DISCUSSION PAPER SERIES

No. 8602

WHEN BIGGER ISN'T BETTER:
BAIL OUTS AND BANK BEHAVIOUR

Marcus Miller, Lei Zhang and Han Hao Li

FINANCIAL ECONOMICS and
INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP8602.asp**

ISSN 0265-8003

WHEN BIGGER ISN'T BETTER: BAIL OUTS AND BANK BEHAVIOUR

Marcus Miller, University of Warwick and CEPR
Lei Zhang, University of Warwick
Han Hao Li, University of Warwick

Discussion Paper No. 8602
October 2011

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **FINANCIAL ECONOMICS and INTERNATIONAL MACROECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Marcus Miller, Lei Zhang and Han Hao Li

CEPR Discussion Paper No. 8602

October 2011

ABSTRACT

When bigger isn’t better: Bail outs and bank behaviour*

The traditional theory of commercial banking explains maturity transformation and liquidity provision assuming no asymmetric information and no excess profits. It captures the possibility of bank runs and business cycle risk; but it ignores the moral hazard problems connected with risk-taking by large banks counting on state bail outs. In this paper market concentration and risk-shifting is incorporated in an analytically tractable fashion; and the modified framework is used to consider measures to restore competition and stability--including, in particular, those recommended for the UK by the Independent Commission on Banking (2011), chaired by Sir John Vickers.

JEL Classification: E41, E58, G21 and G28
Keywords: bailouts, money and banking, regulation, risk-taking and seigniorage

Marcus Miller
Department of Economics
University of Warwick
Coventry
CV4 7AL

Email: marcus.miller@warwick.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=100440

Lei Zhang
Department of Economics
University of Warwick
Coventry
CV4 7AL

Email: ecrsm@csv.warwick.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=101152

Han Hao Li
University of Warwick
Kirby Corner Road
Coventry
West Midlands CV4

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=174312

* While remaining responsible for the content, we would like to acknowledge the benefit of discussions with Sayantan Ghosal, Peter Hammond, John Kay and Frederic Schlosser and comments from seminar participants at the Universities of Heriot-Watt, York and Warwick, the Royal Economic Society Conference and the Bank of England, especially Henrique Basso, Michael McMahon, David Miles, Udara Peiris and Paul Youdell.

Submitted 20 September 2011

‘Expansion in the variety of intermediaries and financial transactions has major benefits…[But] it has potential downsides … to do with incentives’. Rajan (2005)

Following the financial crisis that began in 2007 the Independent Commission on Banking (ICB) was established in June 2010, charged to make recommendations on structural measures to promote stability and competition in UK banking for the benefit of consumers and businesses. As a contribution to this debate[2], this paper starts from the classic model of banking – which analyses how banks achieve maturity transformation and provide liquidity services – and shows how it may be adapted to address the issues examined by the ICB. It concludes with an overview of policy options, based on this modified model of banking, including the recommendations of the Commission in their Final Report.

Current theory, initially developed by Bryant (1980) and Diamond Dybvig (1983), shows how a bank can allocate its portfolio between a liquid short asset and a higher yielding, illiquid long asset so as to achieve maturity transformation; and how – by exploiting the law of large numbers -- it can offer all consumers insurance against liquidity shocks. But this literature assumes that there is free entry into the banking sector, so ‘competition among the banks forces them to maximise the expected utility of the typical depositor subject to a zero profit constraint’, Allen and Gale (2007, p.72).

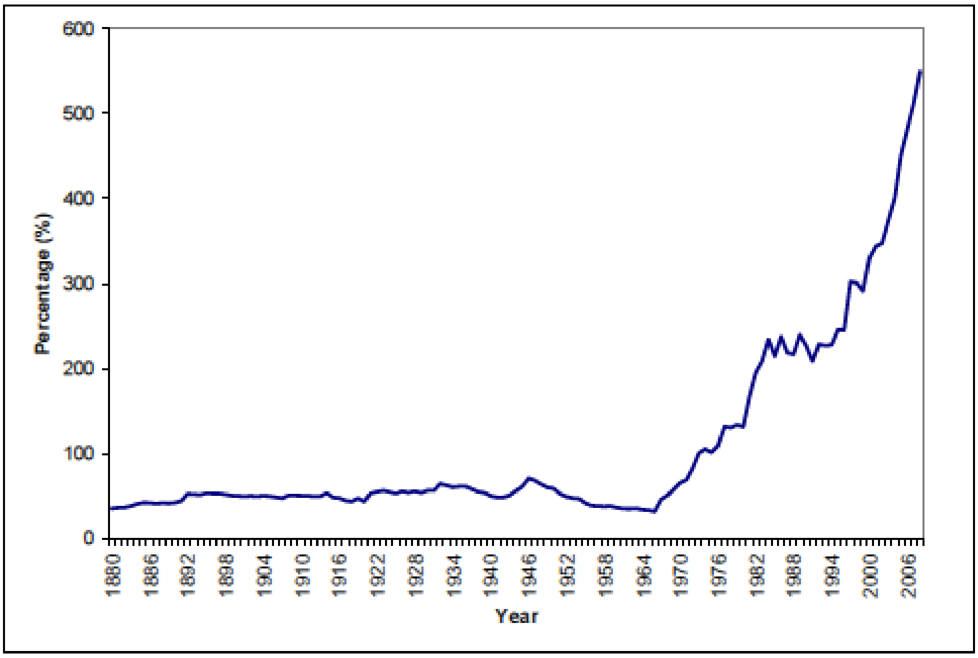
Such an assumption may have been appropriate for banks in the United States for half a century after the passage of legislation in the late 1920s and 1930s. As Haldane (2010) notes, the restrictions on interstate banking imposed in 1927 seemed reasonably effective in controlling the size of the banking industry – at least until deregulation starting in the early 1980s; and the passage of the Glass-Steagall Act separating commercial and investment banking[3] seems to have succeeded in limiting concentration – at least until the passage of Gramm-Leach-Bliley Act of 1999. While the assumptions of perfect competition and utilitarian banking may have been appropriate for US banking when the traditional theory was first developed, the increase in both the volume of bank assets relative to GDP and in industry concentration since the early 1990s suggests this is no longer true. What about the UK?

---

[2] A previous version of this paper, attached as an Annex to Miller et al. (2010), was submitted in response to the ICB’s call for evidence. Submissions are available at http://bankingcommission.independent.gov.uk/bankingcommission/responses/

[3] As discussed in historical detail in Brands (2008), for example.

Two salient characteristics of UK banking are that the key players are universal banks[4] and that the industry is concentrated, 'especially in the retail and commercial sector, where the top six banks account for 88% of retail deposits' ICB (2010, p.9). Two other noticeable features are the rapid expansion of balance sheets prior to the crisis; and the increase in measured value added, especially in profits. As can be seen from Chart 1, banking assets doubled relative to GDP since 1990, rising to more than five times annual GDP – which represents a 10-fold increase above the long run historical average of around 50%.



Source: Sheppard, D. K (1971) and Bank of England.

Note: The definition of UK banking sector assets used in the series is broader after 1966, but using a narrower definition throughout gives the same growth profile.

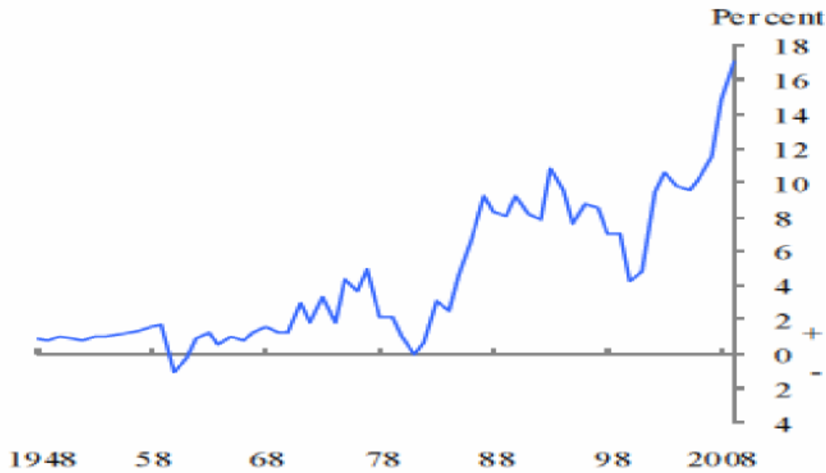
Chart 1: UK banking sector as % of GDP.

Evidence of the sharp rise in the measured contribution of banking to national income in the run-up to the financial crisis is provided by Haldane et al. (2010), where it is reported that, using conventional measures of value added:

> In 2007, financial intermediation accounted for more than 8% of total GVA, compared with 5% in 1970. The gross operating surpluses of financial intermediaries show an even more dramatic trend. Between 1948 and 1978, intermediation accounted on average for around 1.5% of whole economy profits. By 2008, that ratio had risen tenfold to about 15% (See Chart).

[4] i.e. they combine both categories of banking - retail & commercial and wholesale & investment

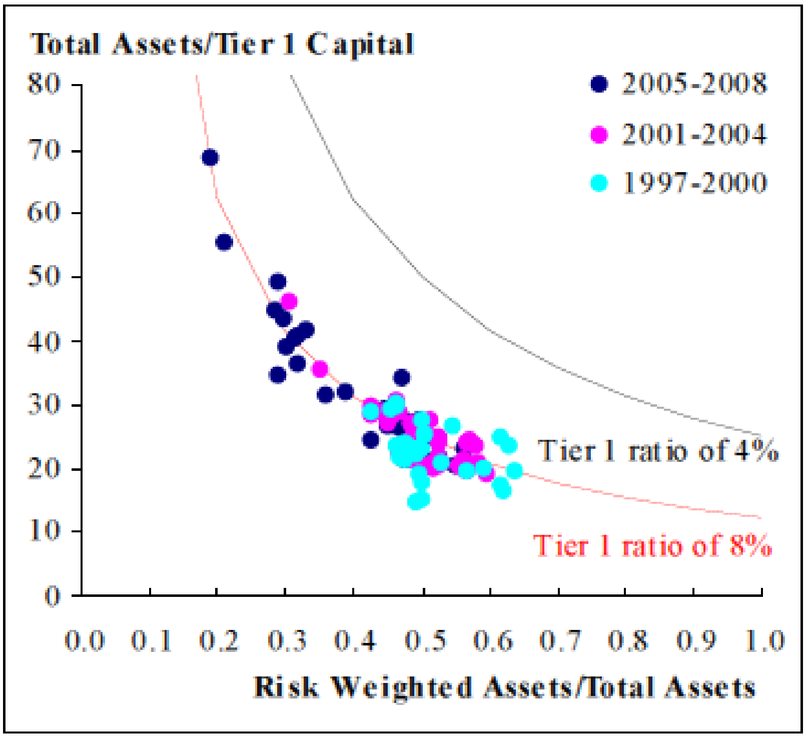
Chart 2 **Gross operating surplus of UK private financial corporations (% of total)**



Sources: ONS and Bank calculations.

Perhaps the most extraordinary feature is, however, that the recent rapid expansion of balance sheets and profitability – what could be termed a productivity miracle -- was accompanied by an apparent reduction in the riskiness of bank portfolios. As shown in Chart 3, the doubling of leverage from the late 1990s until just before the crisis was accompanied by a halving of the fraction of risk-weighted assets. For the period 2005-2008, leverage was around 40 for UK banks on average, considerably higher for some.

Ex-post, this equity cushion turned out to be far too low for the risks actually taken. According to figures from Haldane and Alessandri (2009, Annex, Table 1) official support financial running to 74 % of GDP (including capital injections of more than 5% of GDP) had to be supplied to prevent banking collapse. (They also describe five profit-making strategies -- including taking on tail risk - that contributed to these losses and may have been induced by expectations of state support.)



Source: Published accounts and Bank calculations.

Chart 3: Leverage and risk-taking in UK banks.

To extend a model of consumer-driven bank crisis to accommodate concentration and socially inefficient risk-taking -- a task that involves grafting the gambling model of Hellman *et al*. (2000) onto the base-line model of Allen and Gale (2007) – we modify the traditional theory of banking in three respects. First, by allowing for market concentration -- monopoly in particular – and capitalising these profits into 'seigniorage'; second, by allowing for 'tail-risk' investment by banks not perceived by outsiders on account of asymmetric information. The former transfers the social benefits of banking from depositors to bank shareholders. The latter increases this transfer to include the upside of risky investments whose downside is borne by others. Third we include regulatory activity in the form of required capital, real time monitoring – and 'bail outs' for big banks. This is the framework we use to examine measures designed to address the problems posed by lack of competition and excessive risk-taking – the proposals of the ICB in particular.

The model used in this paper is admittedly stylised and abstract. There is, of course, extensive literature on the market structure of the financial sector from the perspective of industrial organisation, as succinctly summarised in Allen and Gale (2000) and Freixas and Rochet (2008), for example.

Recent research has, however, turned to network theory to analyse the structure of banking system, focusing on how inter-bank connectivity affects the transmission of systemic risk. Using random graphs, Gai and Kapadia (2010a) and May and Arinaminpathy (2010) study the stability of the banking system and find that it is typically "robust-yet-fragile" (robust to the failures of periphery banks but fragile to attacks on the highly connected banks); while Gai and Kapadia (2010b) show how liquidity hoarding can spread through the whole system. For given regulatory regimes, Sui (2010) shows how a core-periphery structure which exhibits "robust-yet-fragile" stability can arise endogenously; and analyses the appropriate policy responses to minimise systemic risk. All these studies indicate that some kind of "structural concentration" is important in explaining how bank failures can become systemic.[5] Our paper does not look at how concentration can arise: we take it as given and investigate how it may impact on banks' incentives to behave prudently.

The paper is organised as follows. Section 1 sets up the competitive banking model and modifies it to allow for market concentration. Section 2 considers the effect of a productivity shock on the profitability of a monopoly bank. In Section 3, we introduce more risky investment opportunities available to a monopoly bank and, in Section 4, we investigate whether the franchise value and/or loss-absorbing capital can ensure prudent behaviour. Section 5 looks at how introducing real-time monitoring can reduce minimum capital requirements. Section 6 extends the model to varying degrees of concentration to show how the acceptance that some banks are "Too Big To Fail" can undermine prudential banking. Section 7 discusses the policy implications for banking reform and looks at the ICB proposals in particular. Section 8 concludes.

1. Utilitarian Banking: competition and concentration

To fix ideas, we first use the basic three-date model with 'early and late' consumers to see how market concentration affects bank profitability. This is done by comparing the optimal

[5] Policy measures in light of network-related research are discussed in Haldane and May (2011) who emphasise the for regulatory requirements to target *systemic* stability.

'take it or leave it' deposit contract offered by a monopoly bank with the competitive equivalent.

Following Bryant (1980), Diamond and Dybvig (1983), and Allen and Gale (2007), each round has three dates, $t = 0,1,2$. There are two assets available to the bank, short and long, all associated with constant return to scale technology. The short asset – representing accessible storage – lasts only one period, and converts one unit of good today into one unit tomorrow. The long asset – representing illiquid but productive investment – takes two periods to mature, and converts one unit invested at t = 0 into $R > 1$ units at t = 2 later. There is a continuum of *ex ante* identical consumers (depositors) with measure 1, each endowed with one unit of good at $t = 0$. At $t = 1$, the types of depositors are known, a fraction $0 < \lambda < 1$ of them being early consumers who derive utility from consumption only at $t = 1$; and $1 - \lambda$ fraction being late consumers who derive utility from consumption at $t = 2$.

The ex ante utility of depositors is

$$U(c_1, c_2) = \lambda U(c_1) + (1 - \lambda)U(c_2) \qquad (1)$$

where c_1 and c_2 are consumptions for early and late consumers, while $U(.)$ is strictly increasing and strictly concave.

Before seeing what a bank can offer, consider what capital markets may achieve in the absence of contingent contracts. Assume specifically that potential depositors can, after the realisation of types, exchange their endowments with each other for early and late consumption goods in capital markets[6] to ensure that $(c_1, c_2) = (1, R)$. This implies an 'outside option' which generates utility of

$$\underline{U} = \lambda U(1) + (1 - \lambda)U(R). \qquad (2)$$

For depositors to participate in banking, the utility from the deposit contract offered should be at least at the level of this outside option,

$$\lambda U(c_1) + (1 - \lambda)U(c_2) \geq \underline{U}. \qquad (3)$$

The other incentive constraint is that the banking contract should be able to separate early and late consumers (so late consumers have no incentive to withdraw earlier), so

[6] See Allen and Gale (2007, pp60-64) for discussion of such a market equilibrium.

$$c_2 \geq c_1 \, . \tag{4}$$

Returns from short and long assets are used to finance early and late consumptions as follows

$$x \geq \lambda c_1 \tag{5}$$

and

$$(1 - \mathrm{x})\mathrm{R} \geq (1 - \lambda)c_2 \, . \tag{6}$$

The sequence of events is such that at $t = 0$, a bank offers a contract (c_1, c_2) in exchange for the depositor's endowment. At $t = 1$, the types of the depositors are realised: and, if they are the early consumers, they receive c_1. At $t = 2$, the late consumers receive consumption c_2.

1.1 Competitive Banking

Before proceeding to monopoly case, we first summarise results under perfect competition.

Proposition 1: The optimal competitive banking contract (c_1^*, c_2^*) satisfies the first order condition for social efficiency, $\frac{U'(c_1^*)}{U'(c_2^*)} = R$, and the zero profit condition, $(1 - \lambda c_1)R - (1 - \lambda)c_2 = 0$. For a constant relative risk averse utility function and risk aversion at least 1, this contract has the feature that: $1 \leq c_1 \leq c_2 \leq R$.

The argument for this proposition, and the properties of the equilibrium, are outlined in detail in Allen and Gale (2007, Chap 3.3). As Diamond and Dybvig (1983) have shown, however, such a model is prone to a 'bank run' if early liquidation of the long asset incurs losses and if there is a sequential service constraint on withdrawal of deposits in t=1. Although this is not a feature we discuss in this paper, it was such a bank run that precipitated the demise of Northern Rock, as discussed in Dewatripoint et al. (2010, p.87-88).

The competitive banking solution is illustrated in Figure 1, where the horizontal axis represents consumption in date 1 and the vertical the consumption at date 2, and the indifference curves represent expected utility of the average depositor. The participation constraint on banking outcomes is indicated by the downward sloping convex curve passing through the point $(1, R)$ labelled Market Equilibrium: so feasible deposit contracts are restricted to consumption points in the convex set defined by (3). The downward sloping straight line l_0 passing through the Market Equilibrium indicates the resource constraint applying to banking equilibria. Bank profitability is zero on l_0 (but positive on positive on l_1, i.e. when the line is shifted to the left).

The competitive contract is illustrated at point A in the figure, where the indifference curve (iso-EU) is tangent to the zero profit line (l_0). For risk aversion greater than 1, it can be seen that $1 \leq c_1 \leq c_2 \leq R$.

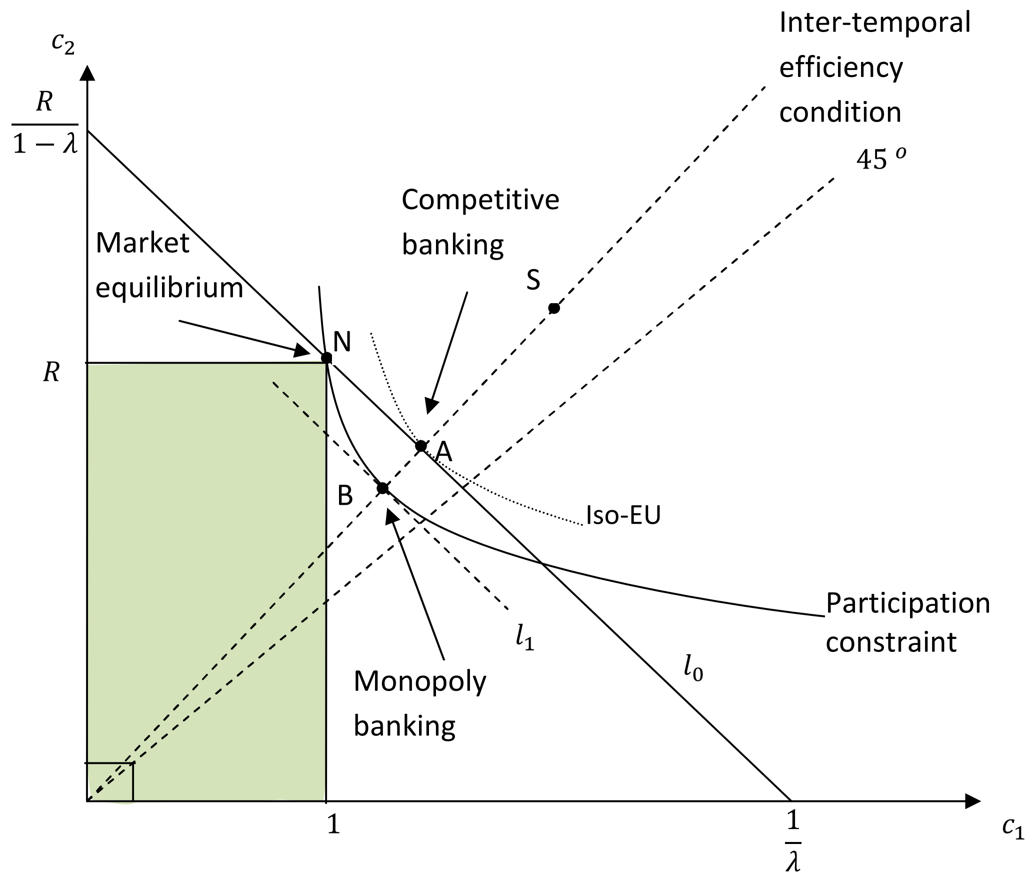


Figure 1. Competitive and monopoly banking

As Bolton and Dewatripont (2005, pp. 112) point out: 'Modigliani and Miller (MM) theorem (1958) states that capital structure (in other words, the form and source of financing) is irrelevant for firms' investment decisions when there are no tax distortions, transactions costs, agency problems, or asymmetries of information).' In the standard model of competitive banking discussed above, the capital structure may indeed be varied without any implications for the asset side of the balance sheet: and Diamond and Dybvig's (1983) model of debt-financed banking was promptly complemented by Jacklin's (1987) version of pure equity banking. The MM Theorem applies because of the assumptions of perfect competition, full information and no risk.

[Where the bank is fully equity financed, with the shareholders paid dividends in each period, one finds:

$$Cost\ of\ Capital = d_2/(1-d_1)$$

where d_i is the per share dividend paid to all shareholders in period *I*, i.e., the cost of capital is the second period dividend per unit invested -- corrected for the interim dividend paid out in period 1. With perfect competition and no risk, this will match the return on capital, R.]

1.2 Monopoly Banking

'It is well known that financial intermediaries can extract rents by exploiting monopoly power through some combination of market share, collusion and barrier to entry' observes Paul Woolley (2010, page 124) in *The Future of Finance*. This can be accommodated in the traditional banking model without much difficulty by allowing for positive profits, as in Chan and Velasco (2001).. Here we explicitly consider the case of monopoly: in addition to being analytically tractable, this has the implication that any failure will be 'systemic'.

A risk-neutral monopoly bank is assumed to maximise its undiscounted, one round, profits by choosing a suitable deposit contract (c_1, c_2) and investment in short asset, x, i.e

$$\Pi = \max_{x,c_1,c_2}\{x + (1-x)R - \lambda c_1 - (1-\lambda)c_2\}, \qquad (7)$$

The first two terms from the profit function are returns from the short and long assets respectively, and the last two terms represent early and late consumption bundles.

The optimal deposit contract is determined when the monopoly bank maximises its profits in (7) subject to constraints (3)—(6). Since the short asset earns lower returns, the bank will have incentive to minimise its holding of x. This implies that (5) must always be binding, i.e.

$$x = \lambda c_1 \qquad (5')$$

Replacing x using (5'), the above problem can be rewritten as

$$\Pi = \max_{c_1,c_2}\{(1-\lambda c_1)R - (1-\lambda)c_2\} \qquad (7')$$

subject to

$$(1-\lambda c_1)R \geq (1-\lambda)c_2 \qquad (6')$$

plus (3) and (4).

The outcome with monopoly can be characterised as follows:

Proposition 2:

The optimal monopoly banking contract (c_1^*, c_2^*) satisfies the first order condition for inter-temporal efficiency, $\frac{U'(c_1^*)}{U'(c_2^*)} = R$, and the participation constraint, $\lambda U(c_1^*) + (1-\lambda)U(c_2^*) = \underline{U}$. This contract exists if and only if

$$\underline{U} \leq U\left(\frac{R}{1-\lambda+\lambda R}\right). \qquad (8)$$

and it must satisfy $c_2^* > c_1^*$.

Proof: The existence condition is trivial because otherwise the feasible set is empty. When (8) is given by a strict inequality, constraint (6') is not binding while (3) binds. In this case, the first order condition is given by $\frac{U'(c_1^*)}{U'(c_2^*)} = R$, which implies $c_2^* > c_1^*$ since $R > 1$ and the utility function is strictly concave. QED

Thus the monopoly bank uses its market power to deny depositors any of the welfare gains available to risk pooling. This monopoly solution is shown at point B in Figure 1. Profit maximisation subject to the participation constraint is achieved when the profit function l_1 is tangent to the indifference curve of the depositor's ex ante utility function, $\lambda U(c_1^*) + (1-\lambda)U(c_2^*) = \underline{U}$.

As regards the distribution of monopoly profits, we assume that these accrue to a limited number of shareholders. Thus, while all members of the population have the same unit endowment of goods, a small fraction of the population , $\sigma << 1$, are also entitled to share in the profits of the monopoly bank.

The final outcome, as shown in the figure, is one of inter-temporal efficiency but income inequality. The majority of the population will expect to achieve the utility associated with no banking, being constrained to consume at point B on the participation constraint. Shareholders, however, will expect to consume an additional amount which takes them to point S, which is the sum of the contract offered by the monopoly bank and their entitlement as shareholders. See Appendix B for further discussion.

2. Productivity shock: Miracle or Mirage?

The extraordinary increase in the size of UK financial services has been noted above. Could this reflect the capture by intermediaries of greater returns available on lending after a positive 'productivity' shock? Not in our benchmark model, even for a monopoly bank, if the productivity gains can be realised *outside the banking sector*. In that case, the increase in R, the rate of return on investments, will lead to a greater allocation of resources to investment – but the benefits will be passed on to depositors because he productivity miracle also leads to a rise in the value of the outside option.

These effects can be seen in the Figure 2 where a productivity shock which raises R to R' swivels the budget line clockwise (as the vertical intercept increases from $R/(1-\lambda)$ to $R'(1-\lambda)$) and swivels the locus of inter-temporal efficiency anti-clockwise. Under the assumption that the outside option is no longer the old market equilibrium at (1, R), but is now (1, R'), i.e. it moves from N to N', the competitive banking contract would shift from A to A', while the profit maximising contract moves from B to B'.

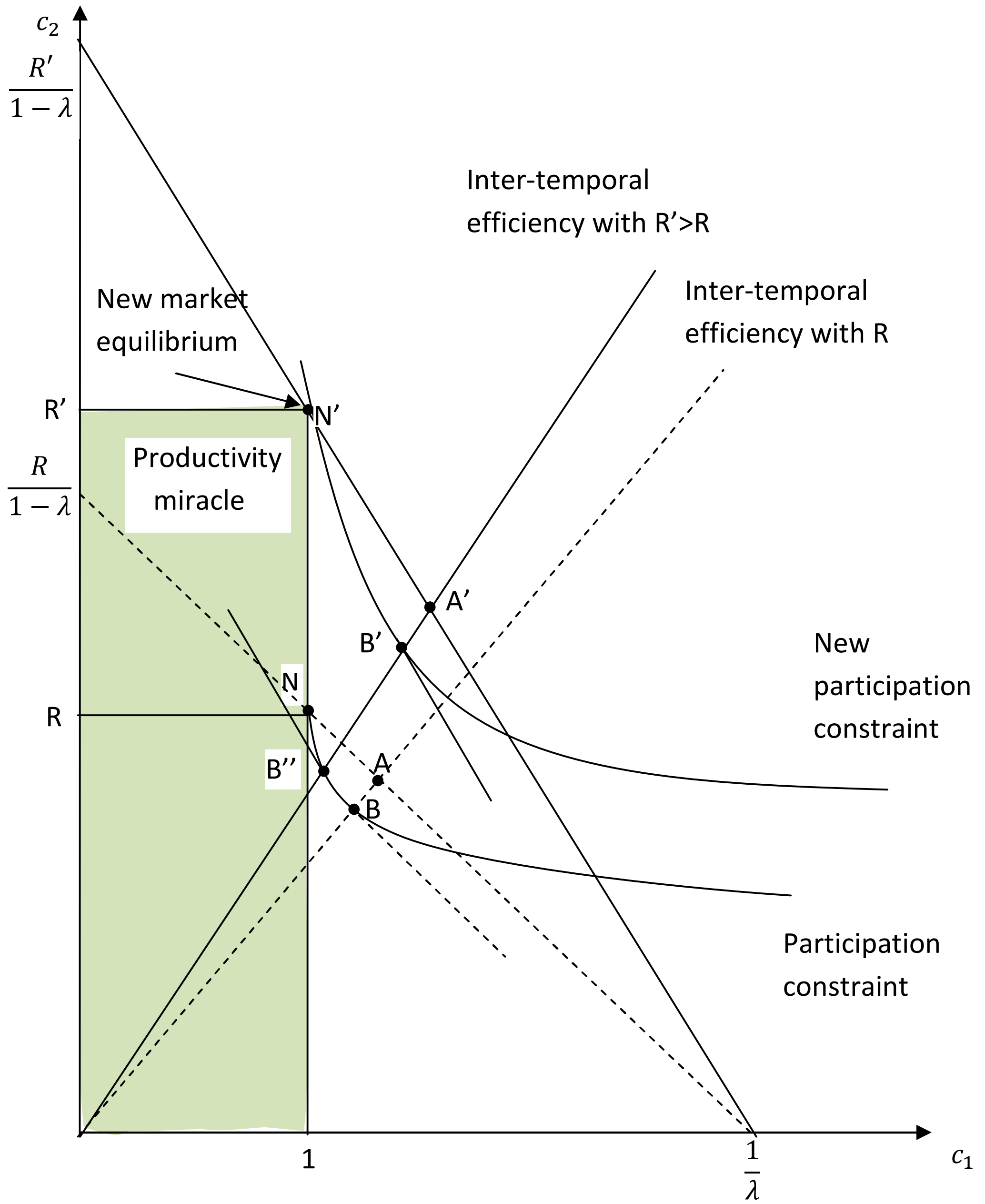


Figure 2. A 'productivity miracle': outside options and profitability

In discussing the extraordinary expansion of the financial sector in the US just before the financial crisis, Reinhart and Rogoff (2009, p.210) comment:

> The size of the US financial sector more than doubled, from an average of 4% of GDP in the mid-1970s to almost 8% of GDP by 2007… Leaders in the financial sector argued that in fact their high returns were the result of innovation and genuine value-added products, and they tended to grossly understate the latent risks their firms were undertaking.

But if, as suggested by those in the financial sector, the productivity gains had their origin *within the banking industry*, then the outside option should surely remain where it was, i.e.

passing through N in Figure 2. This would permit banks to increase their profits substantially (as indicated by the distance from B' to B'') at least for a while – rather like the monopoly profits permitted under patent law to encourage technological innovation. This could represent a 'productivity miracle' due to financial engineering[7].

The gross understatement of latent risk raises another possibility, however, namely that the productivity shock itself has been greatly overstated – with observed returns rising, at least in the short run, far more than expected returns for example. This could lift the quasi-rents available in banking in the same way as raising R to R' for real as banks expropriate rents from depositors due to the latter under-estimating tail-risk. In fact, as Haldane et al. (2010) observe, the inclusion of such earnings in measures of value added may be an error of measurement -- the reason being that the rates of return used are not corrected for risk; and they propose a correction.[8] The implication of the correction is that a substantial fraction of these earnings should be treated not as payments for value added, but as pure 'transfers' from the rest of the economy to the financial sector. A 'productivity mirage' would come from treating such transfers as value added.[9]

How plausible is it that such quasi-rents could be made in finance? In his early warning that financial developments might be making the system riskier, Rajan discussed the distorted incentives governing the allocation of bank resources in conditions of asymmetric information. One of these is:

> 'the incentive to take risk that is concealed from investors – since risk and return are related, the manager then looks as if he outperforms peers given the risk he takes. Typically, the kinds of risks that can be concealed most easily, given the requirement of periodic reporting, are risks that generate severe adverse consequences with small probability but, in return, offer generous compensation the rest of the time. These risks are known as tail risks.' Rajan (2005, p. 316)

[7] On a sceptical note, Reinhart and Rogoff (2009, p.171/2) remark: 'The new delusion was that "this time is different" because there were new markets , new instruments, and new lenders. In particular, financial engineering was thought to have tamed risk by better tailoring exposures to investors' appetites.'

[8] They suggest an adjustment of Financial Intermediary Services Indirectly Measured (FISIM) to allow for risk., and report that, 'According to simulations on the impact of such an approach for the Euro-zone countries, aggregate risk adjusted FISIM would stand at about 60% of current aggregate FISIM for Euro-zone countries over the period 2003-7'.

[9] The overestimation of the value added from the banking system, in the absence of tail-risk, has been documented in Colangelo and Inklaar (2010). Wang et al. (2009) suggest a way to measure properly the contribution of the banking sector in a general equilibrium setting.

Later, in *Fault Lines*, Rajan (2010 p. 139) notes that there are intense pressures to take on tail risk[10] in the banking industry, arguing that neither the traders (who use names such as IBG, 'I'll be gone if it doesn't work' to describe their derivative strategies), nor risk managers (who get fired for worrying about risk), nor the CEOs, nor the Boards and not even shareholders have the incentive to check tail risk. Tail risks are, however, very difficult to control for two reasons: 'first, they are hard to recognize before the fact, *even for those who are taking them*. But second, once enough risk is taken, the incentive for the authorities to intervene to mitigate the fallout is strong', Rajan (2010, p. 152, italics added). As Foster and Young (2011) point out, moreover, there are no non-distortionary compensation packages capable of discriminating between 'true alpha' managers (who consistently generate excess return) and those who are mimicking them by taking on tail risks: efforts to do so will discourage alpha managers themselves.

3. Banking with tail risk

For the UK, Haldane et al. (2009, p. 5-7) argue that two of the five strategies adopted by banks to maximise expected profit involve hidden tail risk: the writing of deep out-of-the-money-options (described as a wolf wrapped in sheep's clothing - beta dressed up as alpha); and high risk lending (on assets yielding a high fixed payoff in good states of the world, but in bad states default generating large losses bunched in the tail in the distribution).

These strategies - together with 'higher leverage', 'higher trading asset' and 'business line diversification' - could help to explain how 'a sector with the utilitarian role of facilitating transactions, channelling savings into real investment and making secondary markets in financial instruments came, by 2007, to account for 40% of aggregate corporate profits in the US , even after investment banks had paid out salaries and bonuses amounting to 60% of net revenues', Wooley (2010, page 121).

For analytical convenience, in what follows we leave to one side the real productivity gains generated by the financial sector, and focus on the profit to be made due to distorted incentives to shift risks into the tail. As in Hellman et al.(2000), we assume that the bank exploits the asymmetry of information to invest in a risky asset with mean return $\tilde{R}$, whose

[10] Tail risks in this paper is used as a metaphor for excessive risk-taking which may have systemic consequences. Using a model of bounded rationality, Gennaioli and Shleifer (2010, 2011) have shown how such systemic risk can arise and endanger financial stability.

true prospects for high and low returns are private information to the bank. These prospects – to be realised in $t = 2$ – are denoted $R_H > R$ and $R_L < R$ respectively, with probabilities π and $(1 - \pi)$; and we only consider the case where $\tilde{R}$ is a mean-preserving-spread[11] of R, i.e., $R = \pi R_H + (1 - \pi)R_L$. Because of the information asymmetry, the downside possibility is not known to the depositors who treat the prospect of high returns as safe- the sweet fruits of innovative financial engineering. As these high, and seemingly safe, returns are not available outside banks, there is no shift to the outside option.

3.1 Monopoly banking

We assume there is concentration in banking and focus especially on a monopoly bank which offers the non-risky contract (c_1, c_2) to consumers. Its expected profits are then

$$\tilde{\Pi} = \max_{c_1,c_2}\{\pi[(1-\lambda c_1)R_H - (1-\lambda)c_2] + (1-\pi)max[(1-\lambda c_1)R_L - (1-\lambda)c_2\,,0]\} \tag{13}$$

where the term $[(1-\lambda c_1)R_H - (1-\lambda)c_2]$ represents the realised profits in the high state, and $max[(1-\lambda c_1)R_L - (1-\lambda)c_2, 0]$ represents the realised profits in the low state. Note that if $(1-\lambda c_1)R_L < (1-\lambda)c_2$, the bank will not be able to fulfil its contract to late consumers, and will be insolvent. What happens in this case is not apparent to the depositors ex ante, however: the low-probability financial crisis will be unanticipated.

To find the optimal deposit contract, one maximises (13) subject to (4) and (5'). Note that here we cannot impose constraint (7'), even in expected terms, because it is possible that the bank is protected by limited liability – and might even be bailed out by the government in the low state, as discussed further below.

The optimal deposit contract is summarised in the following proposition, which covers two cases, only the first being relevant here:

Proposition 3:

[11] Hellman et al. (2000) use the word gamble to describe the taking-on of the tail risk with lower mean return. Note that, we use the term gamble below even when there is no lowering of expected return. Our results would remain the same even if the expected return for taking risky investment is lower than the safe return.

(1) If the bank uses the risky technology, and if $R_L < (1-\lambda)c_2^*/(1-\lambda c_1^*)$, then the optimal contract is a solution to $\frac{U'(c_1^*)}{U'(c_2^*)} = R_H$ and $\lambda U(c_1^*) + (1-\lambda)U(c_2^*) = \underline{U}$.

(2) If $R_L \geq (1-\lambda)c_2^*/(1-\lambda c_1^*)$, the optimal deposit contract is the same as that in Proposition 1.

Proof: See Appendix C.

It is worth noting that, as for the case where there is a productivity miracle, the gambling bank will offer a deposit contract with dated consumptions further apart than for a bank that does not gamble. The optimal deposit contract with a gambling monopoly is shown in Figure 3, using the same axes as in Figure 2. As long as the gamble succeeds, the effective returns for the long asset will apparently have increased to R_H, so the iso-profit functions show a clockwise rotation (see l_0' and l_1') and the efficiency locus also shifts as if there has been a positive productivity shock. But with no change in the outside option, the deposit contract shifts along the original participation constraint. Consequently, the optimal deposit contract offered by the gambling bank is at B'' where the iso-profit function l_1' is tangent to the binding participation constraint (2). Compared with the contract without gambling, date 1 consumption falls and date 2 consumption increases. As long as the gamble succeeds, so bank profits, $(A' - B'')$, will rise sharply, as is suggested by the point S representing consumption of owner-managers of the monopoly bank.

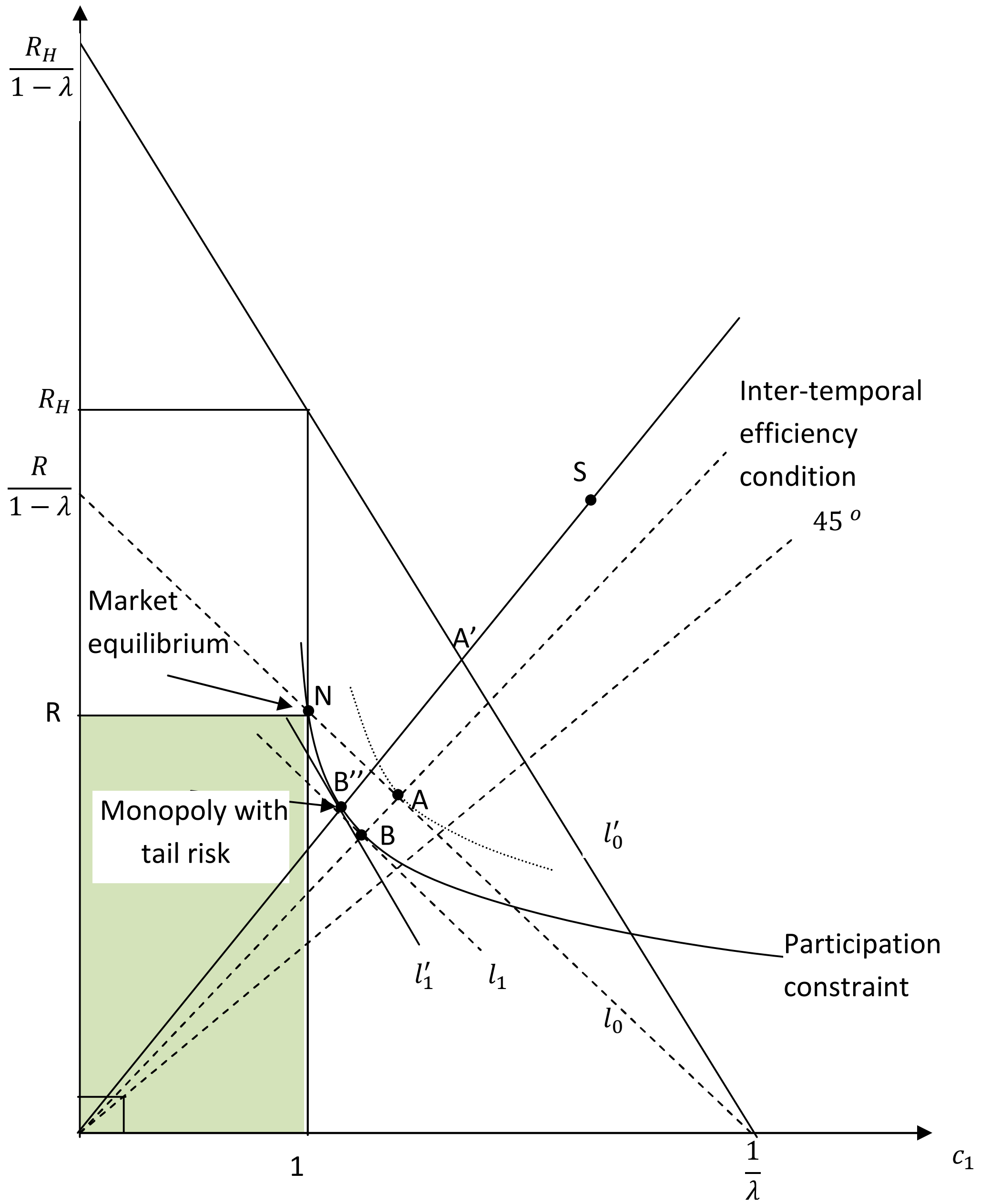


Figure 3. A productivity 'mirage': monopoly banking with tail risk ($R_H > R$)

4. Franchise values and capital buffers as a check on gambling

Given the asymmetry of information, increased competition *per se* would have a dramatic effect on banks' incentive to gamble. With perfect competition, for example, profits to prudent banks will be zero, i.e., there is no franchise value: so all banks will gamble in the equilibrium. Hellmann, Murdock and Stiglitz (2000) consider two regulatory restraints on gambling behaviour: either to impose minimum capital requirements and/or to limit deposit rates so as to allow banks to make excess profits (as with Regulation Q in the U.S.) – subject to the loss of the bank licence if the bank fails in either case.

Before looking at official intervention, consider the possibility of self-regulation. Even without capital requirements or regulation on deposit rates, Bhattacharya (1982) pointed out that the threat of losing its franchise could inhibit gambling by a financial institution; and Allen and Gale (2000, p. 269) note that 'the incentive for banks to take risks in their investments … is reduced the greater the degree of concentration and the higher the level of profits'.[12]

4.1 Franchise value without capital buffers

Will monopoly profits suffice to check gambling without regulation? To compute the franchise value of the monopoly, we consider a repeated game with infinite number of possible rounds. Each round has three dates, and the bank exchanges its deposit contract with consumers at the beginning of each round. There is no discounting within the round but the discount factor between two consecutive rounds is $0 < \delta < 1$. If the bank does not gamble, its capitalised profits are given by the following value function:

$$V_N = \Pi/(1-\delta) \tag{14}$$

In the context of the model we are using, this quantity V_N is the "seigniorage" accruing to the monopoly bank by virtue of its right to create money. Is this seigniorage large enough such that its loss will prevent gambling?

If the bank gambles, the value function is:

$$V_G = \tilde{\Pi} + \delta\pi V_G \tag{15}$$

This means that the gambling bank can capture current-round profits and future discounted profits if the gamble succeeds. But if the gamble fails, losses are taken over by the government and shareholders lose the franchise.

Simplifying (15) yields,

$$V_G = \tilde{\Pi}/(1-\delta\pi). \tag{16}$$

[12] Boyd and De Nicolo (2005) have, however, argued that monopoly behavior which generates franchise values may also have adverse selection effect as loan rates increase and loan quality deteriorates.

To remove the incentive for the bank to gamble, we have to ensure that

$$V_N > V_G. \tag{17}$$

Using (14) and (16), one can rewrite (17) as

$$\widetilde{\Pi} - \Pi < \delta(1-\pi)V_N \tag{18}$$

where the left hand side indicates the one round gain from gambling, and the right hand side represents the cost of gambling, the possible loss of franchise value. Note that (18) is specified for any given feasible deposit contract. Conditioned on this deposit contract, the bank then considers whether to invest in gambling asset or not. This 'no-gambling-condition' is very similar to that in Hellmann, Murdock and Stiglitz (2000).

We may characterise the boundary of the no-gambling-constraint, NGC, (where (18) holds as an equality, the specific form is given in the Appendix) in terms of R_H, π and δ.

Proposition 4:

(1) Given δ, the boundary of the no-gambling-constraint, $R_H(\pi;\delta)$, is downward sloping in π.
(2) An increase in δ will result in a upward shift of the boundary $R_H(\pi;\delta)$.

Proof: See Appendix C.

The boundary of the no-gambling-condition is shown labelled NGC in Figure 4, where the horizontal axis indicates the higher returns for the gambling asset in good state and the vertical the probability of gambling success.

As in Foster and Young (2011), we assume the bank will try to mimic some 'true alpha' investors, subject to a credibility limit[13]. As they do, the targeted 'alpha', $\bar{R}_H$, (shown by the dotted vertical line in the figure) is taken to be a fixed multiple (1.1) of the safe rate R.

[13] It would hardly credible for the bank to claim it can match with certainty on a long term basis the return achieved by Warren Buffet, for example.

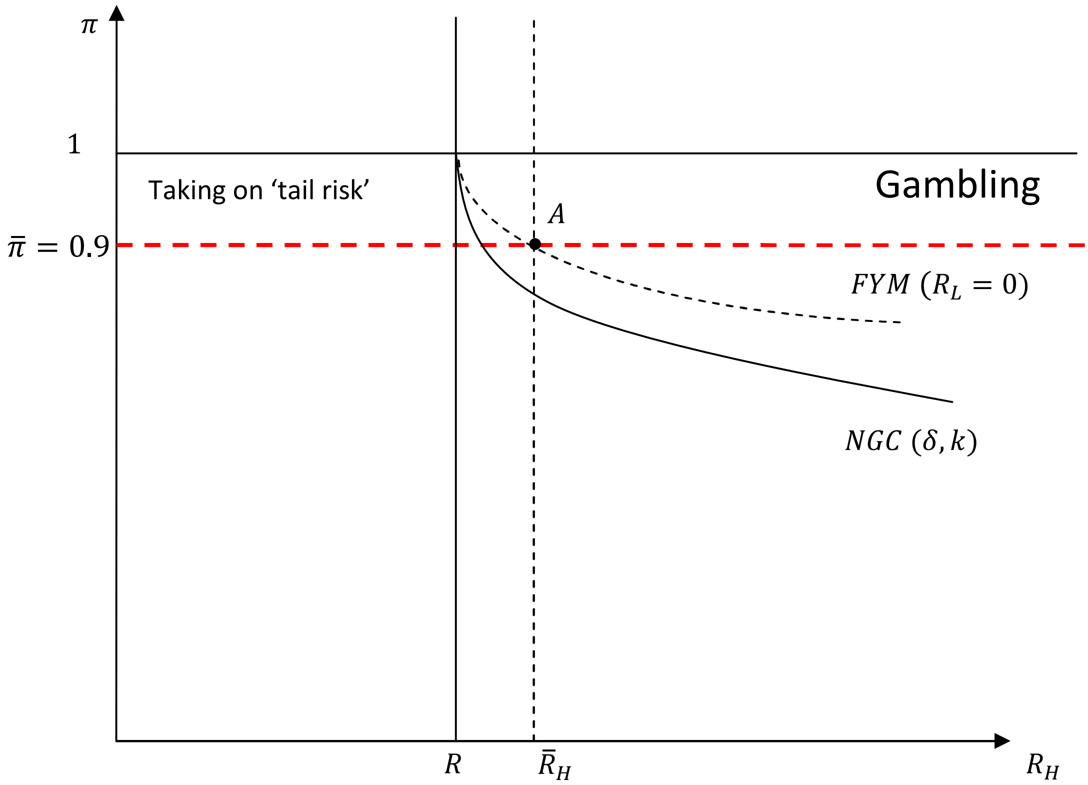


Figure 4. No-gambling-condition (NGC) and the mimicking constraint

The dotted curve, FYM, in Figure 4 indicates the mimicking strategy, which replicates the safe pay-off with risky investments such that $\pi\bar{R}_H + (1-\pi)R_L = R$, where R_L is set to zero and $\bar{R}_H$ is chosen to reflect the targeted alpha investor. For the parameters used in Table 1 below, FYM lies above the NGC, so the bank will gamble, as shown at point A for example, where the alpha target $\bar{R}_H$ is achieved by taking on tail risk[14](see the entry for $\pi = 0.9$ and $R_H = 2.2$). This may be countered by imposing capital requirement, k, as discussed in the next Section.

The essential features of the boundary of no-gambling constraint given in Proposition 4 are illustrated by the numerical example in Table 1, where $R = 2, \lambda = 0.5, \gamma = 2$. For these parameter values, the monopoly bank makes a seignoirage profit of 0.057, measured in date 2 consumption. (Given that $R = 2$, this implies that almost 3 percent of the endowment will be transferred from the depositors to the shareholders even without taking on tail risk.) Entries in the table indicate how profits may be boosted by risk-taking for various values of R_H and π.

[14] Using standard definition for tail risk, the lower threshold for tail risk is $\pi = 0.9$ in our binominal model with mean preserving spread, as shown by the dashed horizontal line in the figure.

R_H	2.1	2.2	2.3	2.4
π=0.95	0.096	0.137	0.180	0.222
π=0.9	<u>0.091</u>	0.13	0.17	0.210
π=0.7	**0.070**	<u>0.101</u>	<u>0.132</u>	<u>0.164</u>
π=0.5	**0.050**	**0.0723**	**0.0945**	<u>0.117</u>

Table 1: Expected flow of profit for monopoly bank that gambles ($R = 2, \lambda = 0.5, \gamma = 2$).

Clearly, the opportunity for a monopoly bank to take on tail risk can substantially undermine its incentive to behave prudently. But the higher the discount factor, the less the incentive for the bank to gamble, as the franchise is valued higher.[15] For a discount factor of $\delta = 0.7$, entries in bold[16] (in blue) towards the lower left of the table satisfy the no-gambling condition, while others fail; however, for a more long-sighted bank with a discount factor of $\delta = 0.9$, the underlined entries (in red) also satisfy no-gambling condition.[17]

4.2 Imposing capital requirements

It is clear from the numerical examples above that when δ is small, the franchise value itself may not be sufficient to deter gambling. In this case, extra measures are needed to ensure correct incentives. So, in what follows, we consider imposing regulatory capital requirements.

With the imposition of a positive capital requirement, k, a gambling bank's expected profit becomes:

$$\tilde{\Pi}(k) = \max_{c_1,c_2}\{\pi[(1-\lambda c_1)R_H - (1-\lambda)c_2] + (1-\pi)max[(1-\lambda c_1)R_L - (1-\lambda)c_2\,, -k]\}$$

where k is measured against total deposits. In the good state, capital will not bring additional cost; but in the bad state, the bank will have more to lose.

The imposition of the capital requirement modifies the no-gambling condition

$$\tilde{\Pi}(k) - \Pi \leq \delta(1-\pi)V_N.$$

[15] The franchise value is the expected no-gambling profit divided by (1-δ), where δ is the discount factor.
[16] Note that equation (18) is used to check whether each entry satisfies the no-gambling-condition.
[17] Raising δ to 0.95 would be sufficient to insure that point A satisfies the NGC.

Note that k has no effect on the optimal deposit contract offered by the gambling bank, and has no effect on Π and V_N, so the no-gambling-condition above can be rewritten as

$$\widetilde{\Pi}(k = 0) - \Pi \leq (1 - \pi)(\delta V_N + k). \qquad (18')$$

It is clear that in checking gambling k is a perfect substitute for the franchise value δV_N. Since $\widetilde{\Pi}$ is increasing in π and R_H, the imposition of the capital requirement shifts the no-gambling boundary, NGC, in Figure 4, upward, reducing the incentive to gamble for any given π and R_H.

It is worth bearing in mind, however, that the efficacy of regulatory capital will also be limited by outside options. Securitisation may be one of these: if regulatory burden on banks becomes excessive, securitisation may be a form of 'regulatory arbitrage', helping to move the business of banking off-balance sheet.

5. Real-time monitoring

For the case discussed above, the regulatory capital required to deter gambling can be substantial, even for moderate gambles (see rows 1 and 3 in Table 2 below). One way to reduce the capital charge is to introduce real-time monitoring. Real-time monitoring will be characterised by a given probability of detecting gambling before it fails; and an associated punishment. For simplicity, we assume that the probability of detection is q and the punishment is the loss of franchise. (Later, we discuss the effect of other punishments.) In this case, the value function of the gambling bank becomes

$$V_G^m = (1 - q)[\widetilde{\Pi}(k) + \pi\delta V_G^m]$$

or

$$V_G^m = \frac{(1 - q)\widetilde{\Pi}(k)}{1 - (1 - q)\pi\delta}$$

Note that introducing real-time monitoring simply scales down the profits of a gambling bank, so the functional form of deposit contracts offered by the gambling bank are unchanged.

Applying the no-gambling-condition $V_G^m \leq V_N$ and re-arranging yields

$$\widetilde{\Pi}(k=0) - \Pi \leq (1-\pi)\{\delta V_N \left[1 + \frac{q}{\delta(1-q)(1-\pi)}\right] + k\} \qquad (18'')$$

Given capital requirements, imposing real-time monitoring reduces gambling profits and so decreases the incentive to gamble. By comparing (18’’) with (18’), it is clear that the net effect is “as if” there is an increase in franchise value. So, in terms of Figure 4, introducing real-time monitoring further shifts the no-gambling-condition upwards in π and R_H space.

Using the no-gambling-condition (18’’), one can obtain the minimum capital requirements as:

$$k^* = \frac{\widetilde{\Pi}(k=0) - [\frac{1}{1-q} - \pi\delta]V_N}{1-\pi}$$

To gauge quantitative significance of real-time monitoring of this form, we compare minimum capital requirements (measured in terms of deposits) for the mimicking strategy, $R_H = R/\pi$, both under perfect competition and under monopoly. Results are summarised in Table 2 where $\gamma = 2$, $\lambda = 1/2$, $q = 0.3$, $\delta = 0.9$ and $\pi = 0.81$. For returns under prudent investment, we choose $R = 2$ as in Allen and Gale (2007), $R = 1.04$ as in Foster and Young (2011) and an intermediate case where $R = 1.5$.

Regime	**R = 2**	**R= 1.5**	**R = 1.04**
Monopoly without monitoring	0.315	0.550	0.528
Monopoly with monitoring	0	0.209	0.522
Perfect competition	0.854	0.693	0.528

Table 2: Minimum capital requirements under different regimes

The first row in Table 2 indicates that, without monitoring, decreasing safe returns, R, increases the minimum prudential capital requirement. This is because smaller R implies a smaller franchise value, so capital requirements have to be increased to offset this and preserve proper incentives. The second row shows that the effect of introducing real-time monitoring depends crucially on the franchise value. When this is high (R=2), a moderate probability of detecting the gamble (30%) can substantially reduce the minimum capital requirement.[18] When the franchise value is low (R=1.04), the effect on the minimum capital requirement is minimal.

The third row illustrates minimum capital requirements under perfect competition: as R decreases, gambling by using mimicking strategy becomes less attractive, so minimum capital requirements decline. Note first that, for any given R, the minimum capital requirements under perfect competition are generally greater than those under monopoly. This is because there is no franchise value under perfect competition; so the gap between them shrinks as the franchise value decreases. Note also that, because of the specific form of the punishment used above, real-time monitoring has no effect on minimum capital requirements under perfect competition.

The numerical examples illustrate how the effectiveness of our form of the real-time monitoring depends on the level of franchise values: becoming ineffective when the franchise value is low (either because of high degree of competition or low returns on prudent investment). One way to overcome this would be to impose an alternative (or additional) sanction in the form of fixed fine, for example, with the effect very much resembling that of a capital requirement. By combining these two forms of sanction, real-time monitoring could be effective regardless of the level of franchise value.

These numerical exercises are, however, subject to a major qualification: they take no account of bank bail outs, so they exaggerate the prudential benefits of concentration.

6. Concentration and 'Too Big To Fail': the U-shaped NGC

If the banking sector is highly concentrated, the failure of one bank is more likely spread to the whole sector, generating systemic risk; so, to prevent a wholesale banking collapse – with all the externalities that will involve - the government may see no alternative to bailing out the failing bank. But seeing itself as "too big to fail" can greatly undermine a bank's

[18] Similar results are found elsewhere. Thus, in a careful calibration of the Hellmann, Murdock and Stiglitz (2000) model, Kuvshinov (2011) shows that real time monitoring can reduce minimum capital adequacy ratio from around 40% to 20%.

incentive to invest prudently , as Haldane and Alessandri (2009) point out in a paper with the suggestive title “ Banking on the State” .

Before discussing in detail how the no-gambling-condition might be modified in the presence of “too big to fail” policy, the strategic elements involved are examined with the aid of a simple two - player game between the banking industry – represented by a monopoly bank - and the state – represented by the taxpayer. (This is very much a short cut , as in practice protecting the interests of society is delegated to a troika of agencies, the Bank of England, Treasury and the FSA.)

In the game tree as drawn, it is the Bank that moves first, choosing to invest in a prudent portfolio or a risky one; while the Taxpayer moves next, offering to bailout of a bank which gambles and loses, or refusing to do so - in which case the bank will fail and its affairs resolved in bankruptcy. The figures in parentheses indicate the notional payoffs to the bank and taxpayer respectively, normalised so that each gets zero if the bank plays safe. (Since the risky strategy may succeed or fail, the payoffs on the right hand branch are to be seen as weighted values of payoffs in the two different cases[19].)

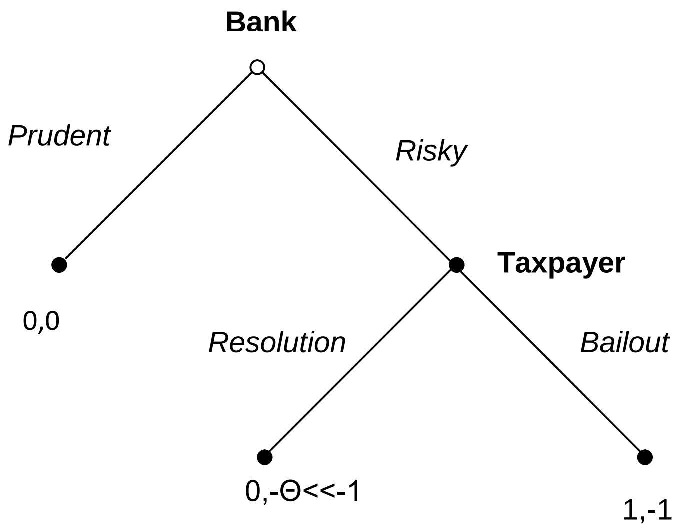


Figure 5. Moral hazard as the taxpayer underwrites risky behaviour

[19] The game could be extended to allow for Nature to act third, randomly choosing success or failure.

The key things to note about the payoffs for the Bank in this game is that it can expect to get more by taking risks so long as it gets bailed out. Otherwise it is indifferent between prudent behaviour and taking risks subject to resolution under laws which limit liability. For the Taxpayer, however, risky behaviour by the bank is unattractive, either because of the costs of bailing out insolvent banks or, more especially, because of much greater social costs (indicated by the symbol Θ) that a banking collapse will inflict on society.

While the Taxpayer would clearly prefer to the Bank to act prudently, this is not the outcome of the game[20]. Using backward induction, the Bank will choose to take risks knowing that the Taxpayer will bail it out: as the Taxpayer (who takes the down-side of risky bets) is effectively providing insurance for the Bank, this is a classic case moral hazard. Note how the high social costs of bank failure, labelled, Θ , appears to act as a threat: once it has embarked on a risk-taking strategy, a High Street Bank can credibly threaten society with the cry 'Your money or your life'![21] . If this is so, a key element of banking reform will be how to limit this threat.

Can these strategic features be incorporated in the framework developed above? For analytical simplicity, let the degree of concentration, β $(0 \leq \beta \leq 1)$ be defined by the fraction of the monopoly franchise value , V_N, that is obtained if a bank plays safe or is bailed out after a failed gamble. To model "too big to fail" policy, TBTF, let the probability the government will come to the bank's rescue, $\tau(\beta)$, increase with β, with $\tau\left(0 \leq \beta \leq \underline{\beta}\right) = 0$ and $\tau(1) = 1$. The rationale for specifying the bail-out policy in this way is as follows: when the degree of concentration is low, no bank is "too big to fail", so the failure of a bank is less likely to have systemic effect; when the degree of concentration increases, any bank failure is more likely to be systemic, so the probability of attracting bail-out increases. The TBTF policy used here specifies that if the bank gambles and fails, it <u>may</u> be bailed out by the government which will honour all deposit contracts. In this case, the bank loses its equity buffer but its franchise is not revoked.

With a given degree of concentration, the bank's profit if it plays safe is a fraction of that under monopoly, so the deposit contract offered by these safe banks will be a scaled-up

[20] Assuming that how the attitude of the taxpayer would behave for a failing bank are irrelevant if the bank is prudent, there will be two Nash equilibria, Prudent, Resolution and Risky, bailout.
[21] Unlike a Highway robber, however, it is acting entirely within the law!

version of that offered by the non-gambling monopoly (though each bank makes less profits per unit of deposit).

For the gambling bank under market concentration, β, its expected profit, assuming $R_L \geq (1-\lambda)c_2^*/(1-\lambda c_1^*)$, is given by

$$\widetilde{\Pi}(\beta,k) = \max_{c_1,c_2}\{\pi[(1-\lambda c_1)R_H - (1-\lambda)c_2] + (1-\pi)[-k]\}$$

$$= \max_{c_1,c_2}\{\pi[(1-\lambda c_1)R_H - (1-\lambda)c_2] + (1-\pi)(-k) \qquad (19)$$

where the probability of losing capital for the gambling banks is $(1-\pi)$. Note that the optimal deposit contract with market concentration of β, if they exist, would be the same as that of a gambling monopoly.

Given the capital requirements and the TBTF policy specified above, the no-gambling condition is then modified to

$$\widetilde{\Pi}(\beta,k) - \Pi(\beta) \leq (1-\pi)(1-\tau)\beta\delta V_N, \qquad (20)$$

where δV_N represents franchise value under full monopoly and $\beta\delta V_N$ the franchise value with market concentration of β and the failed gambling bank will be bailed out with probability τ. To summarise the results of the no-gambling boundary (above which banks will not gamble) in β and k space for some given π and R_H :

Proposition 5:

(i) For $0 \leq \beta \leq \underline{\beta}$, the no-gambling boundary is downward sloping in beta and k space.

(ii) For $\underline{\beta} \leq \beta \leq 1$, the no-gambling boundary is U-shaped in β and k space.

(iii) Increasing R_H and/or π shift the U-shaped no-gambling boundary upwards.

Proof: See Appendix C.

The significance of Proposition 5 (iii) is that if the bailout is restricted to banks with less attractive gambles (i.e., low π and/or low R_H) the U-shaped no-gambling condition will be much less pronounced, as we discuss further below.

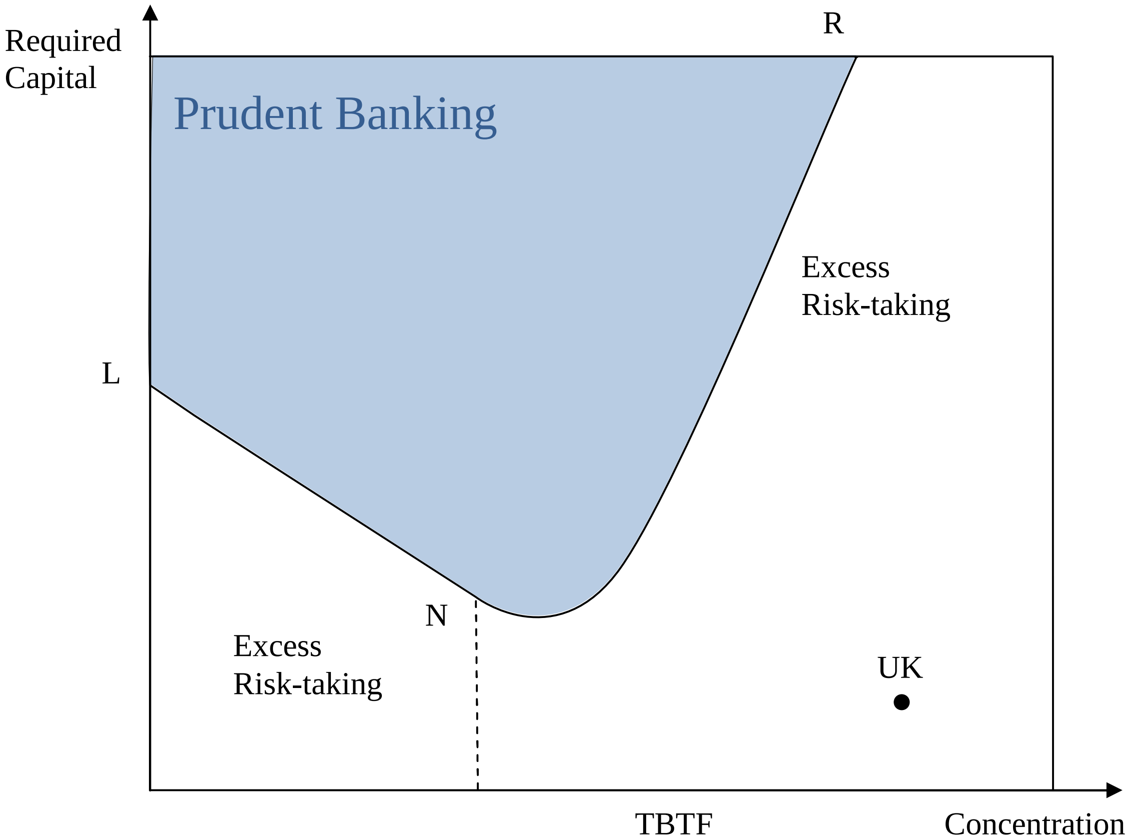


Figure 6. How bailouts increase the risk of imprudent banking

The above framework may be used in a heuristic discussion of options for the reform of the UK banking system, distinguishing in particular between reforms related to structure of bank **balance sheets** (the degree of leverage, for example) and those related to **markets** (such as the degree of concentration). For this purpose, we use Figure 5, with market concentration on the horizontal axis (acting as a proxy for franchise value, assuming that high concentration implies high franchise value), and minimum capital requirement on the vertical axis (to represent variations in bank leverage).

The no gambling condition – defining the shaded area of Prudential Banking -- is the U-shaped schedule LNR in the figure. The downward slope LN reflects the trade-off between bank's profitability (franchise value) and the official capital requirement in terms of prudential behaviour: as banking becomes more competitive and franchise values fall, so the minimum capital requirement will need to be raised to ensure prudence, for any given degree of tail risk.[22] So the point L in Figure 6 will represent the minimum capital requirement needed under perfect competition.

But, as Haldane (2010, p7) points out - on the basis of commercial ratings - an increase in concentration increases the likelihood of an official bailout, as banks become 'Too Big To Fail'. Taking this factor into account will of course greatly encourage gambling, the effect of franchise values being offset by the expectation of a bail-out. In circumstances like these, where banks can, so to speak, have their cake and eat it, the likelihood of banks behaving prudently is sharply reduced, as shown in Figure 6. Let $\underline{\beta}$ indicate the point at which banks become Too Big To Fail. To the left of this point, between L and N, the risk of losing franchise value is sufficient to check gambling: to the right, however, the rise of franchise value increases the probability of bailout which encourages gambling. Consequently, the region for prudential banking becomes U-shaped, as indicated by the shaded area bounded by LNR in the Figure.

7. Regulatory Reform

General points

One reason why banks get bailed out is that their affairs are too complicated to be wound up promptly and efficiently under the normal rules of bankruptcy – Lehman Brothers for example had more than 600 subsidiaries when it filed for bankruptcy. Improvements have already been put in place in provisions for Special Resolution Regime: but to further reduce the moral hazard of bailout, King (2010), Rajan (2010), and Vickers (2011) have proposed further steps , such as the requirement to provide 'living wills', one of the effects of which will be to 'bail-in' debt holders. (If, as is intended, improved resolution procedures will increase the threshold which banks are deemed to be TBTF, this will shift the right hand arm, NR, of the NGC condition.)

[22] And, a shown in Proposition 7, heightened 'tail-risk' (i.e. increasing R_H and/or π) shifts the no gambling frontier LN'R upwards.

What are the implications for the reform in the UK? To start with, it is clear from the evidence that the present level of capital requirements and high levels of concentration in the UK do not ensure prudent banking -- quite the contrary. This is suggested by locating the UK in the bottom right of the figure. What of reform? Take first the need to reduce leverage, as stressed by the Governor of the Bank of England (King, 2010): this can, broadly speaking be achieved by increasing capital requirements – preferably on unweighted assets to limit gaming of the rules. In their evidence to the ICB, Martin Hellwig et al. (2010) commented as follows:

> Basel III is far from sufficient to protect the system from recurring crises. If a much larger fraction, at least 15%, of banks' total, non-risk-weighted, assets were funded by equity, the social benefits would be substantial. And the social costs would be minimal, if any.

In a carefully calibrated study of UK banking, David Miles et al (2010), focusing on risk-weighted assets, report that:

> [O]ur central estimate for the marginal cost and benefit of higher capital suggests an optimal capital ratio of about 50% of risk weighted assets – which might mean a capital to total assets ratio of around 17% and leverage of about 6. This would be about 5 times as much capital – and one fifth the leverage – of banks now… Setting aside risk of GDP fall, our central estimate of optimal capital is 19% of risk-weighted assets.

As for market structure, the evidence suggests that risky M&A activity earns the perverse privilege of increased access to state bail outs (compare the High Street banks with the mortgage banks in the now under intensive care in the UKFI). This could be an argument for breaking up the universal banks as under Glass-Steagall; or, if there are gains from synergy, at least separating the retail and investment arms of the existing universal banks with 'ring-fencing' and 'living wills' so that the latter can go into liquidation if it gambles and fails.

This discussion is, however, subject to an important caveat, namely that such regulatory improvements can be undermined unless decisive steps are taken to reduce the asymmetric information in the financial system. Putting it simply, if asymmetric information is the problem, then transparency must be part of the solution

The Vickers Report in particular

Turning to the key recommendations of the Final Report of the ICB, we note that they involve both market structure and balance sheet restrictions.

Structural separation is recommended in the form of a 'retail ring-fence' designed to isolate and contain banking activities where the continuous provision of service is vital to the economy and a bank's customers so as to ensure that such provision is protected from incidental activities and that it can be maintained in the event of bank failure without government solvency support. 'In essence, ring-fenced banks would take retail deposits, provide payments services and supply credit to households and businesses.' ICB (2011, para. 3.1) Services are divided into those which mandated (involving about 18% of assets as of end 2010); permitted (another 18%); and prohibited (about 64%).

Depending on how much of the second category are taken inside the fence, 'the ring fence might include between a sixth and a third of the total assets of the UK banking sector of over 6tn. pounds.'' ICB (2011, para. 3.40). As banks inside the fence can be linked with those outside (subject to arms length and other restrictions), however, this is not the complete separation required under Glass-Steagall.

In addition several steps are recommended in order to increase competition on the High Street – increased transparency of costs and transferability of accounts, in particular.

Balance sheet requirements involve substantial loss-absorbing capacity in the form of equity and bonds so as to avoid claims on the taxpayer following bank insolvency. Specifically, the Commission recommends that ' large UK ring-fenced banks (and the biggest UK Globally Significant Banks) be required to hold *primary loss-absorbing capacity of at least 17% of RWAs* which can be increased to *a further buffer of up to 3%* of RWAs for a bank to the extent that its supervisor has doubts about its resolvability', ICB (2010, para. 4.118).

This loss absorbing capacity can be split between equity and bail-in bonds, where the equity-to-RWAs ratio is at least 10% for ring-fenced banks with 3% or more of UK GDP in RWAs (falling to 7% for those with RWAs of 1% of UK GDP), ICB (2010, para. 4.132-134).

As regards *monitoring and transparency*, the Commission notes that: "[a] ring-fence of this kind would also have the benefit that ring-fence banks would be more straightforward than some existing banking structures and thus easier to manage, monitor and regulate." (ICB, 2010, para 3.4)

Before discussing these proposals in detail, we use the game tree to see how ring-fencing can change the strategic relation between the state and banks inside the fence .

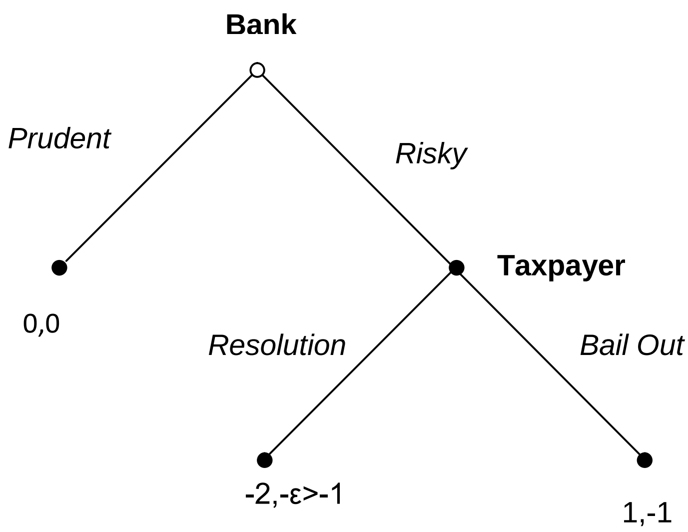


Figure 7. Getting the tax-payer off the hook

In Figure 7, actions and order of play is as before, but the Bank is now the ring-fenced rump of what was a universal bank. For simplicity the payoffs are left unchanged, except for those after resolution. It is still true, therefore, that this Bank will prefer to take risks rather than act prudently if it can count on a bail out. But by regulatory changes on a number of fronts - reducing risks that may be taken, increasing loss-absorbing capacity so as to cover what risk remains, improved monitoring of risk-taking *ex ante* and better resolution procedures *ex post* - the threat to society has been substantially removed (shown symbolically by the replacing the earlier parameter $\Theta >> 1$ by $\varepsilon < 1$ in this figure). So a ring-fenced the Bank can no longer count on being bailed out if it takes risks. Since, in addition, resolution has been made more costly by capital buffers that internalise private losses, the equilibrium of the game is prudent play by the Bank.

This shift in the strategic balance - and the various mechanisms at work to implement it – are designed to reduce the temptation to take excess risk indicated in Figure 5. This can be shown in with reference to Figure 8, which refers only to banks within the ring-fence. Some of the measures should act to expand the region of “Prudential Banking” (beyond that in the earlier figure , indicated here by the dashed U-shape); others to shift the locus of a ring-fenced bank into this enlarged area.

As access to more exotic gambles was found to shift the U-shaped frontier upwards according to Proposition 5, for example, so the *prohibition of many risky assets* – two thirds of the current portfolio of UK banks, in fact – should have the reverse effect, as indicated by the shift from L to L' in the No Gambling frontier. *Improved monitoring* - backed by a threat of losing one's licence if caught – should further reduce the region of excess risk by making the frontier slope down more steeply from L'. Steps to move the locus for ring-fenced banks towards Prudential Banking include both the decisive *increase in the level of capital required* for the operation of a ring-fenced bank and steps to *increase competition* among High Street banks, see the arrow pointing NW in the figure.

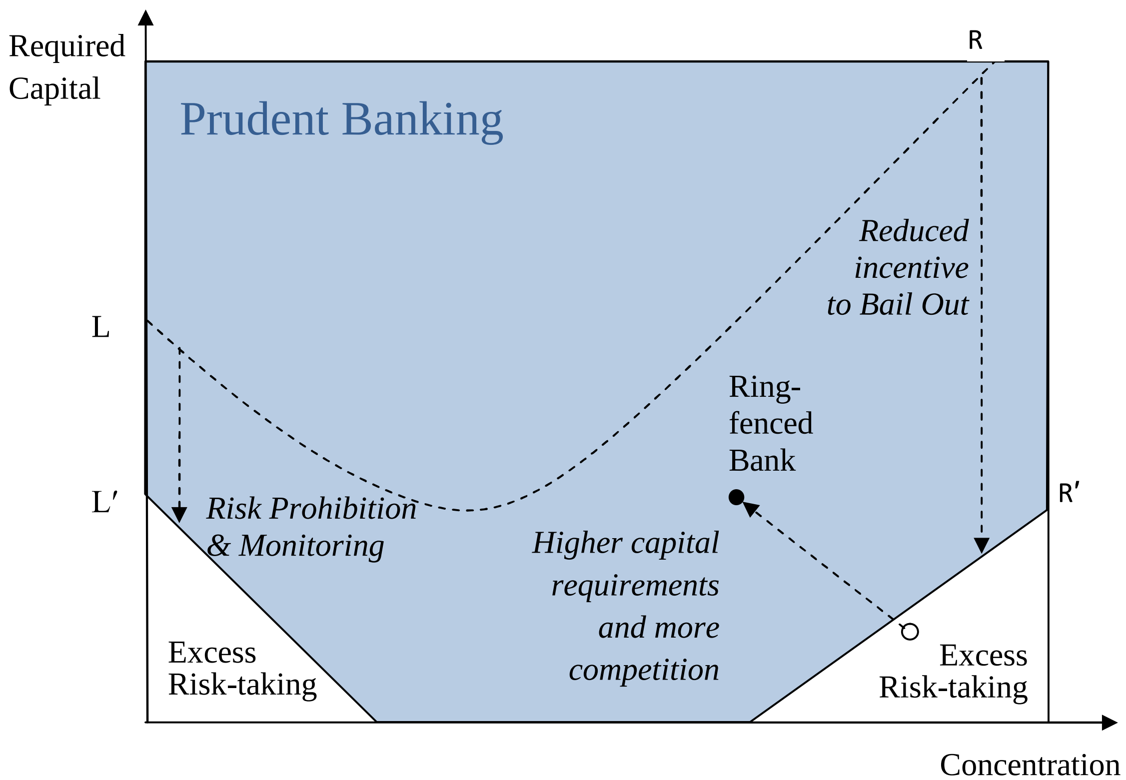


Figure 8. Checking risk-taking in 'ring-fenced' banks

These various regulatory changes - together with arrangements such as 'Living Wills' for *prompt resolution* – are, according to the ICB Final Report, designed to get the taxpayer 'off the hook' of bailing out the High Street banks in trouble. Since the right hand arm of the U-shaped frontier is meant to capture the increased incentive to take risk for those seen as Too Big To Fail, such a shift in strategic balance should, *per contra*, involve a substantial reduction in the area of excess risk-taking to the right of the figure[23].

The stated purpose of retail 'ring-fencing', with the improved transparency and monitoring that it permits by the banning many risky financial products (and the imposition of high-capital requirements on RWAs), is to get such banks back to the business of taking retail deposits, and supplying credit and liquidity to households and businesses. The proposed reforms would, it is claimed:

> put the UK banking system of 2019 on an altogether different basis from that of 2007. In many respects, however, it would be restorative of what went before in the recent past – better capitalised, less leveraged banking more focussed on the needs of savers and borrowers in the domestic economy. ICB(2011, p.18)

8. Conclusion: back to banking basics

There is a dangerous propensity for banks to take on excessive risks in the current regulatory environment: as Vickers (2011, p.2) remarks: "One of the roles of financial institutions and markets is efficiently to manage risks. Their failure to do so – and indeed to amplify rather than absorb shocks from the economy at large – has been spectacular."

Financial innovations, such as securitisation and DCS swaps, have increased the ease with which banks can take risky assets onto their balance sheets while satisfying the regulatory norms set by Basel. There are private incentives for high street banks to expand into investment banking, raising their balance sheets well beyond the needs of households and SMEs borrowers and shifting risk onto depositors by greatly increased leverage. But the social cost of interrupting the nationwide provision of payments services and credit supply associated with bank failures means that banks that combine retail and wholesale activities will be rescued by the government: the threat to the economy effectively puts tax-payers on the hook to underwrite the risks taken by large universal banks.

[23] In terms of the parameters specified in the previous section, this would correspond to a reduction in $\tau(\beta)$ representing the probability of bail out and an increase in $\bar{\beta}$, the level at which this probability becomes non-zero.

Nor has the incidence of crisis changed these incentives. As Diane Coyle (2011), a former member of the UK Competition Commission, noted as bank profitability recovered in 2010-11 despite a still-fragile economy: ‘The truth is that banks are again doing well out of banking, but businesses and consumers are not... Bonuses are back... they are a measure of monopoly rents in the business, it does not take great talent to make a profit by taking excessive risk, safe from effective competition and sure of a bail-out if needed.’

As a contribution to the debate on problems besetting modern banking, we began with a simple model of retail banking and show how, behind the veil of asymmetric information, the incentive to take on risk can easily exceed the threat of losing the franchise - particularly if there is a good chance of an official bail-out. As the prudential benefits of increased concentration are progressively offset by the prospect of rescue, the ‘prudential frontier’ relating capital requirements to concentration becomes U–shaped.

This framework - of concentrated banking with asymmetric information - is used to discuss the impact of regulatory reforms involving changes to market structure, balance sheet restrictions and the efficacy of monitoring. Considering the reforms advocated by the ICB in their Final Report in particular, we note that they are designed to offset excess risk-taking and promote competition, i.e. to eliminate the very features that we have added to the basic banking model to capture current distortions!

A key aim of Mrs Thatcher’s industrial policy was to reduce the threat to the provision of goods and services posed by strikes in the public sector - the confrontation with coal miners being a decisive case in point. An important – perhaps the most important – aspect of the ‘ring-fence’ proposal viewed as industrial policy is how - by reducing the threat of closing High Street banks - it aims to change the strategic balance between banking and the state.

References:

Alan, F. and Gale D (2000), *Comparing Financial Systems*, Cambridge, MA: MIT Press.

Alan, F. and Gale D. (2007), *Understanding Financial Crises*, New York: Oxford University Press.

Bhattacharya, S. (1982) "Aspects of Monetary and Banking Theory and Moral Hazard." *Journal of Finance*, May, 37(2), pp. 371-84.

Bolton, P. and Dewatripont M. (2005, Page 112.), *Contract theory*, MIT press.

Boyd, J. and De Nicolo G. (2005), 'The theory of bank risk taking and competition revisited', *The Journal of Finance*, vol. LX No. 3.

Brands, H. W. (2008), *Traitor to his class.* NY, NY: Doubleday.

Bryant, J. (1980). "A Model of Reserves, Bank Runs, and Deposit Insurance," *Journal of Banking and Finance* 4, 335-344.

Chang, R. & Velasco A. (2001) "A Model of Financial Crises In Emerging Markets," *The Quarterly Journal of Economics*, vol. 116(2), pages 489-517, May.

Colangelo, A, Inklaar, R. (2010). "Banking sector output measurement in the Euro area - a modified approach". ECB working paper 1204.

Coyle, D. (2011) 'Merlin deal will not fix flawed banks', published in FT 8th Feb 2011.

Dell'Ariccia, G., Marquez, R., (2004). "Information and bank credit allocation". *Journal of Financial Economics* 72, 185-214.

Dewatripont, M., Rochet J-C. and Tirole J. (2010), *Balancing the banks: Global lessons from the financial crisis*, Princeton NJ: Princeton university press.

Diamond, D.W., and P.H. Dybvig (1983), 'Bank runs, deposit insurance, and liquidity', *Journal of Political Economy*, 91 (June): 401–19.

Foster D. P. & Young P., (2011) " Gaming Performance Fees by Portfolio Managers' *The Quarterly Journal of Economics*, forthcoming.

Freixas, X. and Rochet J-C., (2008), *Microeconomics of Banking*, MIT Press.

Gai, P. and Kapadia, S., (2010a), "Contagion in Financial Networks", Bank of England Working Paper No. 383.

Gai, P. and Kapadia, S. (2010b), "Liquidity hoarding, network externalities, and interbank market collapse". *Proc. R. Soc.* A 466, 2401–2423.

Gennaioli, N. and A. Shleifer, (2010), “What comes to mind”, *Quarterly Journal of Economics* 125, 1399-1433.

Gennaioli, N. and A. Shleifer, (2011), “Neglected risks, financial innovation, and financial fragility”, *Journal of Financial Economics* 125, forthcoming.

Haldane, A. (2010), ‘The $100 billion question’, Institute of Regulation & Risk, North Asia (IRRNA), Hong Kong, March 2010.

Haldane, A. and Alessandri P. (2009), ‘Banking on the state’, Available at: http://www.bankofengland.co.uk/publications/speeches/2009/speech409.pdf

Haldane, A., Brennan S. and Madouros V., (2010), ‘What is the Contribution of the Financial Sector: Miracle or Mirage?’ *The Future of Finance: the LSE report*, Chapter 2. London: LSE.

Haldane, A. G. and May, R. M, (2011), “Systemic risk in banking ecosystems”, *Nature* 469, 351–355.

Hellmann, T. F. & Murdock K. C. & Stiglitz J. E., (2000). "Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough?" *American Economic Review*, American Economic Association, vol. 90(1), pages 147-165, March.

Independent Commission on Banking (2010), ‘Issues Paper, Call for Evidence,’ London: ICB.

Independent Commission on Banking (2011), *Final Report:Recommendations*, London: ICB.

Jacklin C.J. (1987) ‘Demand Deposits, Trading Restrictions and Risk Sharing.’ In E.C Prescott and N. Wallace (eds), *Contractual Arrangements for Intertemporal Trade*, 26-47, Minneapolis: University of Minnesota Press.

Jonghe, O. d. (2010). “Back to the basics in banking? a micro-analysis of banking system stability”. *Journal of Financial Intermediation* 19, 387–417.

Kuvshinov, D. (2011), “The optimal capital requirement for UK banks: a moral hazard perspective”, MSc Dissertation, Department of Economics, University of Warwick.

LEX Column (2010), ‘Bank non-reform’, Financial Times, October 5 2010.

May, R. M. and Arinaminpathy N., (2010), “Systemic risk: the dynamics of model banking systems”. *J. R. Soc. Interface* 7, 823–838.

Miles, D., Yang J, and Marcheggiano G., (2011), ‘Optimal bank capital’, DP no. 31, External MPC unit, Bank of England.

Miller, M., S. Ghosal, P. Hammond and M. Waterson (2010), ‘Restoring prudent banking in Britain: evidence and policy’, evidence submitted to ICB in November 2010.

Miller, M., Lei Zhang and H.H. Lee (2010) ‘Riding for a Fall: concentrated banking with tail risk.’ Paper submitted to ICB in November 2010.

Rajan, R. G., (2005), ‘Has Financial Development Made the World Riskier?’ Proceedings of the Jackson Hole Conference organised by Kanas City Fed.

Rajan, R. G., (2010), *Fault Lines*, Princeton NJ: Princeton University Press.

Rochet, J-C. and Tirole, J., (1996). “Interbank Lending and Systemic Risk,” *Journal of Money, Credit and Banking*, 28(4), pages 733-62, November.

Sui, Peng (2011), “Financial network formation and the emergence of ‘too connected to fail’”, mimeo, University of Warwick.

Wang, J.C, S. Basu, and J.G. Fernald (2009). “A General Equilibrium Asset-Pricing Approach to the Measurement of Nominal and Real Bank Output”. In W.E. Diewert, J.S. Greenlees, and C.R. Hulten (Eds.) *Price Index Concepts and Measurement*. NBER Studies in Income and Wealth, Volume 70.

Woolley, P., (2010), ‘Why are financial markets so inefficient and exploitative – and a suggested remedy’, *Future of finance: The LSE Report.* London: LSE.

Appendix A. Monopoly Profits with CRRA preferences

To gauge the quantitative significance of monopoly profits, we use a constant relative risk aversion (CRRA) utility function for calibration. Note first that the optimal banking contract under perfect competition results in zero profits for banks. With CRRA utility, $U(c) = \frac{c^{1-\gamma}}{1-\gamma}$, one can obtain the optimal monopoly banking contract as

$$c_1^* = \left[\frac{(1-\gamma)\underline{U}}{\lambda+(1-\lambda)R^{(1-\gamma)/\gamma}}\right]^{1/(1-\gamma)} \tag{9}$$

$$c_2^* = R^{\frac{1}{\gamma}}\left[\frac{(1-\gamma)\underline{U}}{\lambda+(1-\lambda)R^{(1-\gamma)/\gamma}}\right]^{1/(1-\gamma)} \tag{10}$$

Using (7'), one can show that the monopoly profits, measured in terms of deposits, are simply

$$\Pi^* = (1-\lambda c_1^*)R - (1-\lambda)c_2^*$$

$$= R\left\{1 - \left[\frac{(1-\gamma)\underline{U}}{\lambda+(1-\lambda)R^{(1-\gamma)/\gamma}}\right]^{1/(1-\gamma)}\left[\lambda + (1-\lambda)R^{(1-\gamma)/\gamma}\right]\right\} \tag{11}$$

where $\underline{U} = \lambda U(1) + (1-\lambda)U(R) = \frac{\lambda+(1-\lambda)R^{1-\gamma}}{1-\gamma}$.

Given $\gamma \geq 1$, monopoly profits are strictly positive:

$$\Pi^* = R\{1 - \frac{[\lambda+(1-\lambda)R^{1-\gamma}]^{1/(1-\gamma)}}{[\lambda+(1-\lambda)R^{\frac{1-\gamma}{\gamma}}]^{\gamma/(1-\gamma)}}\} \tag{12}$$

Here the size of monopoly profits is limited by the participation constraint – there is an outside option of market equilibrium with no banks. But, as we see in the next section with the aid of numerical examples, monopoly bank profits can be greatly inflated by gambling.

Appendix B. Gambling and Gini Coefficient: Miracle and Mirage?

It is evident that in this simplified model, bank concentration will lead to an increase in the Gini coefficient compared with competitive banking: and this effect will become much more pronounced with gambling. This is illustrated by the stylised Lorenz curves in Figure 4, where σ represents the fraction of the population owning shares in the all-deposit bank. Where ω represents the consumption bundle available to depositors under monopoly banking, and ω(1+μ) is the consumption available to the depositors who are also shareholders enjoying the monopoly premium, μ, in this case $\omega = 1/(1+\sigma\mu)$ and the Gini coefficient[24] turns out to be $(1-\sigma)\sigma\mu/(1+\sigma\mu)$. When the bank gambles, the premium paid to owner-managers will of course rise, say to $\tilde{\mu}$, shifting the Lorenz curve to $O\tilde{L}P$ in the figure. In discussing whether the contribution of financial sector is 'Miracle or Mirage', Andrew Haldane et al. (2010, pp. 79,80) report that the share of financial intermediation in employment in UK is around 4%, and that:

> the measured 'productivity miracle' in finance …has been reflected in the returns to both labour and capital, if not in the quantity of these factors employed. For labour, financial intermediation is at the top of the table, with the weekly earnings roughly double the whole economy median. This differential widened during this century, roughly mirroring the accumulation of leverage within the financial sector.

Using the above formula, a doubling of consumption opportunities for those in finance would add about 4% to the Gini coefficient, i.e. about half the rise in Gini coefficient for the UK from 1986 when the Big Bang took place, to just before the crisis in 2007. (Focusing more narrowly on Investment Banking, however, the Financial Times reports compensation running at 6 times the median income in both US and UK.[25]

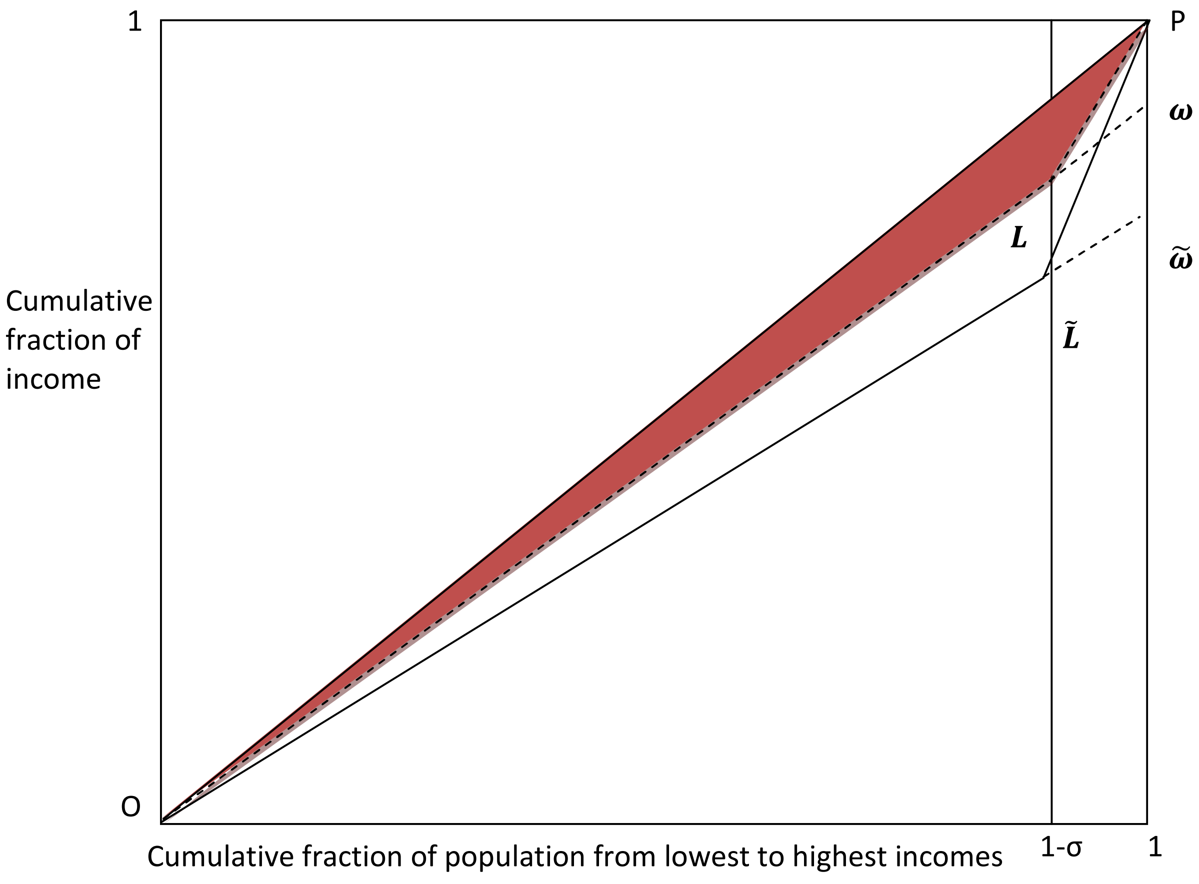


Figure A1: Raising incomes in financial services and inequality

[24] i.e. the area OLP divided by O1P in the diagram.
[25] FT 17th, 2011, 'Feb Banker's pay: time for deep cuts.'

Appendix C. Proof of Propositions

Proof of Proposition 3

If R_L is low enough, the bank cannot honour the contract to the late consumers in the low state (the late consumption in this case is honoured by the insuring agency). So the bank's profits are given by $\pi[(1-\lambda c_1)R_H-(1-\lambda)c_2]$. This changes the first order condition to $\frac{U'(c_1^*)}{U'(c_2^*)}=R_H$. Together with the binding participation constraint, one can then determine the optimal contract as in the second part of Proposition 4. (Since case (2) has the same deposit contract as that under certainty and no default from the bank, we use case (1) to represent gambling.)

If R_L is large, the bank can honour the contract to the late consumers in either state. So the optimal contract satisfies the first order condition $\frac{U'(c_1^*)}{U'(c_2^*)}=\pi R_H+(1-\pi)R_L=R$. With the binding participation constraint the same as in Proposition 1, the optimal contract must be the same.

Proof of Proposition 4

Given the monopoly bank will gamble, it is best for it to choose the deposit contract (C_1^*, C_2^*) specified in Proposition (4). In this case, to ensure that the bank will gamble, condition (18) becomes

$$\tilde{\Pi}(C_1^*, C_2^*)-\Pi(C_1^*, C_2^*)\geq\delta(1-\pi)V_N(C_1^*, C_2^*). \qquad \text{(B1)}$$

Similarly, given the monopoly bank will not gamble, it is best for it to choose the deposit contract (C_1, C_2) specified in Proposition (3.1). To ensure that the bank will not gamble, condition (18) becomes

$$\tilde{\Pi}(C_1, C_2)-\Pi(C_1, C_2)\leq\delta(1-\pi)V_N(C_1, C_2). \qquad \text{(B2)}$$

For some given parameters of π, R_H and δ, it is always possible to have

$$\tilde{\Pi}(C_1^*, C_2^*)-\Pi(C_1, C_2)=\delta(1-\pi)V_N(C_1, C_2). \qquad \text{(B3)}$$

If (B3) is satisfied, then both (B1) and (B2) are true as $\tilde{\Pi}(C_1^*, C_2^*)\geq\tilde{\Pi}(C_1, C_2)$, $\Pi(C_1, C_2)\geq\Pi(C_1^*, C_2^*)$ and $V_N(C_1, C_2)\geq V_N(C_1^*, C_2^*)$. So for the set of parameter values such that (B3) holds, the monopoly bank may choose either to gamble or to play safe.

Now we show that in R_H and π space, for a given delta, the boundary specified in (B3) lies below the boundary where (B2) holds as an equality and above the boundary where (B1) holds as an equality. To simplify comparison, we fix the value for π. Then, we can select the appropriate value for R_H such that (B3) holds. In this case, (B2) is satisfied. To ensure (B2) holds as an equality, we have to increase the value of $\widetilde{\Pi}(C_1, C_2)$. Since $\frac{\partial\widetilde{\Pi}(C_1,C_2)}{\partial R_H} > 0$, the R_H which ensures that (B2) is an equality must be greater than or equal to the R_H for which (B3) holds. So the boundary of (B3) lies below the boundary where (B2) is an equality. Similarly, one can show that (B3) also lies above the boundary where (B1) is an equality.

In R_H and π space, multiplicity of equilibria occurs in the area bounded by the boundaries of (B1) and (B2). So the sufficient condition to ensure no gambling is to choose the parameters of R_H and π such that they lie below the boundary of (B1). In what follows, we characterise the general properties of this boundary.

(1) Rewrite the no-gambling condition as

$$\widetilde{\Pi}(C_1^*, C_2^*) - \left[1 + \frac{\delta(1-\pi)}{1-\delta}\right]\Pi(C_1^*, C_2^*) = 0. \tag{B4}$$

Note that the contract offered by the gambling bank, (C_1^*, C_2^*), must satisfy the first order condition

$$u'(C_1^*) = R_H u'(C_2^*), \tag{B5}$$

and the binding participation constraint

$$\lambda u(C_1^*) + (1-\lambda)u(C_2^*) = \underline{U}. \tag{B6}$$

So, it is clear that $\frac{\partial C_1^*}{\partial \pi} = \frac{\partial C_2^*}{\partial \pi} = 0$, $\frac{\partial C_1^*}{\partial R_H} < 0$ and $\frac{\partial C_2^*}{\partial R_H} > 0$. This implies $\frac{\partial \Pi(C_1^*,C_2^*)}{\partial \pi} = 0$.

Using (B5) and (B6), one can show that

$$\frac{\partial \Pi(C_1^*,C_2^*)}{\partial R_H} = \lambda(R_H - R)\frac{\partial C_1^*}{\partial R_H} < 0 \tag{B7}$$

Applying the envelope theorem, one obtains,

$$\partial\widetilde{\Pi}/\partial R_H = \pi(1-\lambda c_1^*) > 0, \tag{B8}$$

and

$$\frac{\partial\widetilde{\Pi}}{\partial\pi} = (1-\lambda c_1^*)R_H - (1-\lambda)c_2^* > 0. \tag{B9}$$

Differentiating the no-gambling condition (B4) with respect to π and R_H yields

$$\left\{\pi(1-\lambda C_1^*) - \left[1+\frac{\delta(1-\pi)}{1-\delta}\right]\lambda(R_H - R)\frac{\partial C_1^*}{\partial R_H}\right\} dR_H + [(1-\lambda C_1^*)R_H - (1-\lambda)C_2^*]d\pi = 0$$

(B10)

Since both terms before dR_H and dπ are positive, the no-gambling condition must slope downward in R_H and π space.

Finally, note that if $\pi \to 1$ and $R_H \to R$, then (B4) holds. So the no-gambling boundary starts from $(R, 1)$ in the R_H and π space and goes asymptotically towards the R_H axis.

(2) An increase in δ increases the coefficient of the second term in (B4), $\left[1+\frac{\delta(1-\pi)}{1-\delta}\right]$. To maintain equality, R_H has to increase for a fixed π. Note that all no-gambling boundaries start from $(R, 1)$, so an increase in δ swivels the no-gambling boundary upwards in the R_H and π space.

Proof of Proposition 5

As is shown in the proof of Proposition 5 above, it is sufficient to specify the no-gambling condition (20) as

$$\widetilde{\Pi}(C_1^*, C_2^*; \beta, k) - \Pi(C_1^*, C_2^*; \beta) \leq (1-\pi)(1-\tau)\beta\delta V_N(C_1^*, C_2^*), \tag{B10}$$

where (C_1^*, C_2^*) is the optimal deposit contract offered by the gambling bank.

For $0 \leq \beta \leq \underline{\beta}$, $\tau = 0$, so (B10) can be rewritten as

$$\widetilde{\Pi}(C_1^*, C_2^*; \beta, k=0) - \Pi(C_1^*, C_2^*; \beta) \leq (1-\pi)[\beta\delta V_N(C_1^*, C_2^*) + k] \tag{B11}$$

Note that the deposit contract (C_1^*, C_2^*) is unaffected by either β or k, so to keep (B11) as an equality, a reduction in β must be compensated by an appropriate increase in k. This generates the downward sloping section of the no-gambling condition in β and k space.

For $\underline{\beta} \leq \beta \leq 1$, we rewrite (B10) as

$$\widetilde{\Pi}(C_1^*, C_2^*; \beta, k = 0) - \Pi(C_1^*, C_2^*; \beta) \leq (1-\pi)[(1-\tau)\beta\delta V_N(C_1^*, C_2^*) + k]. \qquad \text{(B12)}$$

Since the left hand side is independent of β and k, maintaining (B12) as an equality requires the right hand side to be a constant. Differentiating the right hand side with respect to β or k, one can obtain the slope of the no-gambling condition as

$$\left.\frac{dk}{d\beta}\right|_{NGC} = [\beta\tau'(\beta) + \tau - 1]\delta V_N. \qquad \text{(B13)}$$

It is clear from (B13) that the numerator is negative if $\beta \to \underline{\beta}$ and positive if $\beta \to 1$. So the no-gambling condition is U-shaped.