The Boundary Problem in Financial Regulation
Charles Goodhart
National Institute Economic Review 2008; 206; 48
DOI: 10.1177/0027950108099842

The online version of this article can be found at:
http://ner.sagepub.com/cgi/content/abstract/206/1/48

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
National Institute of Economic and Social Research

Additional services and information for National Institute Economic Review can be found at:

Email Alerts: http://ner.sagepub.com/cgi/alerts

Subscriptions: http://ner.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.co.uk/journalsPermissions.nav

Citations http://ner.sagepub.com/cgi/content/refs/206/1/48
THE BOUNDARY PROBLEM IN FINANCIAL REGULATION

Charles Goodhart*

The current financial crisis has raised queries about the adequacy of the present regulatory regime. Whilst the immediate priority may be to plug the obvious holes in the system, there are some long-term generic problems with almost any system of financial regulation. This paper explores one such concern, i.e. the boundary problem. This arises because effective regulation, one that actually bites, is likely to penalise those within the regulated sector, relative to those just outside, causing substitution flows towards the unregulated. This boundary problem impacts on many proposals, such as ‘narrow banking’ and my own, with Avinash Persaud, for state and time-varying capital adequacy requirements. The question of how and where to set the boundary is considered. Such boundaries will always be criticised as leading to disintermediation, competitive inequality (no level-playing-field), inefficiency and higher spreads and borrowing rates; and such criticisms will be valid up to a point. The paper ends by discussing how best to respond.

Keywords: Regulation; boundary; disintermediation
JEL Classifications: E58; G18; G28

1. Introduction

The recent asset price (housing) bubble was driven by credit expansion and excessive leverage, and the current asset price bust by deleveraging. It is hard to find any evidence that the present regime of financial regulation did much, if anything, to mitigate or lessen the amplitude of the cycle; it may even have exacerbated it. This is leading to a ferment of new ideas on the structure of financial regulation. Many such issues relate to more immediate problems, e.g. relating to deposit insurance and a Special Resolution Regime (SRR) in the UK, and rightly so. But that should not cause one to overlook the fact that there are a number of more fundamental, generic issues relating, at all times and everywhere to financial regulation. This paper, on ‘The Boundary Problem’, is an attempt to elucidate and tackle one such generic problem.

2.6 What is the problem?

If regulation is effective, it will constrain the regulated from achieving their preferred, unrestricted, position, often¹ by lowering their profitability and their return on capital. So the returns achievable within the regulated sector are likely to fall relative to those available on substitutes outside. There will be a switch of business from the regulated to the non-regulated sector. In order to protect their own businesses, those in the regulated sector will seek to open up connected operations in the non-regulated sector, in order to catch the better opportunities there. The example of commercial banks setting up associated conduits, Structured Investment Vehicles (SIVs) and hedge funds in the last credit bubble is a case in point.

But this condition is quite general. One of the more common proposals, at least in the past,² for dealing with the various problems of financial regulation has been to try to limit deposit insurance and the safety net to a set of ‘narrow banks’, which would be constrained to hold only liquid and ‘safe’ assets. The idea is that this would provide safe deposits for the orphans and widows. Moreover, these narrow banks would run a clearing-house and keep the payments’ system in operation, whatever happened elsewhere. For all other financial institutions outside the narrow banking system, it would be a case of ‘caveat emptor’. They should be allowed to fail, without official support or taxpayer recapitalisation.

In fact, in the UK something akin to a narrow banking system was put in place in the 19th century with the Post Office Savings Bank and the Trustee Savings Bank. But the idea that the official safety net should have been restricted to POSB and TSB was never seriously

*Financial Markets Group, London School of Economics, e-mail: c.a.goodhart@lse.ac.uk. I am grateful to Gavin Bingham, Vitor Gaspar. Eva Hupkes, Paul Mizen, Avinash Persaud and Martin Weale for discussion and helpful comments. All remaining errors are, however, my own responsibility.
entertained. Nor could it have been. When a ‘narrow bank’ is constrained to holding liquid, safe assets, it is simultaneously prevented from earning higher returns, and thus from offering as high interest rates, or other valuable services (such as overdrafts), to its depositors. Nor could the authorities in good conscience prevent the broader banks from setting up their own clearing house. Thus the banking system outside the narrow banks would grow much faster under normal circumstances; it would provide most of the credit to the private sector, and participate in the key clearing and settlement processes in the economy.\(^3\)

I was reminded by Vitor Gaspar, to whom I am grateful, that this might be prevented by law, taking legal steps to prohibit broader banks from providing means of payment or establishing clearing and settlement systems of their own. There are, at least, three problems with such a move. First it is intrinsically illiberal. Second, it is often possible to get around such legal constraints, e.g. by having the broad bank pass all payment orders through an associated narrow bank. Third, the reasons for the authorities’ concern with financial intermediaries, for better or worse, go well beyond insuring the maintenance of the basic payment system and the protection of small depositors. Neither Bear Stearns nor Fannie Mae had small depositors, or played an integral role in the basic payment system. Nevertheless, as will be discussed later, with particular respect to time-varying loan-to-value ratios, there may in some instances be an argument for using legal prohibitions to help police the boundary between regulated and unregulated functions.

When a banking crisis does occur, it, usually, first attacks the unprotected sector, as occurred with SIVs and conduits in 2007. But the existence of the differential between the protected and unprotected sector then has the capacity to make the crisis worse. When panic and extreme risk aversion take hold, the depositors in, and creditors to, the unprotected sector seek to withdraw their funds, and place these in the protected sector, thereby redoubling the pressures on the unprotected sectors, who are then forced into fire sales of assets, etc. The combination of a boundary between the protected and the unprotected, with greater constraints on the business of the regulated sector, almost guarantees a cycle of flows into the unregulated part of the system during cyclical expansions with sudden and dislocating reversals during crises.

The genesis for this paper lies not so much in the above arguments, which are reasonably well-known, but in the realisation that exactly the same arguments can be deployed against the proposals that Avinash Persaud and I have made for time-varying capital requirements.\(^4\) And much the same criticism can also be applied to other proposals, such as the reversion to the use of a leverage ratio for capital adequacy requirements advocated by Hyun Shin and Philipp Hildebrand of the Swiss National Bank, that might limit credit expansion and leverage in the boom.

What we observe in this latest financial cycle has been, first, a huge expansion of credit, a massive rise in leveraging during the upswing, followed by the crisis, curtailment of credit expansion and major deleveraging with severe, and continuing, effects on the real economy. For reasons which are by now widely understood (and which I and colleagues at the Financial Markets Group foretold),\(^5\) the present regulatory system (comprising Basel II and the move to mark-to-market accounting practices) not only did nothing to restrain the upswing, but is also exacerbating the downturn. In other words it is highly procyclical. Financial regulation is supposed to improve the working of the system as a whole, but does not seem to be doing so; hardly sensible!

So the obvious answer would seem to be to switch to a system which restrains credit expansion and excess leverage in the upturn, and relaxes such regulatory requirements when managers are themselves more risk averse and cautious in the downturn. What regulation needs to do is to counter the natural proclivities of managers (by the appropriate adjustment of incentives, sanctions and trade-offs), not to try to mimic them (as in the boast that Basel II sets regulatory capital closer to the economic capital desired by bank managers). This is the rationale for having some mechanism, whether time-varying risk-weighted Capital Adequacy Ratios (CARs) or a leverage ratio, or any other such, that restrains the regulated from such credit expansion in an upswing.

Suppose that this can be done successfully.\(^6\) Capital requirements are ratcheted up enough in good times to prevent the regulated expanding as much as they otherwise would. The result, as with narrow banks, would be to lessen the profitability and returns on the regulated, relatively to the unregulated at such times. There is sure to be, or to develop, a boundary problem. During good times funds will flow from the regulated to the unregulated, and the regulated will seek to find ways of transferring business to unregulated associates. During crises the flow will reverse, likely with serious
adverse consequences. Our own proposals are just as subject to this generic boundary problem as any other. If financial regulation is effective, it will have to face the boundary problem.

Since the problem is caused by boundaries between the more and the less regulated, one extreme solution might be to regulate either no financial institution, or all of them alike. Both proposals have some adherents, with some advocating ‘free banking’,7 constrained by market discipline alone, and others a completely controlled financial system, as for example practised in earlier communist countries. Neither extreme would seem feasible; ‘free banking’ would lead to results that would be politically and socially unacceptable, whereas totally controlled financial intermediation is inconsistent with a free market capitalist economy. So if the extremes are eliminated, financial regulators and supervisors will always operate in an interior space, in which there is certain to be a ‘boundary problem’.

The unregulated, however, often depend on services, e.g. payment and administrative, and on back-up lines of credit from the regulated. Indeed, the unregulated are frequently associates, or off-shoots, to the regulated. So cannot one maintain the boundary without excessive difficulty by some combination of prohibition on the regulated maintaining associated unregulated entities, and of limitations on the regulated’s ability to provide credit (even on a contingent basis) (and services?) to the unregulated? We have already noted some of the main arguments against legal prohibition. Moreover, how could one prevent a foreign bank providing such services? Even if one could, and wanted to, draw a strict dividing line between, say, regulated banks and unregulated hedge funds, would it not be possible for hedge funds jointly to establish a separate central institution to provide them with quasi-banking services, including the provision of credit? If the unregulated become, as a result of regulation, more profitable than the regulated, over a long enough run of years for this to become publicly apparent, the unregulated will, one way or another, always be able to attract enough funding for extra expansion, however severely the dividing line between the regulated and the unregulated may be drawn.

Recognition that such a problem is generic may help to mitigate it. Many regulators/supervisors appeared to have been taken largely unawares by banks’ reliance on associated SIVs, etc., in 2007. They should not have been. Any new regulation, such as Basel I, Basel II, or time-varying CARs, will bring with it new boundary problems. Any supervisor must try to learn how the regulated are seeking to avoid the constraints placed upon them (and if the regulated are not doing so, it may be an indication that the regulation is just ineffective!).

Although boundary problems are a generic consequence of effective financial regulation, it does not mean that all such regulation is a waste of time, nor that such problems cannot be mitigated by sensible design. We turn next to some proposals for setting the boundary in a manner that can help to lessen such problems.

3. Where to place the boundary?

Regulation is generally imposed for one, or more, of three main reasons:

(1) to constrain the use of monopoly power and the prevention of serious distortions to competition and the maintenance of market integrity;

(2) to protect the essential needs of ordinary people in cases where information is hard or costly to obtain, and mistakes could devastate welfare; and

(3) where there are sufficient externalities that the social, and final, costs of failure exceed both the private costs of failure and the extra costs of regulation.

(1) above has been the main rationale for the regulation of private utilities, but only enters the financial scene in a few rare cases, e.g. where the network economies of having a single market procedure, e.g. a clearing house, are so great that those who control access to the network could extort huge rents from those trying to join.

(2) above has come to mean that bank deposits have become implicitly, or explicitly, insured and guaranteed, at least up to some upper limit. By the same token there are controls on the behaviour of insurance companies and pension funds. Mutual funds, unit trusts, money market funds, etc., are not guaranteed, but are required to behave in certain required ways. The debate in these cases is not whether they should be regulated, but how this might best be done.

The remaining questions apply mainly to reason (3) where the social costs of failure might, or might not, exceed the private costs (plus the extra costs of regulation). In this respect the causes célèbres of recent years have mostly occurred in the USA; Long Term
Capital Management, Bear Stearns, and Fannie Mae/Freddie Mac. None of these was a monopoly, though the market share of the latter two GSEs (Government Sponsored Entities) was uncomfortably high, and none was a proper repository for the life-savings of the deserving poor. Yet none would have avoided failure but for the good offices of the public authorities, and, in the case of Bear Stearns and the two GSEs perhaps some public sector funding as well.

The latter case of Fannie Mae and Freddie Mac is the most clear-cut. If there are to be institutions guaranteeing a large proportion of housing mortgages, they cannot be allowed to collapse in the middle of a housing downturn. They should always have remained public sector bodies, and have been renationalised now. And if ultimately they are to be sold back to the public, they should be split up first, so that there would be both more competition and a greater ability to allow any one of them to fail. But that just leads on to the issue of monoline insurers in the US? Was the regulation of these sufficient?

The monoline insurers help to underpin the municipal and local government debt market in the US by guaranteeing their debt issues against default; they add value to such debt since their credit rating has been higher than that of the underlying issuers. When the ratings of such monoline insurers themselves weakened in the recent crisis, it caused downward revaluations of existing US municipal debt and problems for new issues of such debt. By much the same token Bear Stearns was too connected to the operation of a number of key wholesale and derivative markets to be allowed to fail. These markets could have ceased to function. Again LTCM was such a large player in a range of major financial markets that its failure would have led to severe dislocation, perhaps shut-down, in them.

What we are approaching is the view that in a system where certain financial markets are crucial, those financial institutions whose failure could lead to the collapse of one or more those markets also need to be regulated, irrespective of the fact that they do not qualify for regulation under rationale (1) and (2). Obviously in this case, size, and market share, matters. It is only when a broker/dealer reaches the size and connectedness of Bear Stearns, and above, that regulation is necessary. The same is probably the case with hedge funds. If they get so big that their failure could threaten the effective continuation of markets, then they need regulation; otherwise not. So what should regulators/supervisors do in this respect? They should start by trying to list the key financial markets and systems in their own country. Having done so, they should review whether and which financial institutions are so important to the functioning of that market, or system, that their failure would seriously disrupt the operations of that market or system. Having done so, they could give the financial institutions involved a choice, either to reduce their exposure to this market (or system), or to be regulated. For example, any hedge fund with a total size beyond some limit, or involvement in any market beyond some scale, would be more closely supervised; otherwise not. Most would avoid any such supervision. But that would be all to the good. The aim would be to keep funds small and diversified enough so that they can be allowed to fail.

In essence, the financial supervisors have got to ask themselves which financial institutions can be allowed to fail and which cannot. Those that they claim cannot be allowed to fail should be specifically regulated. The criteria for regulation should be made public. Any institution which is regulated under rationale (3), too important to fail, should be allowed to appeal against that ruling, and should also have the option to avoid regulation by downsizing.

A danger is that regulators/supervisors might draw the list quite narrowly, out of deference to concerns about moral hazard and market discipline, but then in a crisis seek to rescue institutions not on the original list, and not specifically regulated. A possible antidote would be to require the head of the supervisory body involved automatically to tender his/her resignation whenever an unlisted institution was recommended by the supervisory body for such rescue. What difference would this make? Probably not much. The large US investment houses have already now had to come under the Fed’s umbrella, but it is equally important that the myriad of small broker/dealers do not get lumbered with unnecessary regulation. Similarly, supervisors need to assess when hedge funds, and other financial institutions, e.g. monoline insurers or private equity funds, become so large and prominent in certain key markets that their failure could completely disrupt the functioning of those markets. You could, perhaps, describe this as market-related regulation. Moreover as markets change and develop, so should the boundary change. Also note that the boundary does depend on the estimated effects of failure.
Only those institutions so big and connected that their failure would dislocate the key financial markets would be regulated under rationale (3). There is no case for regulating all broker/dealers or hedge funds, only a handful of those that are so large that their failure would disrupt the financial system.

To conclude this section, we should set the boundary for regulation to include all those institutions where customer protection has been perceived as essential, and those institutions whose size and salience is such that their failure would endanger the continuing operation of key financial markets.

4. How do we police the boundary?

A major problem is that the more effective regulation becomes, the more unpopular it will be, since it will prevent the regulated from doing what they want to do. The boundary problem will worsen such unpopularity. It leads to the following claims; that such regulation is:

(a) ineffective and unfair, resulting in disintermediation;
(b) inefficient and cost enhancing;
(c) complex and capable of being subverted.

Let us take each charge in turn. If there is a boundary problem (and regulation within the boundary is effective), then, almost by definition, there can be no level playing field. The unregulated outside the boundary have a stronger competitive position than those within. Not only could this be described as unfair, but business will be bound to flow from the regulated to the unregulated; in other words disintermediation will occur. Moreover, there will be less information about the unregulated, and their risk management may be even worse. So the resulting financial booms and crises could even be enhanced. What the regulators will have done is to take the business away from the regulated (the good guys in white hats) and given it to the unregulated (the bad guys in black hats).

And all that is true up to a point. The point is that the aim of the exercise is to prevent the key financial institutions from overstretching themselves, and so failing, rather than preventing any financial institution from doing the business, if it thinks it profitable. During property booms and bubbles, a local national regulator ought to be thankful if lending into such a boom does become diverted elsewhere. An example was when Canary Wharf, the large London city office project, was financed by foreign, not British banks; the British bank regulators felt relief.

An example of the difficulty of the boundary problem is the fringe bank crisis in 1973/4. Because of prior constraints (only partially regulatory), much property finance was then done by the ‘fringe banks’ outside the framework of controls, and financed in the wholesale market. When the British property market collapsed in Autumn 1973, so did the fringe banks. The Bank organised a ‘Lifeboat’ to save the better elements of the Fringe; having discovered that it was felt in the event necessary to rescue these, the logical next step was to extend the boundary to cover all banks, as done in the 1979 Banking Act.

So the first problem with effective regulation is that it will induce an unlevel playing field, which is unfair, and will cause disintermediation, which will negate some of the purpose of the exercise. And these criticisms are correct up to a point. Ways of dealing with it include trying to arrange regulation so that its effects only bite some of the time, when additional restraint is really needed, so that the costs and benefits to the regulated are not too far out of line, and trying to limit the potential extent of disintermediation.

The next criticism of effective regulation is that it will often be inefficient and lead to higher costs. The financial intermediaries within the boundary are often the most efficient. If their costs are raised, e.g. by higher CARs, then they will have to respond by raising the spread between interest rates on liabilities and on assets. The interest rates charged to borrowers will rise. When regulation is really needed, in asset bubbles, the outlook is generally optimistic. Everything looks good. As Alan Greenspan noted, no one can easily distinguish between an unsustainable asset bubble and a beneficial change in fundamentals. A regulatory initiative that has the effect of artificially raising interest rates, or tightening other borrowing conditions, such as loan-to-value ratios (LTVs), to borrowers at the height of the boom will be extremely unpopular to borrowers, banks and politicians. Moreover, in so far as the regulation succeeds in averting a future bust, it may then also have appeared to have been unnecessary!

It takes a lot of courage to take away the punch-bowl just as the party gets going. Even if regulators had sufficient instruments (which they do not now have), to
restrain cycles in credit expansion and asset prices, would they have the courage to use them, in the face of uncertainty in need and probable vilification in practice? One partial answer to such a ‘time inconsistency’ problem is to put more reliance on procedural rules, i.e. to state publicly in advance that regulation will be tightened in certain specified conditions (e.g. when housing prices, according to index X, rise at an annual rate faster than Y; when overall bank credit in the country grows faster than Z; when lending by bank I grows faster than annual rate J, etc., etc.). Preferably there should be a ladder of responses, not a single trigger point. The FDIC Improvement Act of 1991 in the USA is an example of a proper regulatory procedure.

The more effective regulation is, the greater the incentive to find ways around it. With time and considerable money at stake, those within the regulatory boundary will find ways around any new regulation. The obvious danger is that the resultant dialectic between the regulator and the regulated will lead to increasing complexity, as the regulated find loop-holes which the regulators then move (slowly) to close. Basel I metamorphoses into Basel II. So the process becomes ever more complex, almost certainly without becoming less porous.

How can one halt the onward march of this dialectic? This is not an easy task. One approach, as already noted, is to limit the periods in which regulation is effectively biting to those few in which it is essential, so that the overall costs, and hence the incentive to avoid, such regulation is lessened. Another, and perhaps more important, solution is to place the boundary at a point where flows across the boundary, substitution between claims on intermediaries within, and without, the boundary are likely to be relatively low. As described earlier, it is such flows that cause the main problems.

The key issue here relates to hedge funds. The aim should be to leave such funds outside the regulatory net, unless they become so large (or so connected with a key market) that their failure would be systemically catastrophic. But if ordinary people should begin to switch en masse between hedge funds and bank deposits, that would no longer be feasible. The authorities should require that all hedge funds operating in their own country impose high minimum limits on inward investment, or have available lock-up conditions on invested funds (so that outflows during crises can be constrained). It is bad enough that pension funds are already investing in hedge funds. The need is to insure that hedge funds will continue to be allowed to fail without public support.19

The other main issue is the incentive for intermediaries caught within the regulatory boundary to establish associated entities outside, to which business can be transferred. This is an obvious response for the regulated. So it was surprising, at least in retrospect, that regulators/supervisors appeared to have been often less than fully apprised, in 2006/7, of the development, and implications, of the chain of associate entities that banks had set up for this purpose. Some of these entities were legally separate, but remained reputationally connected. In that case how far will, or can, the bank within the regulatory system allow its, legally separate, associate outside to fail? If the answer should be that it may not feel able to do so, then the risks have not really been transferred off the balance sheet.

5. Conclusion

These problems of setting, and policing, the regulatory boundary are real and severe. There are no easy answers. But perhaps the first step towards resolving such problems effectively is to be aware of them. A guiding principle would be to design the interface between the regulated and the unregulated in such a way that the resulting incentive to shift business into unregulated channels, because of regulation, was so low that it never became systemic. Perhaps one conclusion from this is that regulation should be designed only to bite occasionally. If so, the time when it should bite is, surely, during periods of optimism, risk-seeking and rapid credit expansion, rather than at present when regulation tends to bite hardest just when the regulated are in any case most risk averse.

NOTES

1 Though not so in every case. In some cases regulation may raise profitability for good reasons (increasing demand as a result of enhanced confidence in the quality of the product) or for bad (creating barriers to entry and restricting competition). I am grateful to Gavin Bingham and to Eva Hupkes for this and other helpful comments.
2 Though it keeps on being revisited; Telser (2008) is a recent example.
3 This does not rule out any role for quasi-public utilities in the financial system. Public sector narrow banks, like POSB in the UK or Postfinance in Switzerland, can continue to provide useful services, especially if their services are priced appropriately. Also there can often be a role for a quasi-public sector utility in financial market infrastructures.
4 In op-ed articles in the Financial Times, ‘A proposal how to
Although the primary reason why all banks need to be regulated is depositor protection, a secondary reason why large (and/or crucially connected) banks need regulation is that their failure would also be systemically important. There is no reason why all banks need to be similarly regulated, even if all banks do need to be regulated. The extent, form and intensity of bank regulation should be related to the risk that each bank would pose to the financial system as a whole were it to fail. The Financial Services Authority (FSA) in the UK has made some attempts to apply this philosophy.

This solution is similar to the one proposed in Hüpkes (2004), ‘Protect functions, not institutions’. In addition, that article suggests that some critical functions could be performed by quasi-public utilities (e.g. a Continuous Linked Settlement bank). Note that the US does have a limit on the market share of banks (5 per cent of the national market), plus additional limits for business in individual states. One consequence of a rule that requires an activity to be regulated once a certain size is reached is a clustering of activities just below the ceiling. This suggests that it serves as a boundary. Such rules exist for some trading activities.

While a supervisor might sometimes feel able to rely on being overruled by higher authority, which would still proceed to a bail-out, and thereby keep the job, the (political) stakes of proceeding with such a rescue against the advice of the chief supervisor would be raised.

Although credit ratings agencies have played an influential role, the failure of one of the big ones would be a nuisance, but would not completely disrupt financial markets. There is no case for their regulation under this criterion.

For such reasons Basel II was rather popular with the large international banks. During booms when these banks wanted to expand, Basel II provided no constraint. During the latest financial crisis, a combination of panic, market forces and self-preservation has been causing banks to cut back on lending and to deliver anyhow and they can then blame the regulators for their restrictive policies.

A major problem with the strategy of originate to distribute was that the distribution was often phony, originate and pretend to distribute to associated conduits, SIVs, etc., which were often artificially beyond the boundary, but where the risk and balance sheet burden flowed back to the main bank as soon as the market turned sour. While forcing all banks to retain some residual proportion of securitised products may well be desirable, in order to encourage properly diligent monitoring, the banks that got into worst trouble with CDOs and RMBS were those that retained, or were forced to take back onto their books, too much of such products.

For example the provision of deposit insurance to bank depositors should allow banks to obtain retail funding more cheaply.

A key component of time-varying regulatory controls could be the imposition of time-varying upper limits on loan-to-value ratios for residential mortgages. Such limits can be easily avoided by having a market for second mortgages, or by booking such mortgages abroad. But this could be deterred by making residential mortgage debts only legally recoverable if financed by a first mortgage issued by a bank sited in the country, i.e. including subsidiaries, but not branches, of foreign banks.

Maybe the need is not so much to regulate hedge funds, but to limit the extent to which pension funds and life insurance companies can invest in them.

REFERENCES


