Bank Competition and the Design of Regulation*

Franklin Allen
University of Pennsylvania

Elena Carletti
Center for Financial Studies

Robert Marquez
University of Maryland

April 20, 2005

Abstract

Market discipline for financial institutions can be imposed not only from the liability side, as has often been stressed in the literature on the use of subordinated debt, but also from the asset side. This will be particularly true if good lending opportunities are in short supply, so that banks have to compete for projects. In such a setting, borrowers may demand that banks commit to "monitoring" by requiring that they use some of their own capital in lending, thus creating a market-based incentive for banks to hold capital that stems purely from the asset side of the bank's balance sheet. Borrowers can also provide a bank with incentives to monitor by allowing the bank to reap some of the benefits from the loan, which accrue only if the loan is in fact paid off. Since borrowers do not fully internalize the costs of capital to the bank and of the deposit insurance, the level of capital required by the market may be above the one chosen by a regulator maximizing social welfare. This implies that capital requirements may not be binding.

1 Introduction

The amount of capital used by banks has varied substantially over time. During the nineteenth century capital ratios were much higher than in recent times. Berger et al. (1995)

*The usual disclaimers apply. Contact information: Robert Marquez, R. H. Smith School of Business, University of Maryland, College Park, MD 20742, *rmarquez@rhsmith.umd.edu*.

report that in the 1840's and 1850's banks in the U.S. had capital ratios of around 40 to 50 percent. After that they fell until they reached the 6 percent to 8 percent range in the 1940's where they stayed until the end of the 1980's. Berger et al. (1995) suggest that this was the result of reforms such as the creation of the Federal Deposit Insurance Corporation (FDIC) in 1933 and restrictions on the interest rates banks could pay on deposits. In the 1980's regulation of bank capital became more important as other kinds of bank regulation were reduced. For example, in December 1981 the Federal Reserve Board and the Comptroller of the Currency announced formal capital requirements depending on the size of the bank and shortly afterwards the FDIC announced a 6 percent requirement regardless of bank size (Ashcraft, 2001). Historically, different countries regulated bank capital in different ways, resulting in a concern that some countries' banks might have a substantial advantage because of lower capital requirements. This concern led to the international coordination of capital requirements with the introduction of the first Basle Accord in 1988. Although capital constraints appeared to be binding during the 1980's, in the 1990's banks increased their capital to a level where at least in the U.S. they were no longer binding for large banks (Flannery and Rangan, 2002).

The most common justification for capital regulation for banks is the reduction of bank moral hazard. Given the presence of deposit insurance, banks have easy access to deposit funds. If they hold a low level of capital, then there is an incentive for them to take on excessive risk. If the risky investment pays off, the banks' shareholders receive the payoff. On the other hand, if it does not, the bulk of the losses are borne either by depositors or by the body providing deposit insurance. Given the widely accepted view that equity capital is more costly for banks than other forms of funds, one would expect banks to ensure that they minimized the use of equity capital so that the capital constraint would always be binding. While in a dynamic environment one might expect there to be a buffer so that banks would be above the statutory requirement, in the 1990's many U.S. banks nevertheless had a surprisingly high level of capital. For many of them the amount of capital in excess of the regulated amount appeared to be well above what would be required as a buffer (Flannery and Rangan, 2002). In general, it seems that capital constraints often do not bind, but there has been little research to understand the likely cause of this important issue.

In this paper we develop a model of a competitive credit market where equity capital is costly but banks may nevertheless choose a level that is above the minimum regulated amount. The market failure is an agency problem within the firm between the shareholders and managers. Banks can help solve this agency problem by monitoring the firm. We model this agency problem by assuming that the greater is the amount of bank monitoring the greater is the probability the firm's investment is successful. An alternative interpretation of the model is that banks perform a screening function. The more effort they exert, the higher is the probability of obtaining a good project.

Bank monitoring thus has two effects. The first is that it increases the probability that the firm's loan is repaid. This provides an incentive for the bank to monitor. The second is that the firm's owners are also better off as a result of the monitoring. Bank loans may therefore be desirable from the firm owner's point of view.

This perspective on bank monitoring implies that higher interest rates on loans can be good not only for the bank but also for the firm. A higher loan rate gives the bank a greater incentive to monitor because it receives a higher payoff on average. This increased average payoff can also benefit the firm's owners if it exceeds the extra amount they pay the bank for the loan. This is not however the only way to provide banks with an incentive to monitor. In addition, the amount of equity capital the bank has affects its incentive to monitor in the usual way. The more capital there is, the greater the loss the bank's owners will face if the loan is not repaid and so the greater is their incentive to monitor. Thus incentives for the bank to monitor are provided by the loan rate and the amount of capital.

The other important determinants of bank monitoring are deposit insurance and the structure of the loan market. Following the rest of the literature on capital regulation, we take it as given that there is deposit insurance. The first case we consider is one where the demand for loans by firms with good projects exceeds the banks' supply of funds so that borrowers compete for funds. The second case is one where there is a shortage of good projects relative to the funds available so that the banks must compete for the firms' business and tailor their contracts to do so.

In the first case, when borrowers compete for funds, banks optimally choose to hold no capital since equity is more costly than deposits and limited liability protects them from having to repay depositors when their loans are not repaid. They also raise the interest rate on loans to its highest possible level and it is this which provides them with an incentive to monitor. In this context, we show that when the cost of equity is close to the cost of deposits a regulator interested in maximizing social welfare would impose a requirement that banks hold a positive amount of capital. This leads to improved monitoring and reduces the cost to the deposit insurance fund, an aspect which is not internalized by the banks. However, when the cost of equity is sufficiently high relative to the cost of deposits, the regulator will prefer to economize on the use of the costly input (capital) and will not impose any capital requirement. Either way, the bank would like to have as low a level of capital as possible and any capital constraint imposed by a regulator will be binding.

The case where banks compete for projects is more complex. Even in the absence of a regulator, banks will be required to hold a positive amount of capital. Capital acts as a commitment device for banks to monitor, which is good for borrowers. The result is that either the level of capital will be such that banks' profits are driven to zero or that banks will be all equity financed. This finding suggests that market discipline can be imposed not only from the liability side, as has been stressed in the literature on the use of subordinated debt, but also from the asset side. As bank capital must be set to attract borrowers, loan rates must also be set at the level which is most advantageous for borrowers. Interestingly, this is not necessarily the lowest rate consistent with non-negative profits for the banks because borrowers want banks to monitor and a higher loan rate provides better incentives for banks to do this. In equilibrium, the incentive effect of the loan rate will be balanced against the cost to the borrowers.

A regulator interested in maximizing social welfare will in general want to choose a different level of capital than the level dictated by market conditions. However, in contrast to the case where borrowers compete for funds, this socially optimal level can be above or below that chosen by the market. In particular, when the cost of equity capital is high and the cost of deposits is low, the regulator may want to impose a lower capital requirement than what would be preferred by the market. The reason is that the borrowers do not internalize the cost of equity capital and the cost of deposit insurance is low. In other words, any capital requirement set by a regulator would not be binding, as competition for borrowers leads banks to hold greater amounts of capital than is socially optimal. By contrast, if the cost of equity capital is close to the cost of deposits, then the capital requirement may be above the market level for the usual reason that the cost of deposit insurance is not internalized.

Section 2 outlines the model. Section 3 considers banks' choice of monitoring taking the loan rates and capital amounts as given. The case where borrowers compete for funds is considered in Section 4 while the case where banks compete for projects is analyzed in Section 5. Section 6 contains concluding remarks.

2 Model

Consider a simple one-period economy, with N banks and M firms. Each bank raises one unit of funds, and extends a loan to a borrower. Loans are risky but can be made safer through monitoring. The level of capital banks raise influences their incentives to monitor, and can be determined either from a competitive market or by a regulator maximizing social welfare, as described below.

Banks finance themselves with an amount of capital k at a cost r_E, and an amount of deposits $1-k$ at a cost r_D, with $r_E \geq r_D$. This assumption captures the idea that bank capital is a particularly expensive form of financing (see, e.g., Berger et al., 1995. Hellmann et al., 2000, and Repullo, 2004, make a similar assumption).

Banks extend loans to borrowers at a gross interest of r_L. Borrowers invest in risky projects with gross payoff of R when successful and 0 when not. Monitoring affects the probability of success of the projects. Specifically, each bank chooses a monitoring effort q, which for simplicity represents also the probability of success of the investment project (see, e.g., Besanko and Kanatas, 1993, Holmstrom and Tirole, 1997, Carletti, 2004, and Dell'Ariccia and Marquez, 2005, for studies with a similar assumption). Monitoring increases the expected return of the project and reduces the probability of failure. This captures the idea that monitoring can help in reducing an agency problem between the shareholders and managers of the firm, and thus increase value. The bank has an incentive to monitor because monitoring increases the probability of loan repayment. Monitoring, however, carries a cost cq^2. The convex cost function reflects the fact that it is increasingly difficult for a bank to find out more about a borrower.

Depositors are fully insured so that the deposit rate r_D does not depend on the risk of bank portfolios and, for simplicity, is exogenously determined. The loan rate r_L and the amount of capital k are set endogenously, and can be determined in one of two ways. They can either both be determined in a competitive market or the amount of capital can be set by a regulator who maximizes social welfare. The market solution will depend on the division of surplus between banks and borrowers, and we will distinguish between two cases. The market is always competitive in that all parties act as price-takers, but the allocation of surplus will depend on whether there is a shortage of firms with good investment projects ($N > M$) or whether there is a shortage of funds available to lend ($N < M$).

The timing of the model can be divided into 4 stages. First, the level of bank capital k is determined, either by the market or by a regulator. Second, the loan rate r_L is set by the market.[1] Third, borrowers choose the loan that is most attractive to them. Finally, banks

[1] Note that, in the absence of regulation, this timing structure is equivalent to assuming that the market

choose their monitoring effort q once the terms of the loan have been set and they have raised capital and deposits.

3 Bank monitoring

Before proceeding to the main analysis, it is useful to characterize banks' choice of monitoring effort taking as given the amount of capital, k, the pricing of the loans, r_L, as well as the cost of deposits and capital, r_D and r_E, respectively. Each bank chooses a monitoring effort so as to maximize expected profits. Since $r_L - (1-k)r_D$ is the bank's revenue if the loan is repaid, but it gets nothing if the loan defaults, these expected profits can be expressed as

$$\max_q \Pi = q(r_L - (1-k)r_D) - kr_E - cq^2$$

The solution to this problem yields

$$q^* = \min\left\{\frac{r_L - (1-k)r_D}{2c}, 1\right\}$$

as the optimal level of monitoring for each bank. Note that, for values such that $q < 1$, bank monitoring effort is increasing in the return from lending (r_L) as well as in the level of capital (k) the bank holds, but is decreasing in the deposit rate (r_D), and in c, a measure of the marginal cost of monitoring.

We note that this framework implies a moral hazard problem in the choice of monitoring when banks raise a positive amount of deposits. Since banks repay depositors only when their portfolios succeed, they do not internalize the full cost of default on depositors. This limited liability biases bank monitoring downwards. Capital, however, forces the bank to bear some of the burden associated with a non-performing loan, and therefore provides an incentive for the bank to monitor. Thus, a possible rationale for regulation is to limit moral

sets k and r_L simultaneously.

hazard and raise the level of monitoring. This is illustrated by noting that, in the absence of limited liability, the equilibrium level of monitoring would be $\widehat{q} = \min\left\{\frac{r_L}{2c}, 1\right\} \geq q^*$, with the inequality strict whenever $q^* < 1$.

4 Excess demand for credit

We begin with the case where there is an overall shortage of loanable funds relative to the demand for credit, which implies that banks will be able to obtain their preferred terms. This case reflects a situation where there are fewer banks than investment projects ($N < M$), so that borrowers compete away the return on their projects in order to attract funding.

Banks set k and r_L so as to maximize expected profits, taking into account their subsequent monitoring choice and the fact that borrowers accept the loans only if they have a non-negative surplus. Thus, the profit-maximizing contract solves the following problem:

$$\max_{k,r_L} \Pi = q(r_L - (1-k)r_D) - kr_E - cq^2$$

subject to

$$\begin{aligned} q &= \min\left\{\frac{r_L - (1-k)r_D}{2c}, 1\right\}; \\ CS &= q\left(R - r_L\right) \geq 0; \\ 0 &\leq k \leq 1. \end{aligned}$$

The first constraint represents the monitoring effort that banks choose to maximize expected profits after lending to borrowers, which was obtained above. The second constraint is the participation constraint of borrowers, labelled as consumer surplus (CS), and states that borrowers are willing to accept loans only if they can earn a non-negative expected return. The last constraint is simply a physical constraint on the level of capital, in that banks can choose between raising only deposits, a mixture of deposits and capital, or being entirely equity financed.

The solution to this maximization yields the following result.

Proposition 1 *When borrowers compete for funds in a competitive market, the profit-maximizing contract allows banks to not hold any capital ($k = 0$), and set the loan rate equal to the maximum return on the project ($r_L = R$). Banks exert a level of monitoring effort $q = \min\left\{\frac{R-r_D}{2c}, 1\right\}$ and earn positive expected profits ($\Pi > 0$).*

Proof: Substituting q in bank profits, we obtain

$$\Pi = \frac{(r_L - (1-k)r_D)^2}{4c} - kr_E.$$

Differentiating it with respect to k gives

$$\begin{aligned} \frac{d\Pi}{dk} &= \frac{(r_L - (1-k)r_D)r_D}{2c} - r_E \\ &= qr_D - r_E \leq 0, \text{ as } q \leq 1 \text{ and } r_E \geq r_D. \end{aligned}$$

This implies that $k = 0$. Furthermore, for $r_L \geq r_D$,

$$\frac{d\Pi}{dr_L} = \frac{(r_L - (1-k)r_D)}{2c} \geq 0 \quad \forall\, r_L \leq R,$$

implying that bank profits are always increasing in the interest rate r_L and are therefore maximized at $r_L = R$. $\square$

The intuition behind Proposition 1 is simple. When there is an excess supply of profitable lending opportunities, banks will retain all the surplus from investment projects as borrowers compete away their own returns in order to attract funds. Since equity is more costly to banks, they choose to finance themselves entirely with deposits. Banks benefit from a high loan rate in two ways. First, a high loan rate provides a large return to the bank, all things equal. Second, a high loan rate also provides greater incentives for the bank to monitor.

Loan rates and capital are indeed two alternative ways to provide banks with monitoring incentives, but they differ in their impact for the bank. Raising capital entails a direct cost only for banks, whereas increasing loan rates has a negative impact only for borrowers, which the bank does not internalize.

Given banks' desire to minimize their holdings of capital, there may be a scope for capital regulation in this context. Due to limited liability, banks do not internalize the full cost of default, and therefore choose their level of capital and loan prices so as to maximize their expected profits. A regulator, by contrast, would likely be interested in maximizing social welfare, which includes the cost borne by the deposit insurance fund, or by depositors if they ultimately bear the consequences of default. The regulator would therefore solve the following problem:

$$\begin{aligned}\max_k \; SW &= \Pi + CS - (1-q)(1-k)r_D \\ &= qR - (1-k)r_D - kr_E - (1-q)(1-k)r_D\end{aligned}$$

subject to

$$\begin{aligned} q &= \min\left\{\frac{r_L - (1-k)r_D}{2c}, 1\right\}; \\ r_L &= \arg\max_r \Pi(r); \\ CS &= q\left(R - r_L\right) \geq 0; \\ 0 &\leq k \leq 1. \end{aligned}$$

The optimization problem is similar to before, with the important difference that the regulator chooses only the level of capital, and that it does so in order to maximize social welfare. The loan rate is still set by the market and will be equal to the maximum return of the project, as established in Proposition 1.[2]

Proposition 2 *When borrowers compete for funds, capital regulation that maximizes social welfare requires that banks hold capital equal to* $k = 1 - \frac{2c}{r_D^2}(r_E - r_D)$, *which is positive as long as* $r_D > \max\left\{R - 2c, \sqrt{c(c + 2r_E)} - c\right\}$.

Proof: Substituting $r_L = R$ and $k = 0$ in the expression for q gives

$$q = \min\left\{\frac{R - r_D}{2c}, 1\right\}$$

Thus, $q = 1$ if $R - r_D \geq 2c$, and $q < 1$ if $R - r_D < 2c$.

Substituting $r_L = R$ and keeping $k > 0$, social welfare becomes

$$SW = \frac{(R - (1-k)r_D)^2}{4c} - kr_E - [1 - \frac{(R - (1-k)r_D)}{2c}](1-k)r_D.$$

Differentiating SW with respect to k, we have

$$\begin{aligned} \frac{dSW}{dk} &= \frac{(R - (1-k)r_D)r_D}{2c} - r_E - [-\frac{(1-k)r_D^2}{2c} - r_D + \frac{(R - (1-k)r_D)r_D}{2c}] = 0 \\ &= \frac{(1-k)r_D^2}{2c} + r_D - r_E = 0. \end{aligned}$$

Calculating this expression at the two extreme levels of capital gives

$$\left.\frac{dSW}{dk}\right|_{k=1} = r_D - r_E \leq 0,$$

and

$$\left.\frac{dSW}{dk}\right|_{k=0} = \frac{r_D^2}{2c} + r_D - r_E \gtreqless 0,$$

[2] The loan price in this model represents a pure transfer between banks and borrowers, so that setting a high price is non-distortionary. We believe that allowing for an incentive effect for the borrower as well, so that requiring that some rent be left to the borrower, would not qualitatively affect our findings.

implying that the welfare-maximizing level of capital is $k^* \in (0,1)$ if $r_D > \sqrt{c(c+2r_E)} - c$, and is given by

$$k^* = 1 - \frac{2c}{r_D^2}(r_E - r_D) < 1,$$

thus establishing the proposition. □

Proposition 2 implies that capital regulation requires a positive level of capital only when an increase in k has a positive effect on banks' incentive problem, thus reducing the cost to the deposit insurance fund. This occurs when the required return for depositors r_D is sufficiently high that banks would not monitor fully when they have no capital (i.e., when $r_D > R - 2c$), and also high enough that the positive incentive effect on social welfare of raising capital outweighs the cost r_E (i.e., when $r_D > \sqrt{c(c+2r_E)} - c$).

Comparing Propositions 1 and 2 leads to the following immediate result.

Proposition 3 *When borrowers compete for funds, capital regulation requires banks to hold a higher amount of capital than the market if* $r_D > \max\left\{R - 2c, \sqrt{c(c+2r_E)} - c\right\}$.

This result establishes that a regulator will often require a higher amount of capital than what maximizes banks' profits, and never a lower amount. Regulation can thus be beneficial as it increases social welfare relative to what would be obtained under the market solution. In these instances, there is a rationale for capital regulation as a way of providing banks with incentives to monitor.

5 Excess supply of funds

We now turn to the case when there is a shortage of good lending opportunities for banks relative to the funds the banking system has available to lend. In this case, banks will have to set contract terms competitively in order to attract borrowers, who will generally be able to appropriate most, if not all, of the surplus associated with their projects. In contrast to the previous section, this case reflects the situation when there are fewer investment projects than banks ($M < N$).

The contract that maximizes consumer (e.g., borrower) surplus solves the following problem:

$$\max_{k,r_L} CS = q\,(R - r_L)$$

subject to

$$\begin{aligned} q &= \min\left\{\frac{r_L - (1-k)r_D}{2c}, 1\right\}; \\ \Pi &= q(r_L - (1-k)r_D) - kr_E - cq^2 \geq 0; \\ 0 &\leq k \leq 1; \\ 1 &\leq r_L \leq R; \end{aligned}$$

where, as before, Π represents bank profits, q is the monitoring effort that each bank chooses as a function of r_L and k, and CS represents consumer surplus. Note that, relative to the previous section, we impose a participation constraint for banks, in that they must earn non-negative profits, and a constraint on the loan rate which cannot be higher than the return from the project. We can now state the following result.

Proposition 4 *The contract that maximizes borrower surplus in a competitive market is characterized as follows:*

1) For $R \geq 4c$, then monitoring is $q = 1$. The loan rate is $r_L = (1 - k^{CS})r_D + 2c$, and banks are required to hold capital k^{CS} equal to $k^{CS} = \min\left\{\frac{c}{r_E}, 1\right\}$. For $k^{CS} = 1$ (i.e., if $c > r_E$), banks earn profits $\Pi = c - r_E > 0$, otherwise $\Pi = 0$.

2) For $R < 4c$, then monitoring is $q = \frac{R-(1-k^{CS})r_D}{4c} < 1$. The loan rate is $r_L = \frac{R+(1-k^{CS})r_D}{2}$, and banks hold capital equal to $k^{CS} = \min\left\{\frac{8cr_E - Rr_D + r_D^2 - 4\sqrt{r_E c\left(4cr_E - Rr_D + r_D^2\right)}}{r_D^2}, 1\right\}$, which is less than 1 for $c > \frac{R^2}{16r_E}$ and equal to one otherwise. For $k^{CS} = 1$, $\Pi = \frac{R^2}{16c} - r_E > 0$, *and* $\Pi = 0$ *for* $k^{CS} < 1$.

Proof: Start by noting that, since $q = \min\left\{\frac{r_L-(1-k)r_D}{2c}, 1\right\}$, if $r_L < (1-k)r_D + 2c$ then $q = \frac{r_L-(1-k)r_D}{2c} < 1$. Since $CS = q(R-r_L)$, we have that $\frac{\partial CS}{\partial k} = \frac{\partial q}{\partial k}(R-r_L) = \frac{r_D}{2c}(R-r_L) > 0$ for $q < 1$. Therefore, more capital increases consumer surplus.

We proceed in two stages, starting by maximizing CS with respect to the loan's price, r_L, for a fixed k, which yields

$$\frac{\partial CS}{\partial r_L} = \frac{\partial q}{\partial r_L}(R-r_L) - q = \frac{R - 2r_L + (1-k)r_D}{2c} = 0$$

Solving the FOC yields $r_L = \frac{R+(1-k)r_D}{2}$.

We can now maximize CS with respect to the choice of capital, k. However, we know from above that the combination of $r_L = \frac{R+(1-k)r_D}{2}$ and the highest possible k will be optimal for borrowers. We therefore introduce the participation constraint for the bank, that $\Pi = q(r_L - (1-k)r_D) - kr_E - cq^2 \geq 0$. Substituting for $q = \frac{r_L-(1-k)r_D}{2c}$ as well as for r_L, we obtain

$$\Pi = \frac{(R-(1-k)r_D)^2}{16c} - kr_E \geq 0, \quad k \leq 1$$

We can solve this for the value of k that satisfies the constraint with equality ($\Pi = 0$). Since Π is strictly convex in k, $0 \leq k \leq 1$, and consumer surplus is increasing in k, the relevant solution must be either the smaller root or a corner solution at $k = 1$. The solution is then

$$k^{CS} = \min\left\{\frac{8cr_E - Rr_D + r_D^2 - 4\sqrt{r_E c\left(4cr_E - Rr_D + r_D^2\right)}}{r_D^2}, 1\right\}$$

Note that if $k^{CS} = 1$, then $r_L = \frac{R}{2}$.

We now check when in fact $q < 1$. From the definition of the optimal level of monitoring $q = \min\left\{\frac{r_L-(1-k)r_D}{2c}, 1\right\}$, we see that, for $r_L \geq (1-k)r_D + 2c$, $q = 1$. Substituting in the optimal value for r_L gives the following condition:

$$\frac{R+(1-k)r_D}{2} \geq (1-k)r_D + 2c$$

The right hand side is maximized at $k = 0$, so a sufficient condition for $q = 1$ is that $R - r_D - 4c \geq 0$. In this case, there is no benefit in terms of greater monitoring to having a higher interest rate on the loan, and so borrowers should just require the lowest possible interest rate consistent with $q = 1$, which is satisfied by $r_L = (1-k)r_D + 2c$. If we again substitute this value of r_L into the expression for bank profits we obtain

$$\Pi = \big(r_L - (1-k^{CS})r_D\big) - k^{CS}r_E - c = c - k^{CS}r_E,$$

which, after setting equal to zero, yields $k^{CS} = \frac{c}{r_E}$ as long as $c < r_E$. In this case, we have $r_L = (1-\frac{c}{r_E})r_D + 2c = (\frac{r_E - c}{r_E})r_D + 2c$.

Otherwise, for $c > r_E$, $k^{CS} = 1$, which implies that $r_L = 2c$. Moreover, substituting this value of r_L into $q = \min\left\{\frac{r_L - (1-k^{CS})r_D}{2c}, 1\right\}$ and observing that $k^{CS} = 1$, we obtain that $\Pi > 0$ and $q = \min\left\{\frac{R}{4c}, 1\right\} = 1$ for $R \geq 4c$, and is less than 1 otherwise. All together, this implies that $q \geq 1$ for $R \geq 4c$, and $q < 1$ for $R < 4c$. $\square$

The results in Proposition 4 highlight the incentive mechanisms for bank monitoring provided by a competitive credit market. There are two ways of providing banks with incentives to monitor: by requiring that they hold a minimum amount of capital k^{CS}, and by setting the rate r_L on the loan so as to compensate them when the project is successful and the loan is repaid. Both of these two variables increase bank monitoring, but differ in terms of their costs and their effects on borrower surplus and bank profits. Borrowers would like banks to be fully equity financed and exert a high level of monitoring, as their returns increase with q and they do not internalize the cost of capital and of monitoring. By contrast, since capital is a costly input for them (i.e., $r_E \geq r_D$), banks would prefer not to use it and rather receive incentives through a higher loan rate, r_L. However, while increasing r_L is good for incentive purposes, its direct effect is to reduce the surplus to the borrower. Raising r_L enough will therefore eventually reduce borrower surplus, and this occurs when the positive incentive effect of a higher r_L on bank monitoring, q, is dominated by the negative direct effect on borrower surplus, $R - r_L$. Thus, when borrowers obtain the surplus, banks have to raise a positive amount of capital to have sufficient monitoring incentives and attract borrowers by offering them the maximum possible surplus.

The exact amounts of monitoring and capital in equilibrium depend on the return of investment projects R, the cost of capital r_E and of monitoring c. When projects are very profitable ($R \geq 4c$) and capital is not too costly ($c > r_E$), banks exert the maximum effort, $q = 1$, and raise the highest level of capital, $k^{CS} = 1$. Borrowers want banks to monitor fully as projects are very profitable, and can induce banks to do so by raising only capital, as long as this is not too costly and banks' profits are positive. When projects are less profitable ($R < 4c$), banks still raise the highest level of capital, but no longer monitor fully as this is too costly. For all the other cases, market incentives will lead banks to choose a lower level of capital ($k^{CS} < 1$), less monitoring ($q < 1$), or both. The participation constraint of banks prevents banks from raising the highest level of capital, thus leading to a lower level of monitoring when projects are not very profitable.

Interestingly, borrowers may be willing to give up some of the return on the loan to the bank in order to provide the bank with incentives to monitor. They accomplish this by allowing the loan rate, r_L, to reflect the returns of the project, and to be increasing in such returns as long as there are incentive effects from doing so (as long as $q < 1$): $\frac{\partial r_L}{\partial R} > 0$. In other words, the loan's price need not be set only to compensate the bank for the credit risk associated with granting the loan, but also to induce the bank to exert effort in monitoring the project and thus improve the expected return of the loan. Furthermore, since capital and the loan rate are alternative instruments for providing banks with an incentive to monitor, we note that the equilibrium value of r_L is decreasing in the level of capital k. This implies that these are substitute instruments from the point of view of borrowers, who only trade off their relative costs from the perspective of reducing consumer surplus.

The complement to Proposition 2 is to analyze the optimal choice of capital from a social welfare perspective when we assume that loan contracts are determined as part of a market solution. In other words, a regulator would solve the following problem.

$$\begin{aligned}\max_k SW &= \Pi + CS - (1-q)(1-k)r_D \\ &= qR - (1-k)r_D - kr_E - (1-q)(1-k)r_D\end{aligned}$$

subject to

$$\begin{aligned} q &= \min\left\{\frac{r_L - (1-k)r_D}{2c}, 1\right\}; \\ r_L &= \arg\max_r CS = q(R-r) \\ 0 &\leq k \leq 1. \end{aligned}$$

Note that we here explicitly assume that the loan rate is set competitively so as to maximize the return to borrowers, but that the regulator can set the level of capital banks are required to hold.

Proposition 5 *When banks compete for projects, capital regulation that maximizes social welfare requires that banks exert an amount of monitoring and hold capital k^{reg} equal to:*

1) For $R > 2c\frac{2r_E - r_D}{r_D}$, then $q = \min\left\{\frac{R}{4c}, 1\right\}$, and $k^{reg} = \min\left\{\frac{4c + r_D - R}{r_D}, 1\right\}$, which is less than 1 for $R > 4c$ and equal to 1 otherwise.

2) For $R < 2c\frac{2r_E - r_D}{r_D}$, then $q = \frac{R - (1-k^{reg})r_D}{4c} < 1$ and $k^{reg} = \min\left\{\frac{Rr_D + r_D^2 - 8c(r_E - r_D)}{r_D^2}, 1\right\}$, which is less than 1 for $R < 8c\frac{(r_E - r_D)}{r_D}$ and equal to 1 otherwise.

Proof: Start by maximizing social welfare with respect to k, assuming that the loan rate is set to maximize CS, i.e., that $r_L = \frac{R+(1-k)r_D}{2}$. Social welfare is given by

$$\begin{aligned} \max_k SW &= \Pi + CS - (1-q)(1-k)r_D \\ &= qR - (1-k)r_D - kr_E - cq^2 \end{aligned}$$

We can now take the first order condition to get

$$\frac{\partial SW}{\partial k} = \frac{\partial q}{\partial k}(R - 2cq) - r_E + r_D$$

We know that, for $q < 1$, the solution for q is $q = \frac{r_L-(1-k)r_D}{2c}$. Substituting in the value of r_L above we get $q = \frac{R-(1-k)r_D}{4c}$. We therefore have that

$$\frac{\partial SW}{\partial k} = \frac{r_D}{4c}\left(\frac{R+(1-k)r_D}{2}\right) + r_D - r_E = 0$$

The solution we obtain is

$$k^{reg} = \min\left\{\frac{Rr_D + r_D^2 - 8c\,(r_E - r_D)}{r_D^2}, 1\right\}$$

From this expression, we obtain that $k^{reg} < 1$ for $c > \frac{Rr_D}{8(r_E-r_D)} \Leftrightarrow R < 8c\frac{(r_E-r_D)}{r_D}$. Otherwise, for $R > 8c\frac{(r_E-r_D)}{r_D}$, we have that $k^{reg} = 1$.

This is assuming that $q < 1$. To get the bounds on when $q = 1$, substitute the solution for k^{reg}, assuming $k^{reg} < 1$, into

$$q = \frac{R-(1-k^{reg})r_D}{4c} = \frac{Rr_D - 4c\,(r_E - r_D)}{2cr_D}$$

From here, we see that for $c > \frac{Rr_D}{4r_E-2r_D} \Leftrightarrow R < 2c\frac{2r_E-r_D}{r_D}$, $q < 1$. Otherwise, for $R > 2c\frac{2r_E-r_D}{r_D}$, $q = 1$ and k should be set such that $q(k) = 1 \Leftrightarrow k^{reg} = \frac{4c+r_D-R}{r_D}$. Note, however, that for $R < 4c$ this solution would imply that $k^{reg} > 1$, which is not feasible. Therefore, for

$R < 4c$, we obtain that $k^{reg} = 1$, which implies that $q = \frac{R}{4c} < 1$.

One final point that needs to be verified is that, for $R < 2c\frac{2r_E - r_D}{r_D}$, then $q = \frac{R-(1-k)r_D}{4c} < 1$, but that for $R > 8c\frac{(r_E - r_D)}{r_D}$, we have that $k^{reg} = 1$, which would imply that $q = \frac{R}{4c}$. Note, however, that for both of these conditions to be true at the same time requires that $8c\frac{(r_E-r_D)}{r_D} < 2c\frac{2r_E-r_D}{r_D}$. This will be satisfied if and only if $4(r_E - r_D) < 2r_E - r_D \Leftrightarrow r_E < \frac{3}{2}r_D$. We can now use this in the necessary condition for $q < 1$, which is $R < 2c\frac{2r_E-r_D}{r_D}$. Given the restriction on r_E and r_D, the right hand side must be less than $2c\frac{2\left(\frac{3}{2}r_D\right)-r_D}{r_D} = 4c$. Therefore, the joint assumption that $R < 2c\frac{2r_E-r_D}{r_D}$ and $R > 8c\frac{(r_E-r_D)}{r_D}$ implies that $R < 4c$, and consequently that $q = \frac{R}{4c} < 1$, as desired. □

While the terms of the loan contract (i.e., the loan rate, r_L) are set in a competitive market setting and not subject to regulatory interference, a regulator may want to impose a capital requirement for banks in order to ensure they have a sufficient incentive to monitor. In contrast to Proposition 2, now the regulator requires in general that banks hold a positive amount of capital. The reason is that the market sets a lower loan rate when the borrowers obtain the surplus than when the banks obtain it and, therefore, the regulator has to use more capital to provide banks with incentives to monitor. Interestingly, however, optimal regulation does not necessarily call for "narrow banking" in the sense of having a fully capitalized intermediary, but rather allows for a mix between capital and deposit-based financing for the bank. This will generally be true when the cost of capital relative to deposits, $r_E - r_D$, is high, or when the aggregate return from encouraging greater monitoring, R, is relatively low.

We now turn to one of the main results in the paper, which is whether a pure market-based system is likely to provide sufficient incentives for bank discipline and monitoring, and whether capital regulation can be an effective tool for providing such incentives. For that, we have the following:

Proposition 6 *For all R, r_E, and c, there exists a value $\widetilde{r}_D(R, r_E, c) > 0$ such that $k^{reg} <$*

k^{CS} for $r_D < \widetilde{r}_D$, and $k^{reg} \geq k^{CS}$ for $r_D \geq \widetilde{r}_D$.

Proof: We begin with the case of parameter values such that $q, k < 1$, and show that there exists a value $\widetilde{r}_D > 0$ such that $k^{reg} < k^{CS}$ if and only if $r_D < \widetilde{r}_D$. Consider the solution that maximizes borrower (consumer) surplus, k^{CS}, and assume that $R < 4c$ and $c > \frac{R^2}{16r_E}$, which implies that $q, k^{CS} < 1$. From the condition defining k^{CS},

$$\Pi = \frac{\left(R - (1 - k^{CS})r_D\right)^2}{16c} - k^{CS} r_E = 0$$

one can clearly see that, as $r_D \to 0$, $k^{CS} \to \frac{R^2}{16cr_E} < 1$ for $c > \frac{R^2}{16r_E}$.

By contrast, k^{reg} is defined by

$$\frac{\partial SW}{\partial k} = \frac{r_D}{4c}\left(\frac{R + (1 - k^{reg})r_D}{2}\right) + r_D - r_E = 0$$

For $r_D \to 0$, $k^{reg} \to 0$ as well, since it is optimal to just have deposit-based finance. These two results together imply that there is some threshold $\underline{r}_D$ such that, for $r_D < \underline{r}_D$, $k^{reg} < k^{CS}$.

At the other extreme, we consider the solutions as $r_D \to r_E$. For $c > \frac{R^2}{16r_E}$, $k^{CS} = \frac{8c - R + r_E - 4\sqrt{(4c - R + r_E)c}}{r_E} < 1$. By comparison, $k^{reg} \to 1$ as $r_D \to r_E$ for all parameter values. Therefore, we can also conclude that there must exist some threshold $\overline{r}_D$ such that, for $r_D > \overline{r}_D$, $k^{reg} > k^{CS}$.

Comparing the two values of k, $k^{reg} < k^{CS}$ if and only if

$$k^{reg} = \frac{Rr_D + r_D^2 - 8c\left(r_E - r_D\right)}{r_D^2} < \frac{8cr_E - Rr_D + r_D^2 - 4\sqrt{r_E c\left(4cr_E - Rr_D + r_D^2\right)}}{r_D^2} = k^{CS}$$

Rearranging, we obtain the condition for $k^{reg} - k^{CS}$:

$$\frac{2}{r_D^2}\left(Rr_D + 4cr_D - 8cr_E + 2\sqrt{cr_E\left(4cr_E - Rr_D + r_D^2\right)}\right) < 0$$

Since we know that for low values of r_D this condition will be satisfied, but not for higher values, we can establish that there is a unique threshold where the inequality flips (i.e., that $\underline{r}_D = \overline{r}_D$) if the difference $k^{reg} - k^{CS}$ is either concave or convex in r_D. For this, we only need the second derivative of the term inside the parenthesis, which yields

$$\begin{aligned} & \frac{\partial^2}{\partial r_D^2}\left(Rr_D + 4cr_D - 8cr_E + 2\sqrt{cr_E\left(4cr_E - Rr_D + r_D^2\right)}\right) \\ = \; & \frac{\frac{1}{2}\left(16cr_E - R^2\right)\left(\sqrt{cr_E}\right)}{\left(4cr_E - Rr_D + r_D^2\right)\left(\sqrt{4cr_E - Rr_D + r_D^2}\right)} > 0, \end{aligned}$$

since by assumption $c > \frac{R^2}{16r_E}$.

The finding that the function $k^{reg} - k^{CS}$ is convex implies that $k^{reg} - k^{CS}$ can at most cross zero twice, the first time from above and the second from below. However, two crossings are inconsistent with the finding in the proposition above that for low values of r_D, $k^{reg} - k^{CS} < 0$, while for high values of r_D, $k^{reg} - k^{CS} > 0$. Therefore, $k^{reg} - k^{CS} = 0$ at one unique point, which implies that $\underline{r}_D = \overline{r}_D = \widetilde{r}_D$, and we have just one threshold, as desired.

We next proceed to the case where $q = 1$ in both cases, which is true for sufficiently large R, but that $k < 1$. Start with the case of consumer surplus maximization, where, for $R > 4c$, $q = 1$ and $k^{CS} = \frac{c}{r_E}$. For the case with regulation, we have that for $R > \max\{2c\frac{2r_E - r_D}{r_D}, 4c\}$, $q = 1$ and $k^{reg} = \frac{4c + r_D - R}{r_D}$. Therefore, $k^{reg} < k^{CS} \Leftrightarrow$

$$\frac{4c + r_D - R}{r_D} < \frac{c}{r_E}$$

This last inequality can be solved for r_D to yield the condition

$$\widetilde{r}_D < r_E\left(\frac{R - 4c}{r_E - c}\right),$$

which establishes that $k^{reg} - k^{CS} < 0$ if and only if $r_D < \widetilde{r}_D$, as desired.

The last case is a possible "mixed" case, in which monitoring may be at a maximum for one solution but not the other. It is straightforward to show that the only case of relevance is where, with a slight abuse of notation, $q^{CS} = 1$ but $q^{SW} < 1$. This occurs for $\frac{Rr_D}{2(2r_E - rD)} < c < \frac{R}{4}$, and in this range $k^{CS} = \frac{c}{r_E}$ and $k^{reg} = \frac{Rr_D + r_D^2 - 8c(r_E - r_D)}{r_D^2}$. The difference $k^{reg} - k^{CS}$ simplifies to

$$(r_E - c)r_D^2 + (R + 8c)r_E r_D - 8cr_E^2 = 0$$

The relevant solution is

$$\widetilde{r}_D = \frac{r_E(8c + R - \sqrt{32c^2 + R^2 + 16c(R + 2r_E)})}{2(c - r_E)},$$

which implies $k^{reg} > k^{CS}$ only if $r_D > \widetilde{r}_D$. □

While the exact expressions for $\widetilde{r}_D$ are provided in the proof, the interpretation of this result can be stated quite generally for all parameter values as follows. When capital is much more expensive than deposits ($r_D << r_E$), it is socially optimal to economize on capital and instead rely more heavily on deposits for financing the bank. The market solution, by contrast, will demand that banks hold an excessive level of capital when there is a shortage of good lending opportunities and credit markets are competitive. At the other extreme, when bank capital is not significantly more costly than deposits, regulators desire that banks hold more capital than what the market requires. The reason is that the regulator internalizes the cost of raising capital and of the deposit insurance fund. Thus, when the difference $r_E - r_D$ is high, the regulator chooses not to impose a higher capital requirement because this is socially more costly than repaying the depositors in case of bank default. By contrast, when the difference $r_E - r_D$ is low, the regulator prefers to require a high level of capital and reduce the costs of bank default. Since the market does not take the cost of capital and of bank default into account, it requires that banks hold the amount of capital which maximizes consumer surplus without any consideration for the social costs this may impose on banks or the deposit insurance fund.

The result of Proposition 6 also suggests that, when capital is relatively expensive, capital regulation is not likely to be binding as market incentives will induce banks to hold greater amounts of capital than what is socially optimal. In other words, a minimum capital requirement imposed by a regulator, as well as changes in that requirement, would have no effect on banks' aggregate holdings of capital.[3] In fact, social welfare maximization would call for a ceiling being placed on the level of capital, or a tax on its use so as to discourage banks from holding excessive capital.

The contrast between the finding in Proposition 6 and that in Proposition 3 is clear. When there is a shortage of bank funds and banks are able to appropriate most of the surplus from lending, capital regulation plays a clear role in increasing bank monitoring and reducing the probability of failure. When banks compete for projects, the market may increase bank monitoring even more by requiring that they hold a higher amount of capital. The effectiveness of capital regulation, therefore, clearly depends on the structure of the market for bank credit. When borrowers compete for funds, establishing a capital adequacy requirement can be a useful way of imposing bank discipline, reducing the burden to the insurance fund and raising social welfare. By contrast, when banks compete for projects, the incentives provided in the market as banks compete to attract borrowers may lead banks to hold excessive amounts of capital, so that capital adequacy requirements become ineffective and unnecessary.

6 Concluding remarks

The standard view of capital regulation is that it offsets the risk-taking incentives provided by deposit insurance. If equity capital is more costly than other forms of finance as is generally believed, the implication is that capital regulations should be binding. However, in many cases such as the U.S. in the 1990's they appear not to be binding. In this paper we have developed an alternative view of capital which is consistent with the observation that capital constraints may or may not be binding. In particular, when there is an excess supply

[3] This is consistent with the findings of Ashcraft (2001), who finds little evidence that tougher capital requirements were responsible for the increase in capital ratios throughout the 1980's.

of funds relative to the number of attractive projects available so that banks compete for projects, the level of capital determined by the market can be high.

We have considered the case where the existence of deposit insurance is taken as given. It is also important to consider the case where there is no deposit insurance. We conjecture that our main results would be unaffected by this change. The market equilibrium can still involve a positive level of capital. In addition the optimal amount of capital from a social welfare point of view can be above or below the market level.

References

[1] Ashcraft A.B., 2001, "Do Tougher Bank Capital Requirements Matter? New Evidence from the Eighties", mimeo, Federal Reserve Bank of New York.

[2] Berger A.N., R.J. Herring and G.P. Szego, 1995, "The Role of Capital in Financial Institutions", *Journal of Banking and Finance*, 19, 393-430.

[3] Besanko, D. and G. Kanatas, 1993, "Credit Market Equilibrium with Bank Monitoring and Moral Hazard", *Review of Financial Studies*, 6, 213-232.

[4] Carletti, E., 2004, "The Structure of Relationship Lending, Endogenous Monitoring and Loan Rates", *Journal of Financial Intermediation*, 13, 58-86.

[5] Dell'Ariccia, G. and R. Marquez, 2005, "Competition Among Regulators and Credit Market Competition", forthcoming, *Journal of Financial Economics.*

[6] Flannery, M.J. and K.P. Rangan, 2002, "Market Forces at Work in the Banking Industry: Evidence from the Capital Buildup from the 1990s", mimeo, University of Florida.

[7] Hellmann, T., Murdock, K., and J. Stiglitz, 2000, "Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough?", *American Economic Review*, 90, 147-165.

[8] Holmstrom, B. and J. Tirole, 1997, "Financial Intermediation, Loanable Funds, and the Real Sector", *Quarterly Journal of Economics*, 112, 663-691.

[9] Repullo, R., 2004, "Capital Requirements, Market Power, and Risk-Taking in Banking", *Journal of Financial Intermediation*, 13, 156-182.