent. The government bailout, when it came, was large, and market participants feared that it might be still larger.15

The ability to intervene early in fragile banks, or to exercise judgement as to the best time to do so, was denied to regulators of UK banks before the crisis. Prior to the passage of the Banking Act 2009, British insolvency law did not distinguish banks from other failing companies. As a result, insolvency proceedings were triggered by default; the systemic consequences of bank default were of such potential significance that it proved impossible to trigger default during 2008. As a result, the UK’s tripartite authorities16 did not allow Northern Rock, RBS, or Lloyds to enter formal insolvency proceedings. And, so long as


16 The tripartite authorities in the United Kingdom are HM Treasury, the Bank of England, and the Financial Services Authority (FSA), who together share responsibility for financial stability.
these institutions were not subject to formal insolvency proceedings, there was very limited scope to alter their financial contracts by fiat. As a result, the adverse consequences of financial distress were somewhat ameliorated for investors in these banks; if they interpreted this experience as confirmation that they would never bear the costs of future failure, then the likelihood of future financial crises was probably increased.

The UK’s Banking Act 2009 attempts to ensure that regulators have the power to intervene in failing banks in the future, and that uninsured creditors do not benefit from future bailouts. It institutes a Special Resolution Regime for failing banks, under the terms of which the regulator’s judgement can be used to force financially distressed banks into a formal resolution process. The process allows for bank assets to be transferred to other institutions, to be temporarily managed by the Bank of England, or for temporary bank nationalization; it should not impair the day-to-day payment operations of banks and, hence, it is hoped that supervisors will in future be more willing to close fragile banks, and to impair their uninsured investors’ claims. If the regime works as intended, it should go some way towards addressing the TBTF problem.

United States legislators have attempted to build on the experience of the 2007–9 crisis, and to plug holes in the framework created by the FDICIA. The Dodd–Frank Wall Street Reform and Consumer Protection Act (‘Dodd–Frank Act’) of July 2010 attempts to reduce the scope of the TBTF policy. The Act provides for the creation of the Financial Stability Oversight Council (FSOC), with responsibility for assessing systemic significance; it grants the Federal Reserve new powers to supervise institutions that are systematically significant, and requires institutions to develop ‘living wills’ that will make liquidation simpler. The US Secretary of the Treasury Tim Geithner, testifying before the Congressional Oversight Panel (COP) in June 2010, said that the reform would end the TBTF policy. But, as I argue above, it seems highly unlikely that any reform could ever achieve this. At the end of 2010, the largest bank holding company in the United States controlled assets amounting to 16 per cent of US GDP, and the largest five institutions accounted for nearly 60 per cent of GDP (Hoenig, 2010). It seems inconceivable that these institutions could be allowed simply to fail. However, the Dodd–Frank Act may reduce the direct cost and administrative upheaval of a bailout, and, critically, lower the recovery rate for uninsured creditors in failed financial firms. If it achieves these things then it will go some way towards fixing the cost-of-capital distortions that lie at the heart of the TBTF problem.

To summarize, a TBTF policy is probably a political and economic inevitability. But careful institutional design can limit the damage that this policy causes. The adverse incentive effects of a TBTF policy can be countered directly when the regulator is better able to intervene early in fragile financial firms. The boundary between banks that are TBTF and those that are not can be shifted outwards by designing clear procedures for liquidating failed financial institutions; such procedures can improve banker incentives when they minimize the pay-outs received by the uninsured creditors and shareholders of failed institutions. Finally, there is a case for separating the prudential regulation of banks from the decision to bail them out, so that bailout choices are made by individuals who are not implicated in decisions that will be revealed by a bank closure to have been misguided (Mailath and Mester, 1994; Repullo, 2000; Morrison and White, 2010; Freixas and Santomero, 2004). Arguably, the European Union moved in this direction after the introduction of the single

17 For a brief discussion of UK insolvency processes in the context of the financial crisis, see House of Lords (2009, pp. 52–3).
currency and the creation of the European Central Bank, when monetary policy was removed from the ambit of national bank regulators. It is to this effect that Sironi (2003) attributes an increase in risk sensitivity of European bank debenture spreads in the second half of the 1990s.

(ii) Bank scope restrictions

I argue above that banks may have been moved by the TBTF policy to inefficient scope expansion. In line with this suggestion, a number of supervisors and commentators have argued that the largest universal banks are now too-complex-to-manage, or, indeed, to understand. If a perceived extension of the TBTF policy beyond the commercial banking sector is to blame for this change, then perhaps legislators should force large universal banks to separate commercial from investment banking businesses.

The most widely cited proposal along these lines was advanced by Paul Volcker, a former Chairman of the Federal Reserve System. Volcker argued that US banks should be prohibited from undertaking certain risky investments, unless acting on behalf of customers. A modified version of his original proposal made it into law as section 619 of the Dodd–Frank Act. The section amends the Bank Holding Company Act of 1956 by adding a new section 13 that prohibits banking entities from engaging in proprietary trading, or from acquiring or retaining any equity, partnership, or other ownership interest in a hedge fund or a private equity fund.

The Volcker Rule has not yet been implemented, and at the time of writing, it is too early to comment upon its effectiveness. However, early indications are that the Rule is having an effect: for example, Bank of America closed a proprietary trading desk in order to comply with the Rule,18 Goldman Sachs began in February 2011 to run down its macro proprietary trading team, and J.P. Morgan has spun its proprietary trading division into asset management.19 If the new Rule is to have the desired effect, it is clearly important that regulators be able to draw a clear line between banking and non-banking firms, and to commit not to bail out those that are classified as non-banks. But establishing and policing a clear boundary between regulated and non-regulated entities is difficult. Brunnermeier et al. (2009, Appendix) discuss this ‘Boundary Problem’. They note that, once a boundary has been drawn, institutions that lie near to it will naturally attempt to place themselves outside the ambit of the regulator. Nevertheless, these institutions may benefit from bailouts of regulated entities, either because, in the event, they are so large and so interconnected that they cannot be allowed to fail, or because an important political constituency has a strong interest in their survival.

The most obvious example of the boundary problem in the years prior to the crisis arose in the shadow banking sector. The shadow banking sector comprises entities that perform the type of maturity transformation between short-term assets and long-term liabilities associated with commercial banks, without the benefit of deposit insurance. Although the shadow banking sector was not subject to bank regulation prior to the crisis, many of the assets in the

---


sector were securitizations of loan assets originated by banks. Immediately prior to the crisis, the shadow banking sector had more assets than the regulated sector (Adrian and Shin, 2009).

The shadow banking system had many of the features of the regulated banking sector. In particular, it was subject to panics and runs (Gorton, 2009). This was particularly apparent during the financial crisis in the money market mutual fund market, which experienced a run after the apparently safe Reserve Primary Fund ‘broke the buck’: in other words, when it failed to repay investors at par, as a standard commercial bank deposit account promises to do (Kacperczyk and Schnabl, 2010).

The Volcker Rule draws a clear line between banking and non-banking entities, and it insists that banking firms cannot engage in certain risky proprietary deals. This may serve to lower the riskiness of banking firms. But it will also shift a great deal of risky financial business into firms that are not closely regulated. It may even attract this type of business from regulated banks that are seeking to find ways of circumventing formal regulation. For example, Acharya et al. (2010) demonstrate that securitizations in the asset-backed market did nothing to lower their issuers’ risk, although they did serve to reduce regulatory capital requirements, and hence to increase the value of the implicit government guarantee.

In short, if the Volcker Rule is to succeed in reducing the deleterious effects of a TBTF policy, it is important that informal as well as formal links between deposit-taking institutions and the risk-taking part of the financial sector be severed. If they are not, the Volcker Rule may serve to create pockets of unregulated firms taking risky and highly complicated positions that, like AIG, have to be rescued when they are close to failure.

(iii) Capital regulation

If, as I have argued, the damage caused by a TBTF policy manifests itself mostly in distorted funding costs for uninsured debt, then it may be possible to draw the sting from the TBTF policy by reducing the level of debt liabilities on banks’ balance sheets: in other words, by increasing the regulatory requirement for equity capital. Provided institutional arrangements allow for a bank’s equity holders to be wiped out in the event of its failure, they will react more strongly to the bank’s risk than its debt-holders and, as a result, if they bear more of the bank’s risk, its cost of capital will better reflect the risks that it takes.

Regulators have already started to move to increase capital requirements. The new Basel III capital regulations require banks to hold more and better capital (Basel Committee on Banking Supervision, 2010). As I remark in section III(i), bankers have argued that this will increase their overall cost of funding. But this argument is bogus. Debt funding is cheaper than equity funding for banks for two reasons: first, disbursements to debt holders are tax-advantaged relative to those to equity holders; and second, debt holders anticipate state support in some states of the world, and so lower the price that they charge for their capital. Both of these reasons reflect taxpayer subsidies of the banking system, so arguing against higher equity capital requirements on the basis of increased costs of capital is tantamount to arguing against a withdrawal of state subsidies. As the rationale for increasing regulatory capital requirements for banking firms is precisely that doing so would shrink the level of distortionary state support extended to the banking sector, this counter argument is plainly ludicrous.

Another argument against increased capital regulation is that, in other contexts, high levels of debt can serve to counter agency problems between shareholders and managers. As Jensen and Meckling (1976) argue, high levels of debt might serve to curb managerial tendencies
towards perquisite consumption, because such behaviour raises the possibility of bankruptcy when firms are levered. But banks are different, because many of the costs of failure are borne by the deposit insurance fund. Because of this, as I have argued above, bank shareholders naturally experience less capital market discipline than shareholders in other types of firm. Indeed, the standard argument in favour of higher equity requirements in banking is that they will resolve agency problems between shareholders and regulators, by forcing shareholders to bear more of the costs of risk-shifting.

A better argument against increased capital regulation is that it might leave the door open to regulatory capital arbitrage, of the type identified by Acharya et al. (2010) and discussed in section IV(ii). This type of regulatory capital arbitrage took advantage of capital allowances for risk reduction. These allowances reflected the fact that debt finance appears cheaper to banks. Since, as I argue in the previous paragraph, this appearance was a result of state subsidy, the case for these allowances seems rather weak, particularly because, as Acharya et al. (2010) note, they were frequently granted for illusionary risk reduction.

In summary, the case for higher regulatory capital requirements is a strong one. At the same time, regulators should carefully consider the scope that these requirements present for regulatory arbitrage. Simple, crude, rules might be a better response to the TBTF problem than more complex ones that could give a false impression of precision.

V. Conclusion

Some banks and other financial institutions are so big and so interconnected that their failure would impose unacceptable costs on other economic actors. These firms are ‘Too-Big-to-Fail’ (TBTF). Irrespective of the comments that their politicians and regulators make before a financial crisis, it seems that all developed countries have a TBTF policy, albeit frequently a tacit one. We received direct evidence of this fact during the 2007–9 financial crisis, when large financial firms all over the world received huge levels of taxpayer support.

Executing a TBTF policy involves very high direct costs. But the most important costs of a TBTF policy arise ex ante, in the form of distorted incentives in the financial sector. I have argued that the direct source of incentive distortion is a reduced risk-sensitivity of capital costs, rather than moral turpitude among bankers. Nevertheless, its effect is excessive risk-taking, economically inefficient expansion in bank scale and scope, and, ultimately, a degradation of the state’s fiscal position.

Governments and regulators cannot credibly abolish the TBTF policy. They should instead concentrate upon policies that can ameliorate its effects. I identify three important foci for policy-makers.

First, whether or not a bank is TBTF, and the direct and indirect costs of operating a TBTF policy, both depend upon the institutional environment within which the bank operates. Institutions and laws that make it easier for regulators to intervene in fragile banks, to impair uninsured investors in banks, and to challenge banks’ lending and risk management policies diminish the scale and the incentive effects of a TBTF policy.

Second, regulators could consider scope restrictions for large financial firms. The Volcker rule, now passed as part of the Dodd–Franks Act, is such a restriction. Such restrictions run up against the boundary problem in financial regulation; effective legislation in this area must therefore allow for the endogenous consequences of its passage.
Third, if the incentive problems deriving from the TBTF policy are reflected in the cost of capital, then policy-makers could reduce their impact by increasing statutory equity capital requirements. Equity capital is currently viewed as more costly than debt capital in banks because the latter is subsidized by the taxpayer. Hence, I argue that banker objections to increased capital requirements reduce to the circular argument that state subsidies should not be withdrawn from banks because to do so would be tantamount to withdrawing state subsidies from banks. A more serious concern is that changed capital requirements will open the door to regulatory arbitrage, and, hence, exacerbate the boundary problem. As with regulations on bank scope, alterations to capital requirements should allow for their effects upon the behaviour of the regulated.

References


House of Lords (2009), Banking Supervision and Regulation, 2nd Report of Session 2008-09, Select Committee on Economic Affairs.


