Riding for a fall? Concentrated banking with hidden tail risk[1]

14/11/2010

Marcus Miller, Lei Zhang, and Han Hao Li

University of Warwick

October, 2010

Abstract

The traditional theory of commercial banking explains maturity transformation and liquidity provision under assumptions of free entry, no asymmetric information and no excess profits. These assumptions seem entirely at odds with evidence from banking in the US and UK, prior to the recent crisis. So we extend the traditional theory to allow for market concentration (monopoly banking) and principle-agent problems in the form of excess risk taking, via the use of derivatives to boost measured profits. The paper ends with a brief discussion of possible regulatory changes to limit concentration and gambling.

Key Words: Money and banking, Seigniorage, Risk-taking, Regulation

JEL Classification: E41 E58 G21 G28

[1] While remaining responsible for the content, we would like to acknowledge the benefit of discussions with Sayantan Ghosal, John Kay and Frederic Schlosser and of attending FMG meetings at LSE organised by Charles Goodhart,.
Corresponding author: marcus.miller@warwick.ac.uk

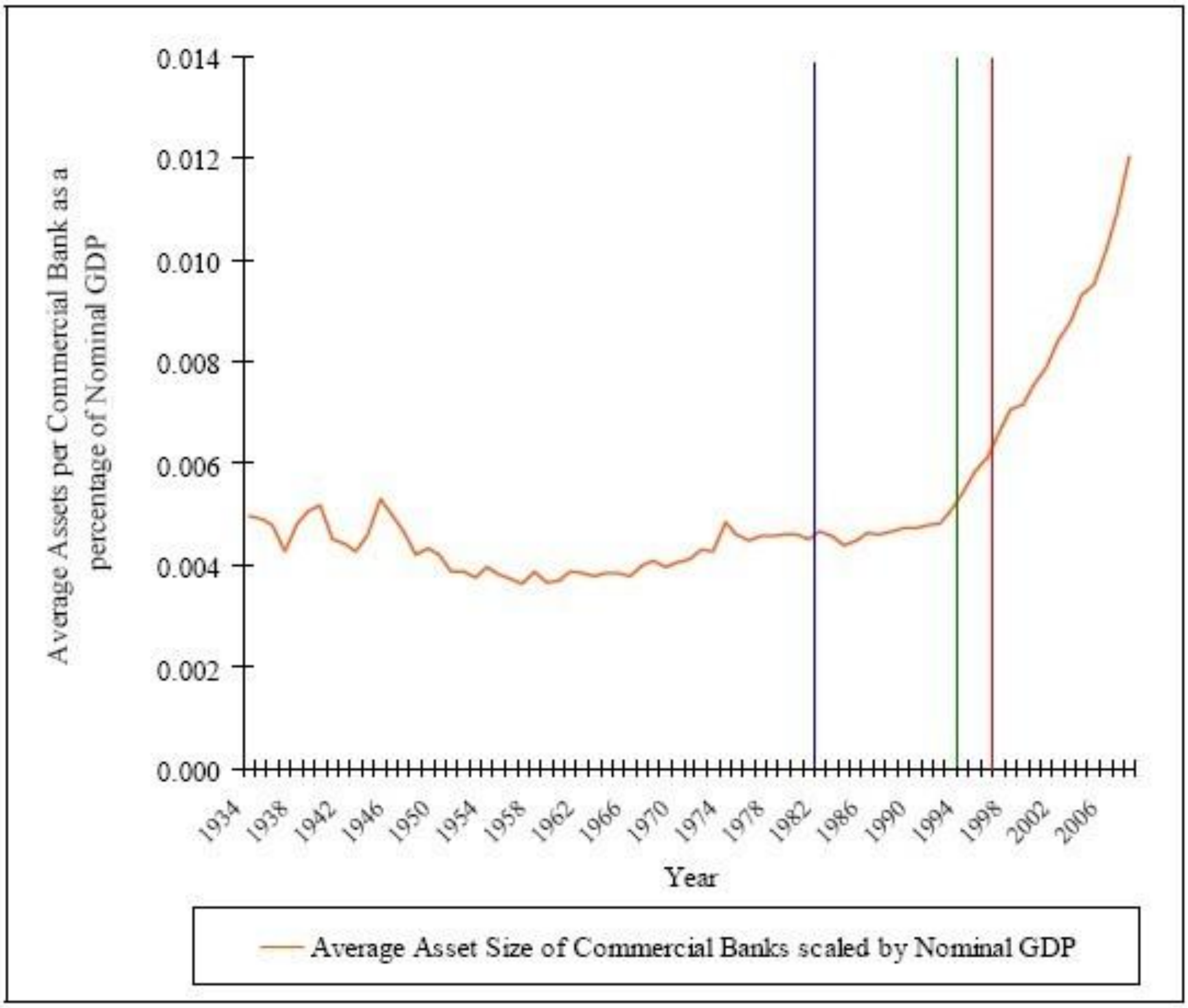
Introduction:

The recently established Independent Commission on Banking (ICB) is to study the UK banking sector, and to make recommendations on structural measures to promote stability and competition in banking, for the benefit of consumers and businesses. Before turning to measures that might be taken, we first consider the business of banking – how banks achieve maturity transformation and provide liquidity services.

Current theory, based on 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 the 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 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. For as Andy 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 80s, see Chart 1.

Chart 1: Average assets relative to GDP of US commercial banks[(a)]

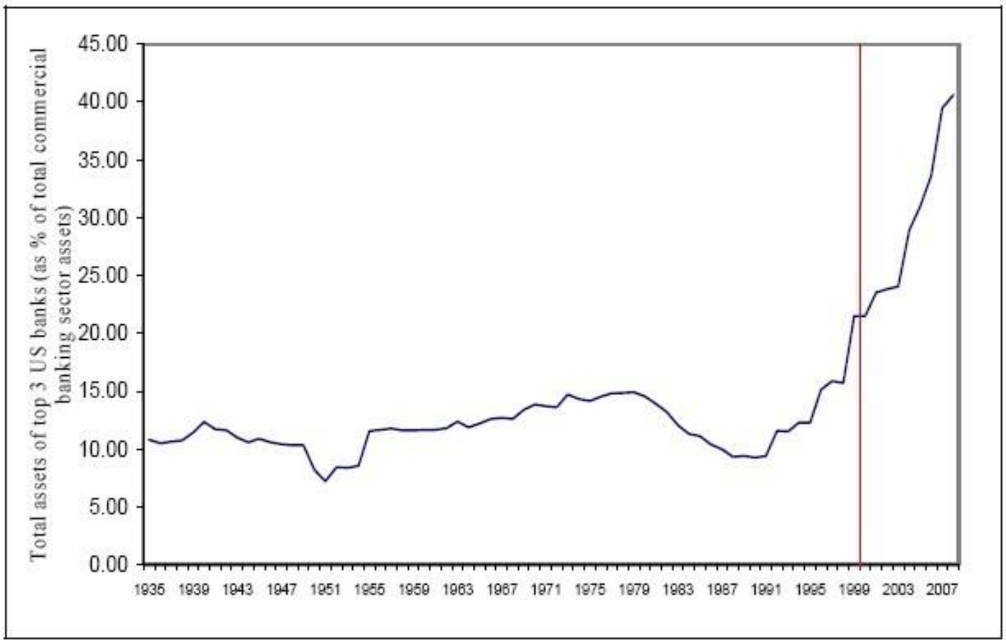


(a) Blue vertical line represents the 1982 Garn-St Germain Act, green vertical line represents the 1994 Riegle-Neal Act, red vertical line represents the Riegle-Neal Act coming into effect in 1997.
Source: FDIC and www.measuringworth.org

Moreover, the passage of the Glass-Steagall Act separating commercial and investment banking[2] seems to have succeeded in *limiting concentration* – at least until the passage of Gramm–Leach–Bliley Act of 1999, see Chart 2.

[2] As discussed in historical detail in Brant (2007) , for example.

Chart 2: Concentration of the US banking system[b]



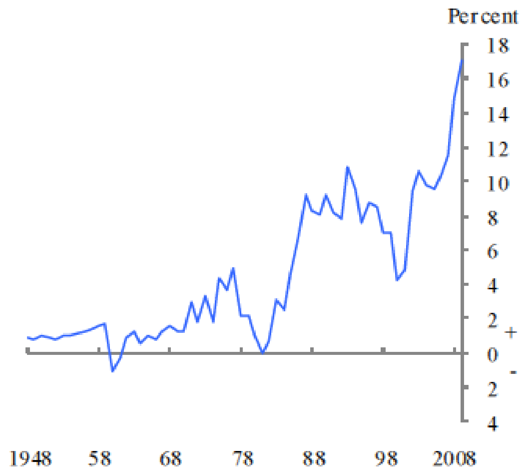
(a) Red line represents the Gramm-Leach-Bliley Act (1999) which revoked restrictions of Glass-Steagall
(b) Top 3 banks by total assets as a % of total banking sector assets
(c) Data includes only the insured depository subsidiaries of banks to ensure consistency over time - for example, non-deposit subsidiaries are not included.
Source: FDIC

Even if the assumptions of perfect competition and utilitarian banking may have been appropriate for US banking in the past, the simultaneous increase in both the volume of bank assets relative to GDP and in industry concentration evident from these charts suggest this is no longer true. What about the UK?

Markets for financial services in UK are quite concentrated. Among the Cruickshank Report (2000) on the state of competition in UK banking was the recommendation for 'a monopoly reference for SMEs under the Fair Trading Act 1973 in order to identify the existence of complex monopolies within the banking industry.' Evidence of concentration is provided in ICB (2010, section 3.11) where the market shares of the top 5 banking groups in the markets for each of personal current account, residential mortgages and SME banking is reported as 85%, 82% and 91% respectively. In the *The Future of Finance*, Paul Woolley (2010, page 124) notes that this may lead to excess profits: '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.' Evidence of the sharp rise in the profitability of banking in the UK is provided by Haldane et al. (2010), where they report 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). [Italics added.]

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



Sources: ONS and Bank calculations.

In this paper therefore, the traditional theory of banking is modified in two respects. First, in addition to competitive behaviour, we allow for market concentration: in particular, we study the effect of monopoly in the banking system. We find that a monopoly bank restricted to normal banking business of matching savings with real investment will continue to satisfy the first order conditions for inter-temporal efficiency; but it will transfer the social benefits of banking from its customers to the bank. These benefits can be capitalised into a 'franchise value', which represents, crudely speaking, the value of 'seignorage' - the right to create money.

Allowing banks to depart from the basic business of commercial banking by taking on substantial unsecured risk changes things more radically. Banks may be tempted to 'gamble' - in particular by taking on tail risk - collecting the upside, and leaving the downside to the depositors or (given deposit guarantees) to the tax payer. With asymmetry of information and limited liability, there is no guarantee of social efficiency.

Could the seignorage profits accruing to monopoly bank check the incentive to gamble? Not if gambling is endogenous in the sense that the banks can reshape the distribution of risk in favour of current high profits and large future losses borne by other parties. Nor, of course, if market dominance brings with it access to official bail-outs for those 'too big to fail'.

While the principal objective of the paper is to extend the traditional theory of banking by adding concentration and excessive risk-taking as described above - a task that involves grafting the gambling model of Hellman et al. (2000) onto the base-line model of Allen and Gale (2007) - we also briefly examine in conclusion various regulatory measures that might address the problems posed by lack of competition and excessive risk taking. Particular attention is drawn to two features: that of creating tail risk and of banks that are too big to fail.

Something we do not study in this paper is the impact of liquidity shocks on asset prices and banking stability. A recent paper by Douglas Gale does so,[3] but this too assumes competitive banking and full information. There are therefore no principal-agent problems, but the externalities involved in fire-sales call for regulation to increase liquidity holdings.

Section 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 - without any asymmetry of information - affects the size and profitability of banking. This is achieved by comparing the optimal 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 dates to mature, and converts one unit today into $R > 1$ units two dates 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

[3] In Gale (2010), a rise in liquidity preference on the part of these depositors forces banks to sell long term assets and acquire liquidity from a second group, namely risk neutral long-term investors. In the process, the fire-sales of long-term assets can have a substantial negative effect on asset prices, all the greater if depositors are more risk-averse. Fire-sale externalities suggest the need for liquidity requirements.

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

$$U(c_1, c_2) = \lambda U(c_1) + (1 - \lambda)U(c_2) \tag{1}$$

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

Before seeing what a bank can offer, consider what capital markets may achieve. Assume specifically that potential depositors can exchange their endowments with each other for early and late consumption goods in capital markets to ensure that[4] $(c_1, c_2) = (1, R)$. This implies an ‘outside option’ which generates utility of

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

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

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

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)

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

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

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

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.

Competitive Banking

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

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, $(1-\lambda c_1)R - (1-\lambda)c_2 = 0$. This contract has the feature that: $1 \leq c_1 \leq c_2 \leq R$.

The argument for this proposition, and properties of the equilibrium, are outlined in detail in Allen and Gale (2007, Chap 3.3). As Diamond and Dybvig (1983) have shown, such a model is prone to a bank run if early liquidation of 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 just such a bank run that precipitated the demise of Northern Rock, as discussed in Dewatripoint et al. (2010, p.87-88).

This competitive banking solution can be illustrated with the aid of 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 bounded from below 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 averse utility function, it is clear that $1 \leq c_1 \leq c_2 \leq R$.

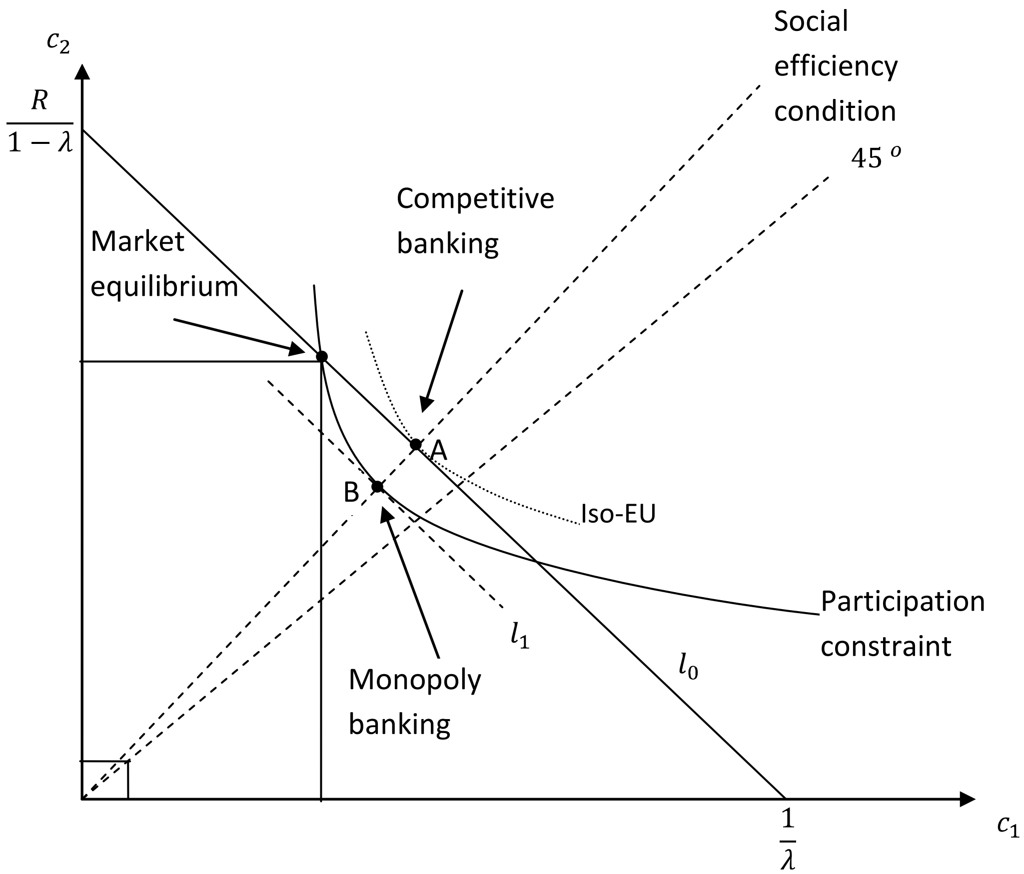


Figure 1. Competitive monopoly banking with no gambling

Monopoly Banking

How would market concentration affect the outcome? Essentially by allowing for positive profits, as can be seen from examining the case of monopoly. A risk-neutral monopoly bank is assumed to maximise its 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

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

subject to

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

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 social 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

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

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. ◊

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, $\lambda U(c_1^*) + (1-\lambda)U(c_2^*) = \underline{U}$.

The following proposition summarises some results of comparative statics for the competitive equilibrium and for optimal monopoly banking contract.

Proposition 3:

(1) An increase in R will increase the value of the outside option. Consumption at date 2, c_2^*, increases, with a rising spread between c_1^* and c_2^*.

(2) An increase in λ, the fraction of early consumers, has no effect on the outside option. So c_1^* goes down and c_2^* goes up.

(3) An increase in the utility, $\underline{U}$, associated with the outside option will result in an increase in consumption in both dates.

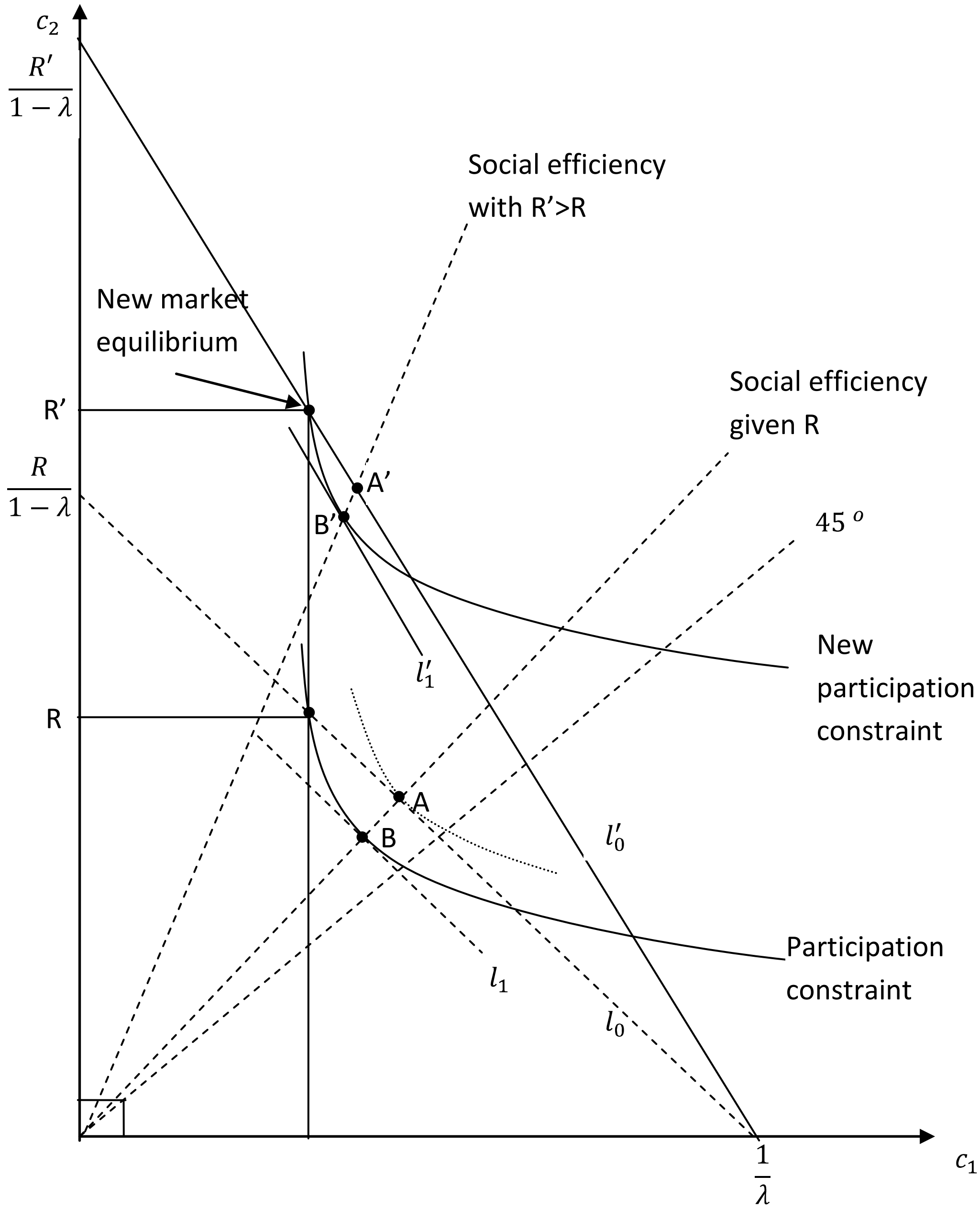


Figure 2. Effects of a 'productivity shock'

The impact of a positive ‘productivity shock’ is shown in Figure 2. An increase in R shifts market equilibrium upwards, shifting the resource constraint for the bank and improving the outside option for the depositors. This is reflected in an upward shift of the participation constraint shown in the Figure and the clock-wise rotation of the iso-profit functions. The effect on banking outcomes is to shift them from A and B to A’ and B’. The impact on competitive equilibrium is straightforward: the benefits of higher returns on the long asset are passed on to consumers, with the higher ‘spread’ of consumption reflecting more inter-temporal substitution. For the monopoly outcome, explicit account has to be taken of improved outside option, as shown at B’ where the new iso-profit function, $l_1^{'}$, is tangent to the new participation constraint. This limits ability of the monopolist to extract the benefits of productivity as private.

[The effects of changing λ and $\underline{U}$ can also be seen with the aid of Figure 2. An increase in λ implies a clockwise rotation of the iso-profit function l_1. Given that the feasible set remain the same, the optimal contract will shift to a new tangency point on the (old) participation constraint to the left of B. So the difference between the two dated consumptions is increased. With an increase in $\underline{U}$, the feasible set shifts upward to the right, resulting in increases in consumptions at both dates.]

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.

Section 2: Banking with tail risk

In the previous section, bank behaviour under different competitive conditions is calculated on the assumption that the banks were investing safely. But, as became apparent in the financial crisis, banks in the US and elsewhere had in fact been taking on a great deal of risk.

To see why, consider the incentives facing those allocating bank resources, noting that 'investors will reward a manager handsomely only if the manager consistently generates excess returns, that is, returns exceeding those of the risk-appropriate benchmark. In the jargon, such excess returns are known as "alpha".' Rajan (2010, p.138). In this context, there is it would seem a strong temptation to invest in risky assets where the downside is very rarely observed, i.e. take on tail risk[5]. In their 2008 paper on hedge funds , Foster and Young describe how writing out-of-the-money puts can generate high rates of profit for a while, but - given the asymmetry of information characterising hedge funds, for whom investment strategies are proprietary information - these high rates of return may well be mistaken for riskless profit, and paid out as bonuses and dividends. This is what they call the 'hedge fund game' – the creation of 'fake α' so as to get supernormal profits [6] . But what about banks?

In the chapter entitled 'Betting the Bank', Rajan (2010) describes how tail risk is taken on as follows:

> 'Suppose a financial manager decides to write earthquake insurance policies but does not tell her investors. As she writes policies and collects premiums, she will increase

[5] Tail risk is the risk of an asset or portfolio of assets moving more than 3 standard deviation from its current price in a probability density function.

[6] Lo (2001) gives an example: the probability of S&P 500 being lower than 7% of current value in a 1-3 month period is pretty low, so a short position in 7% out of the money put option has a good chance of making the fund manager 'excess returns' when the option expires worthless. He claims that strategies of this type do not only exist in theory, but a 'well-known artifice employed by unscrupulous hedge-fund managers to build an impressive track record quickly..." (Lo, 2001, p.23)

> her firm's earnings. Moreover, because earthquakes occur rarely, no claim will be made for a long while. If the manager does not set aside reserves for the eventual payouts that will be needed … all the premiums she collects will be seen as pure returns, given that there is no apparent risk. The money can all be paid out as bonuses or dividends… Because she has set aside no reserves [when the earthquake occurs] she will likely default on the claims, and her strategy will be revealed for the sham it is. But before that… she may have salted away enough in bonuses to retire comfortably.'

Although the manager in question is supposed to be working in a regulated financial institution, the strategy employed is, of course, essentially the same as that in Foster and Young's hedge fund game. How could this be? Possibly, as Woolley argues, it is because there is also substantial asymmetric information between the banks and their customers.

The use of such strategies could also 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 and UK, even after investment banks had paid out salaries and bonuses amounting to 60% of net revenues' Wooley (2010, page 121).

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.[7] The implication of the correction is that a substantial fraction of these earnings should be seen not as payments for value added, but as pure 'transfers' from the rest of the economy to the financial sector.

To see how 'tail-risk' investment (or more generally 'gambling'[8]) can generate such transfers to banks, we assume that the bank exploits the asymmetry of information to invest in a risky asset with mean return $\tilde{R}$, whose the true prospects for high and low returns are private information to the bank. The risks captured by the high and low returns to be realised in $t = 2$, $R_H > R$ and $R_L < R$ respectively, with probabilities π and $(1 - \pi)$. For simplicity, we only consider the case where $\tilde{R}$ is a mean-preserving-spread 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. As these high returns are not available outside banks, however, so there is no shift to the outside option.

---

[7] 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'.

[8] So as to allow for the downside to be less than to be less than 3 standard deviations from the mean.

Monopoly banking

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

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

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 cannot fulfil its contract to late consumers, and will be insolvent. What happens in this case is not apparent to the depositors ex ante, however.

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 even might be bailed out by the government in the low state.

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

Proposition 4:

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

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

Proof: 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.

If R_L is low enough, the bank cannot honour the contract to the late consumers in the low state. 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 in the second part of Proposition 4. (Since case (1) has the same deposit contract as that under certainty and no default from the bank, we use case (2) to represent gambling.)

It is worth noting that the gambling bank would offer a deposit contract with dated consumptions further apart than that by a non-gambling bank. This is because, the effective returns on the long asset for the gambling bank is R_H larger than that for the non-gambling bank. The difference between the dated consumption must increase. As c_1^* is reduced, the gambling bank will, moreover, keep less liquidity than the non-gambling bank.

The optimal deposit contract with a gambling monopoly is shown in Figure 2, using the same axes as in Figure 1. Since the effective returns for the long asset have apparently increased to R_H, the iso-profit functions show a clockwise rotation (see l_0' and l_1') and the efficiency locus shifts as if there has been a positive productivity shock. With no change in the outside option, however, the deposit contract shifts along the participation constraint. Specifically, the optimal deposit contract offered by the gambling bank is at C 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. Clearly bank profits will rise sharply

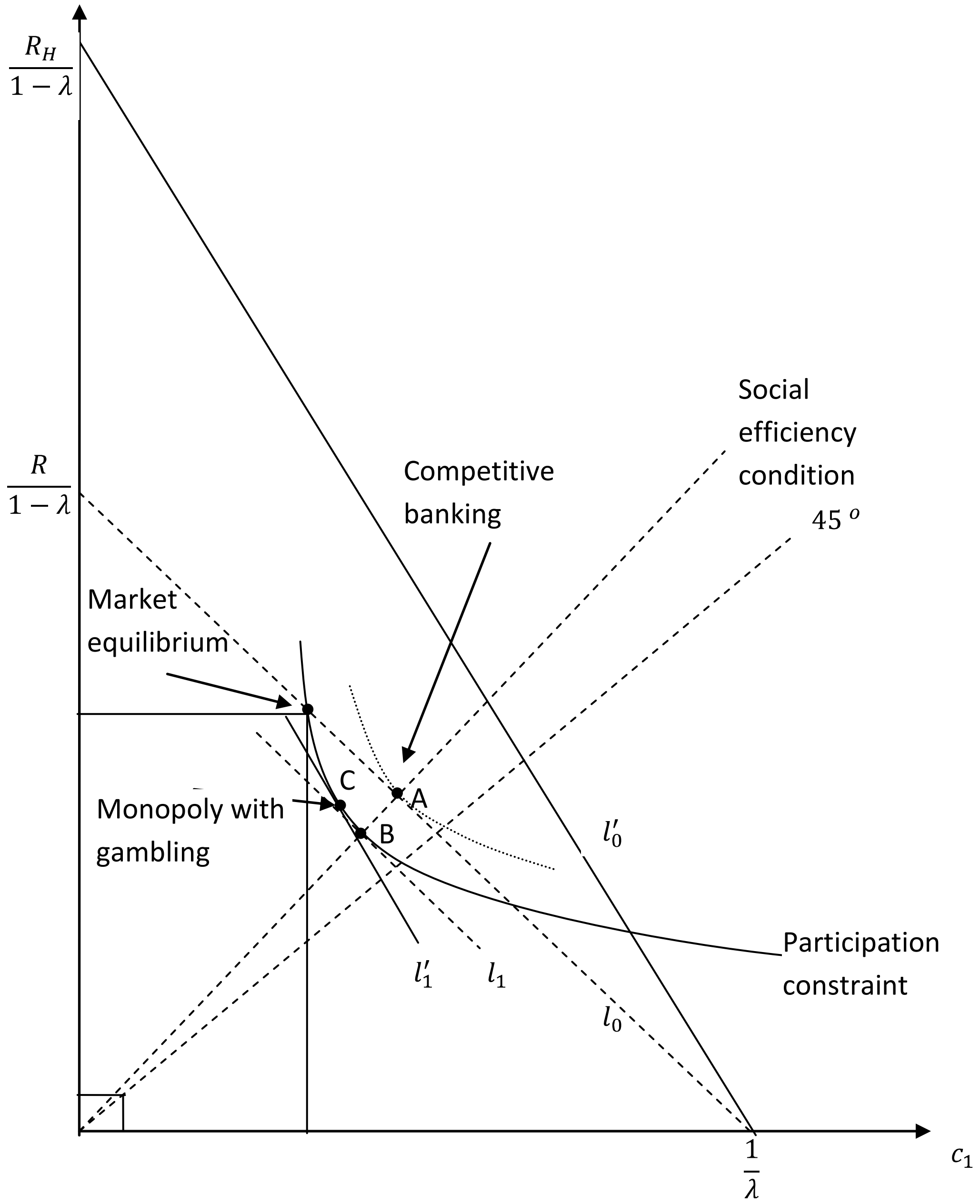


Figure 3. Monopoly with "fake α" ($R_H > R$)

Competitive banks

Note that increased competition would have a dramatic effect on banks' incentive to gamble. With perfect competition, for example, profits to prudent banks are zero, i.e. there is no franchise value. So all banks will gamble in the equilibrium. Hellmann, Murdock and Stiglitz (2000) consider two checks on gambling behaviour: either to impose minimum capital requirements or to limit deposit rates so as 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.

Section 3: Franchise values as a check on gambling

Even without capital requirements or regulation on deposit rates, Bhattacharya (1982) pointed out that the threat of losing its franchise could nevertheless 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'.

We investigate this possibility in a context where hidden tail risk can be taken on. We assume that the bank can choose either to gamble or not to gamble; and at the end of date 2, the bank will distribute all its profits – if any - to its shareholders. But if the bank chooses to gamble and fails, losses are taken over by the government and the bank is closed down: i.e. there is no bail out. Our calculations suggest that it is an illusion to believe that franchise values based on monopoly profits will prevent gambling – even when there are resolution procedures in place. The reason is that the extent of gamble is endogenous – it reflects the portfolio choices of the banks. Given the protection of asymmetric information, the bank can – in principle - always find tail risk returns that provide sufficient current profits to justify risking the franchise value of seigniorage.

To compute this franchise value, we consider a repeated game with infinite 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:

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

If the bank gambles, the value function is

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

This simply means that the gambling bank can capture current round profits and future discounted profits if the gamble succeeds. Simplifying (15) yields,

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

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. The no-gambling-condition is essentially the same as that in Hellmann, Murdock and Stiglitz (2000).

Now we characterise the boundary of the no-gambling-constraint (where (18) holds as an equality) in terms of R_H, π and δ.

Proposition 5:

(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: (1) First we show that for given values of π and δ, there is a threshold such that (18) holds as an equality. Using envelope theorem, one can show that, $\partial\widetilde{\Pi} / \partial R_H = \pi(1 - \lambda c_1^*) > 0$. For small enough R_H, (18) holds; for large enough R_H, the inequality of (18) is reversed; so these must be a R_H, (18) holds as an equality. Second, using the envelope theorem, we also have $\frac{\partial\widetilde{\Pi}}{\partial\pi} = (1 - \lambda c_1^*)R_H - (1-\lambda)c_2^* > 0$. So the boundary of the no-gambling-condition must be downward sloping. (2) An increase in δ increases the discounted franchise value on the right hand side of (18) without affecting the left hand side. Therefore, it will shift the no-gambling-constraint upwards.

The boundary of the no-gambling-condition is shown in Figure 3, where the horizontal axis indicates the higher returns for the gambling asset and the vertical the probability of gambling success. When δ is low, say at δ_L, the boundary is as represented by the downward sloping curve $R_H(\pi; \delta_L)$, below which the no-gambling-constraint is satisfied. For higher values of δ, representing more far-sighted banking, the boundary will swivel upwards as shown by the arrow.

As profits for the gambling monopoly increase in both π and R_H,so - in the absence of appropriate regulatory controls - the bank will have the incentive to invest in a long asset with substantial downside risk.

Using the standard definition for tail risk, i.e., that the low return, R_L, is at least 3 standard deviations away from the expected return, $\tilde{R} = \pi R_H + (1-\pi)R_L$, our binomial model with mean preserving spread, (i.e. $\tilde{R} = R$) implies that *all investments with* π *above a critical value involve tail risk*. Straightforward calculation indicates that this threshold π is 0.9, as shown by the dashed horizontal line in the Figure.

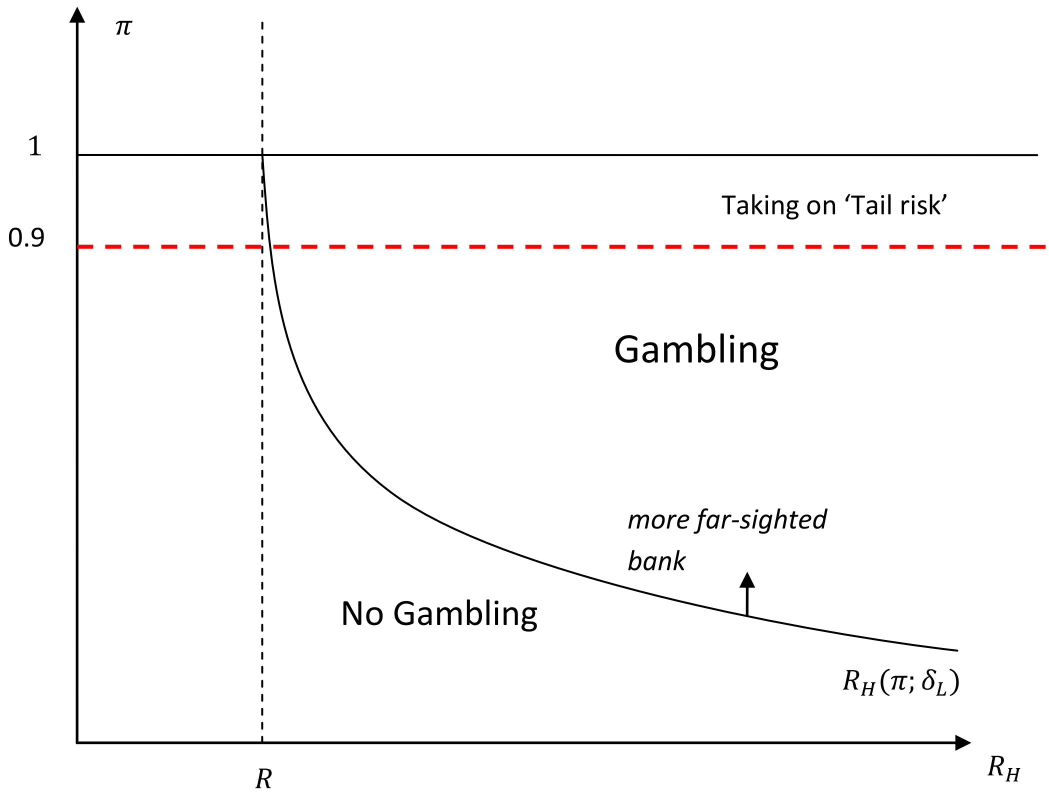


Figure 4. Boundaries of no-gambling-constraint

How increasing the discount factor can reduce incentive to gamble is illustrated numerically in the table below, which reports the expected profit in each round for the gambling monopoly bank (for which the upside is given by R_H, but the downside limited to the loss of franchise). For the parameters $R = 2, \lambda = 0.5, \gamma = 2$, (see Gale 2010), the monopoly profit without gambling turns out to be 0.057, while profits with gambling are shown below:

	$R_H = 3$	$R_H = 4$	$R_H = 5$
π=0.98	0.197	0.392	0.624
π=0.95	0.191	0.38	0.605
π=0.9 (TR threshold)	0.181	0.36	0.573
π=0.7	0.141	0.28	0.446
π=0.5	0.1	0.2	0.318

Table 1: Expected flow of profit for monopoly bank that gambles.

Note that the entries with $\pi \geq 0.9$ are the expected one round profits by taking on tail risk. In a context where monopoly profit without gambling is about 0.06, the increase in expected profit provided by gambling is enormous: the expected profit given in the top right hand corner is more than 10 times that figure. An example of how this might be achieved is by writing out-of-the-money puts on the stock market.

The higher the discount factor, the less the incentive for the bank to gamble, as the franchise is valued higher.[9] Thus, for a discount factor, $\delta = 0.61$, only the bottom left corner entry will satisfy the No-Gambling-Constraint given by equation (18); but if δ rises closer to 1, (e.g. $\delta = 0.94$) all entries except the three top cells in the last column will satisfy the No-Gambling-Constraint. Clearly, the opportunity for a monopoly bank to take on tail risk can substantially undermine its incentive to behave prudently.

Although the model does not include a capital requirement per se, the franchise value associated with such market power plays much the same role as capital requirements in a competitive environment. The franchise measures the bankers' 'skin in the game' – what s/he will lose if the gambling fails – and this is reflected in the spread between the return on asset and effective rate paid to depositors.

In the bank's balance sheet, the equity will simply be the present discounted value of the monopoly profits. Loosely speaking, the ratio of equity to deposits is the capital adequacy ratio. In the benchmark simulation, where the monopoly profits are about 6% of deposits, the

[9] The franchise value is the expected no-gambling profit divided by (1-δ), where δ is the discount factor.

capital adequacy ratio would be about 12% if the discount factor is half. As can be seen from the simulations, such a capital adequacy ratio cannot deter even moderately attractive gambling. If one wants to introduce additional regulatory capital requirement on the part of the bank's shareholders (as in Hellmann, Murdock and Stiglitz (2000)), the incentive to gamble will be reduced, but our results will remain qualitatively the same.

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.

Section 4: Tentative thoughts on Regulatory reform:

In this section, the preceding analysis is used in a heuristic discussion of options for the reform of the UK banking system. We distinguish in particular between reforms related to structure of banks (the degree of leverage, for example) and those related to markets (such as the degree of concentration). For this purpose, we use figures with minimum capital requirement on the vertical axis (to represent variations in bank leverage), and market concentration on the horizontal axis (acting as a proxy for franchise value, assuming that high concentration implies high franchise value).

The no gambling condition – defining the shaded area of prudential banking - is the downward sloping schedule LN in Figure 6. It's slope reflects the finding in Hellman et al.(2000) that - in ensuring prudential behaviour - there is a trade-off between bank's profitability (franchise value) and the official capital requirement: 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. The point L in Figure 6 so will represent the minimum capital requirement needed under perfect competition.

How will higher tail risk impact on prudential banking? The analysis of this paper has confined itself to the case where there is no capital requirement, i.e. to the horizontal axis in the diagram, focusing on the special case of a monopoly bank as at t point M in the diagram. What we have shown above is that heightened 'tail-risk' (i.e. increasing R_H and/or π), shifts the no gambling condition to the right: and if point N shifts far enough, the monopoly bank will be willing to risk its franchise. As Hellman et al (2000) have shown that a more attractive gamble will call for increased capital for competitive banks (i.e. shift L upwards), we may conclude that, with higher tail risk, the no gambling condition NN' schedule as a whole shifts to the right, as indicated by the arrow in the diagram.

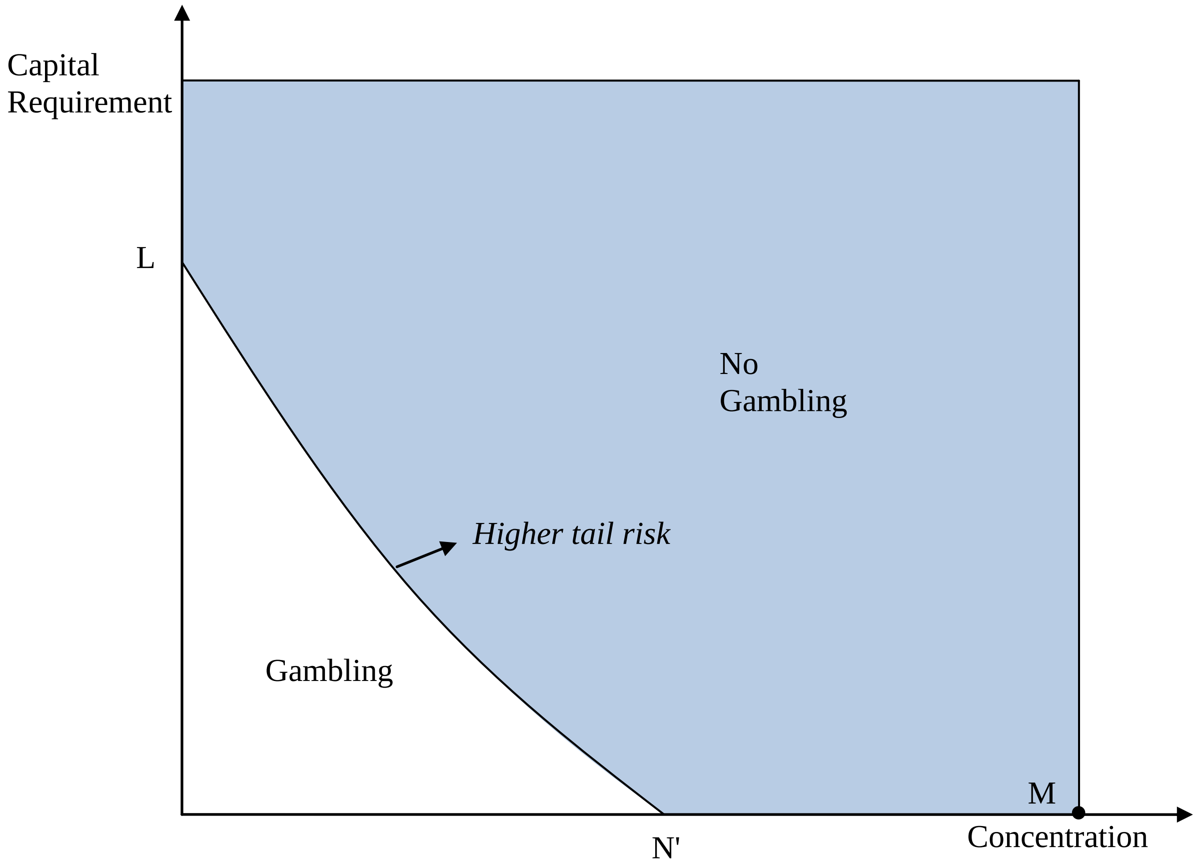


Figure 5. Capital Requirements, Market Concentration and Gambling

As Rajan points out, however, an increase in concentration increases the likelihood of an official bailout, as banks become ‘Too Big To Fail’ (TBTF). Taking this factor into account will of course greatly encourage gambling, as the effect of franchise values is 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 we show in Figure 6. Let TBTF indicate the point at which banks become Too Big To Fail. To the left of this point, the risk of losing franchise value is sufficient to check gambling: to the right, however, the probability of bailout offsets monopoly rents and encourages gambling. So the region for prudential banking becomes more U-shaped - as indicated by the shaded area bounded by LN’R in the Figure.

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 others 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, N'R, of the NGC condition rightwards as shown by the arrow in the figure.

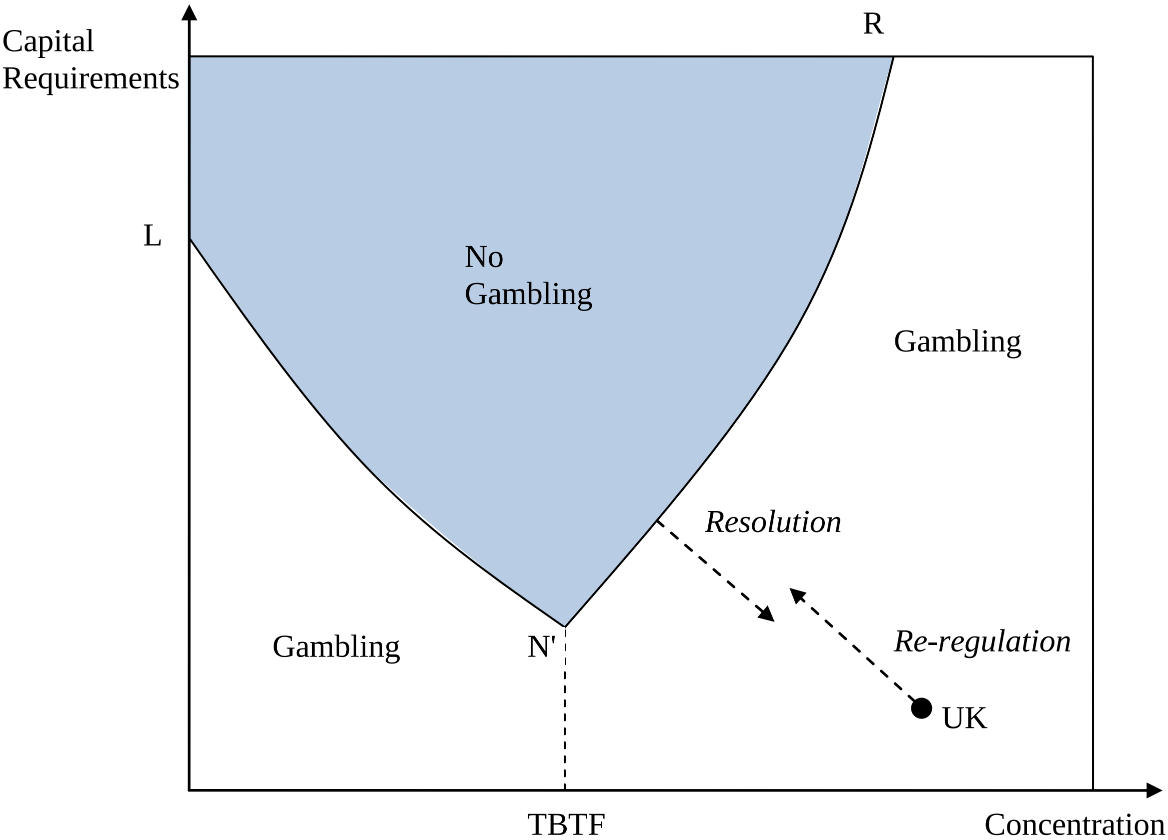


Figure 6. How bailouts increase the risk of imprudent banking

Policy implications for policy reform in the UK may be discussed heuristically using the figure. To start with, it is clear from the evidence that, given the present level of capital requirements, 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 achieved by increasing capital requirements

– preferably on unweighted assets, to limit gaming of the rules. As for market structure, the evidence suggests that risky M&A 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) – a powerful argument for reducing concentration, namely to limit gaming of the state! Together, these imply a shift to the northwest as shown by the arrow labelled re-regulation. The regulatory agenda will, evidently, be considerably assisted by the improvement in the resolution regime.

This whole discussion is, however, subject to an important caveat, namely that such regulatory improvements can be undermined unless draconian steps are taken to reduce the asymmetric information in the financial system. As Rajan (2010 p.152) observes:

> ‘The problem of tail risk taking is particularly acute in the modern financial system, where bankers are under tremendous pressure to produce risk-adjusted performance. Few can deliver superior performance on a regular basis, but precisely for this reason, the rewards for those who can are enormous. The pressure on the second-rate to take tail risk, thus allowing them to masquerade as superstars for a while, is intense.’

As Rajan shows 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. Putting it simply, if asymmetric information is the problem, then transparency must be part of the solution - together with changes in the law governing what is legitimate business in financial services.

The ‘doom loop’ and ‘doomsday’ scenarios sketched by Haldane (2010) and Wolf (2010) provide graphic warning of what can happen without decisive action to correct distorted incentives in a sector which has been described as ‘the brain of the modern economy’.

Section 5: Conclusion

Concentration in UK banking is higher than in the US: and while US banks are roughly equal in size to GDP, in the UK they are 5 times as large as GDP. Some outside observers have argued that ‘bank reform may not be a top priority in UK’, and that, ‘the September 2011 report by the ICB will shortly be placed in a file marked “Worthy but irrelevant”.’ (FT LEX Column 2010). A positive analysis of banking behaviour with high concentration and excessive risk taking -- not to mention recent calamitous developments in Iceland and Ireland -- suggest this would be a great mistake.

References:

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

Allen, Franklin & Douglas Gale, 2004. "Financial Intermediaries and Markets," *Econometrica*, Econometric Society, vol. 72(4), pages 1023-1061, 07.

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

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

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.

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

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

Ehrenberg, Tim () 'paper on banking using DD model'.

Foster Dean P. & H. Peyton Young, 2008. "The Hedge Fund Game: Incentives, Excess Returns, and Piggy-Backing," *Economics Series Working Papers* 378, University of Oxford, Department of Economics.

Gale, Douglas (2010), 'Liquidity and Monetary Policy', forthcoming in Jagjit Chadha (ed.) *Quantitative Easing*, Cambridge: Cambrdige University Press.

Gertler M. & Nobuhiro Kiyotaki (2010), 'Financial Intermediation and Credit Policy in Business Cycle Analysis', Working paper.

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

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

Hellmann, Thomas F. & Kevin C. Murdock & Joseph E. Stiglitz, 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.

HM Treasury (2010), ‘A new approach to financial regulation: judgement, focus and stability’, Presentation to Parliament, July.

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

IMF (2010), ‘A Fair and Substantial Contribution of the Financial Sector’. Interim report for the G-20, April 2010.

Johnson, Simon and James Kwak (2010), *13 Bankers*, New York: Pantheon Books.

Kay, John (2010), ‘Finance spreads its own risks but left ours alone’, Financial Times, July 6 2010.

Keen, Michael (), ‘Paper on bank levies and IMF plans’.

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

Lockwood, Benjamin (2010), ‘How Should Financial Intermediation Services be Taxed?’ paper presented at the 2010 CBT Summer Symposium.

Lo, Andrew W. (2001), “Risk management for hedge funds: introduction and overview,” *Financial Analysts’ Journal*, Nov/Dec, 2001, 16-33.

Korinek, Anton (2009), “Systemic Risk-Taking: Amplification Effects, Externalities, and Regulatory Responses”, mimeo, University of Maryland.

Moore, John (2009), ‘Contagious Illiquidity’, presented at Leo Hurwicz Lecture, Latin American Meeting of the Econometric Society, Buenos Aires Oct.

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

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

Rochet (2003), "Why are there so many Banking Crises?", *CESifo Economic Studies*, vol. 49, n. 2, Ifo Institute for Economic Research, Munich , p. 141-156.

Sachs, Jeffery (2010), ‘Rob rich bankers and give money to the poor’, The Times, March 10 2010.

Turner et al. (2010), *Future of finance: The LSE Report.* London: LSE.

Westlake, Melvyn (2010), *Regulators to stand their ground over Basel III*, Global Risk Regulator, May 2010 Volume:8 Issue:5.

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