WP 36

Optimal Structure of the Financial Intermediation Industry

Sonja Daltung*
Sveriges Riksbank

This version: February 10, 1997

Abstract

The optimal structure of the financial intermediation industry in a finite economy, where the role of intermediaries is to reduce information costs in the lending of resources from investors to entrepreneurs, is characterized. Banks are owned by entrepreneurs in order to exploit informational gains from close relationships between banks and firms. It is shown that this form of ownership is a reason for allowing more than one bank to operate, although there are economies of scale in intermediation. It is also argued that free entry into intermediation generally would not provide the optimal number of banks.

*This paper was written while I was a research fellow at IIES. I thank IIES for its hospitality. I am very grateful to Nils Gottfries and Henrik Horn for very helpful suggestions and many stimulating discussions. I have also benefited from discussions with Lars Hörngren, Anthony Santomero, and Hans Wijkander. Helpful comments on earlier drafts were provided by Thomas Björk, Mathias Dewatripont, Per Krusell, Harald Lang, Tomas Sjöström, the participants of the Bertil Danielson Symposium on "Banking in Sweden", seminar participants at IIES, and the participants of the Nordic Workshop on "Asymmetric Information and Financial Markets". Financial support from Bankforskningsinstitutet is gratefully acknowledged. The views in this paper are those of the author and do not necessarily reflect those of Sveriges Riksbank.

Contents

1 Introduction

Most governments regulate the financial sector and have pronounced ideas concerning the structure of a sound financial sector. Entry is regulated by charter proceedings, the activities of financial firms are regulated by legal rules of operation, and interest rate regulations are not uncommon. Economic theory, on the other hand, gives hardly any guidance on the subject of the optimal structure of the financial sector. This deficiency has become more apparent as governments being under pressure to change regulations in financial markets and is not least a problem for the former socialist economies, which are in the phase of building up private financial systems and introducing financial regulation.

The main reason for the optimal structure of financial systems to be a relatively unexplored research area has been the lack of theory explaining the functions of banks and other financial intermediaries. However, the development of informational economics made it possible to rigorously analyze the *raison d' être* of financial intermediaries, and since the beginning of the 80's a great deal of effort has gone into the study of financial intermediaries as optimal institutional responses to information problems in financial markets.[1] The purpose of this paper is to take these theoretical developments as a starting point in characterizing the socially optimal structure of the financial intermediation industry.

A recurring theme in the research on intermediation and banking is that diversification of the portfolio of the intermediary firm is crucial for its ability to mitigate information problems, and that there are economies of scale in intermediation. This suggests that financial intermediation is a natural monopoly, and that the optimal structure is a regulated monopoly. However, even though authorities in many countries do limit the number of banks, in all developed countries, they permit more than one bank to operate. Nor does the equilibrium number of banks appear to be one.

In this paper, I argue that one reason for it to be optimal to give charter to more than one bank is the informational value of close relationships between banks and firms, a value which has not been recognized by the previous literature on endogenous intermediation.[2] As will be shown, a single bank is not able to realize the full value of such relations.

Close relationships between banks and firms are theoretically motivated by the existence of informational asymmetries. Costs of asymmetric information arise when contracting parties have different interests, and the establishment of close relationships between the parties reduce information costs by bringing interests more in line. Close relationships between banks and firms are also an important feature of actual banking. This is most apparent in countries such as Germany, which allow universal banking,[3] but there are close relations between banks and firms also in other countries. The most widely known example is probably the *keiretsu* system in Japan.[4] There is also

[1] See Bhattacharya and Thakor (1993) and Gertler (1988) for surveys of this research.

[2] Other explanations are aggregate risk (Krasa and Villamil, 1992a) and bank capitalization diminishing with bank size (Winton, 1995).

[3] Universal banks may offer the entire range of financial services, and they may own shares in non-financial firms (Benston, 1994). See for instance Cable (1985) for a discussion of the German system.

[4] Also in Sweden there are strong links between banks and industrial firms. One of the largest

empirical evidence supporting the view that close relationships between banks and firms are motivated by information problems and improve the efficiency of banks in their role as credit providers.[5]

The relationship analyzed in this paper is common ownership of firms and banks. I will refer to this as "relationship banking", a term often used for banking systems which build upon close relations between banks and firms. The potential gain of a common ownership is that it brings the interest of the firm in line with the interest of the bank, so that a loan from the bank to this firm does not give rise to any costs of asymmetric information.

The structure of the financial intermediation industry is examined in the context of the model of delegated monitoring developed by Diamond (1984), in a situation where banks are owned by entrepreneurs. The next section describes the economy, which is similar to that analyzed by Williamson (1986), but with the important difference that the total number of agents is finite. In an infinitely large economy there is room for infinitely many intermediaries, each of infinitely large size. In this setting, there is no trade-off between the diversification of banks and the number of banks, as each bank can perfectly diversify independently of the number of banks in the economy. Hence, the industrial structure of the financial intermediation industry is relevant only in a finitely large economy.

There is "costly state verification" as in Townsend (1979): only the entrepreneur freely observes the outcome of his project, while other agents can verify the outcome at some cost, an action referred to as monitoring. Section 3 describes direct lending in an economy without financial intermediaries.

As shown in section 4, in this paper there are two ways in which financial intermediation by a sufficiently large monopoly bank economizes on monitoring costs. Firstly, as in Diamond (1984) and Williamson (1986), delegating the monitoring task to the bank is a way of avoiding duplication of effort when the scale of inputs for the project exceeds the wealth of any single investor. Secondly, since the entrepreneur who owns the bank has no incentive to be dishonest with his own bank, nobody has to monitor the project of the bank's owner. These benefits of delegated monitoring increase with the scale of investments. There are also costs of delegating monitoring to the bank. The bank finances its lending to entrepreneurs by issuing debt contracts to investors. In case the bank cannot honour its debt to investors, each investor has to monitor the

Swedish banks, Skandinaviska Enskilda Banken, is controlled by the Wallenberg family, which also controls one of the largest conglomerates in Sweden. This relation goes right back to the foundation of Stockholms Enskilda Bank, which in 1972 was merged into Skandinaviska Enskilda Banken (Lindgren, 1990). Another Swedish example is Östgöta Enskilda Bank, which is owned almost entirely by the Lundberg family, which also owns a conglomerate firm. See Sundin and Sundqvist (1995) for the owner structures in Sweden in 1995. See for instance Hoshi et. al. (1990, 1991) and Berglöf and Perotti (1993) for discussions of the financial system in Japan.

[5]Mayer (1990) provides evidence suggesting that the financial system of Germany and Japan are more efficient in funding their industries than other systems. Analyzing German companies, Cable (1985) finds that there is a significant, positive relationship between the degree of bank involvement and the financial performance of the company. Investigating a panel data set of Japanese manufacturing firms Hoshi et. al. (1991) find that investment by firms in industrial groups - those with close financial relationships to their banks - is much less sensitive to their liquidity than it is for nongroup firms, and Hoshi et. al. (1990) find that firms in industrial groups invest more and sell more after the onset of distress than nongroup firms.

bank, otherwise its owner would always claim a low outcome and pay a small amount to investors. However, these costs decrease with the size of the bank in terms of the number of loans to entrepreneurs, so that, if there are sufficiently many projects in the economy, financial intermediation will dominate direct lending.

The optimal structure of the financial intermediation industry is characterized in section 5. Because the costs of delegated monitoring decrease with bank size, there are economies of scale in intermediation. Does this imply that there should be only one bank? The analysis shows that it is generally *not* the case. The informational gain from relationship banking can be a reason for letting more than one bank operate.

The optimal number of banks depends on the parameters of the economy, especially the size of the economy. A numerical example is provided in section 6. Naturally, a larger economy has room for more banks than a smaller economy, but it is shown that the optimal number of banks does not increase in proportion to the size of the economy. The bank industry should be safer in a larger economy than in a smaller one.

The industrial structure in a free market is briefly discussed in section 7. The conclusion that could be drawn from this section is that there is no reason to believe that a free market would provide the optimal number of banks; on the one hand, the threat of fierce competition can result in too few banks, while, on the other hand, collusive pricing behaviour can result in too many banks.[6]

2 The economy

The economy lasts two periods and consists of a finite number of risk-neutral agents, who are either investors or entrepreneurs. There exists a single consumption good. Each investor is initially endowed with one unit of the consumption good. Entrepreneurs receive no endowment, but have access to an indivisible investment project which converts k units of goods in the first period into kx units of consumption in the second period. Here k is a given integer with $k \geq 2$, and x is a random variable with positive support on the interval $[0, X]$. Project returns are independently and identically distributed across entrepreneurs according to the probability density function $f(x)$ and probability distribution function $F(x)$.

All consumption takes place in the second period. Investors have access to an investment technology that yields a certain return of y units of the consumption good in the second period. The expected return of the project, $\bar{x}$, is larger than y. Each entrepreneur is able to carry out only one project. Thus, there is a given demand for credit equal to km, where m is the number of entrepreneurs. The number of investors is much larger than km so that the total endowment of the investors is more than sufficient to finance the projects of the entrepreneurs.

The outcome of the project is freely observed only by the entrepreneur, but all agents know the distribution of x. Investors can learn the return on a given project

[6] Another reason for market failure, which is not analyzed here, is coordination problems among agents. Yanelle (1995) argues that due to coordination problems of entrepreneurs in their choice of bank, there may be several active banks in the market equilibrium, while Winton (1993) focuses on the adoption externality among depositors, and show how this lead to multiple equilibria with several active banks. In both cases the market fails to reach the optimal solution, which in their cases is one active bank.

by expending c units of effort, an action referred to as monitoring. The outcome of monitoring is private information. Utility is additive separable in consumption and effort.

3 Direct lending

In order to achieve the necessary scale, the entrepreneur needs finance from k investors. For investors to be willing to fund the project, the entrepreneur must offer them an incentive compatible contract. The optimal contractual arrangement between the entrepreneur and the investor provides the entrepreneur with incentives to be truthful at a minimum of expected monitoring costs. A rigorous analysis of this contracting problem can be found in Williamson (1986), where the optimal symmetric contract is shown to be a so called "standard debt contract".[7]

According to the standard debt contract the entrepreneur either repays his investors a fixed amount (the amount borrowed plus interest), or he defaults, each investor monitors, and the investors split the realized output of the project. This contract is completely characterized by its fixed payment, which I will denote by r and refer to as the loan rate. I will also now and then refer to the debt claim of the investor as a loan. The expected return on a loan of one unit to a project hence is

$$Y^d(r) \equiv r(1 - F(r)) + \int_0^r xf(x)dx - cF(r). \tag{1}$$

The idea of the debt contract is to economize on resources spent on monitoring by investors not verifying all announcements by the entrepreneur. For all announcements that do not trigger monitoring the only implementable payment schedule is a fixed repayment, since, if the stipulated payment differed among these announcements, the entrepreneur would always choose to report the one associated with the lowest repayment. However, the entrepreneur can pay no more than the realized output to his investors. Thus, he will default whenever the outcome is lower than the fixed payment. Investors monitor the entrepreneur when he reports an outcome lower than the promised payment, since otherwise he would always declare bankruptcy and claim a low outcome. In case of bankruptcy, investors recover whatever there is to recover. Through maximum recovery, a given expected return to investors is obtained at lowest possible promised payment, so that the probability of bankruptcy and thereby the expected monitoring costs are minimized.

Due to the monitoring cost, the expected return to investors is not monotonically increasing in r. The first derivative of Y^d with respect to r is

[7] More precisely, the standard debt contract is shown to be optimal given that attention is restricted to non-stochastic monitoring and the precommitment to monitor for a given announcement is taken as binding. Diamond (1984) also derives the standard debt contract, but he supports it with non-pecuniary penalties on the borrower instead of monitoring by the lender. The term "standard debt" was first used by Gale and Hellwig (1985) for the optimal contract between an entrepreneur and an investor. Williamson (1986) extends their analysis to multiple investors assuming that the entrepreneur offers the same contract to all investors. If the entrepreneur is allowed to offer different contracts to investors, the optimal contract will change; see Winton (1995).

$$\frac{dY^d}{dr} = 1 - F(r) - cf(r),$$

which is negative for $r = X$, since $f(r) > 0$ for $r \in [0, X]$. Thus, $Y^d(r)$ reaches a maximum for $r < X$. The monitoring cost is assumed to be small enough for the derivative to be positive for some values of r, so that there is an interior maximum. Moreover, if c is sufficiently small this maximum rate will give the lender an expected return which is higher than the alternative return, y, in which case direct lending is feasible.

Investors are assumed to behave competitively. Since there is a large enough amount of credit available, entrepreneurs face a horizontal supply schedule for credit. Hence, the equilibrium direct lending rate, denoted by $\hat{r}$, is equal to the smallest r for which $Y^d(r) = y$. That is $\hat{r}$ is given by investors individual rationality (IR) constraint:

$$\hat{r}(1 - F(\hat{r})) + \int_0^{\hat{r}} xf(x)dx - cF(\hat{r}) = y. \tag{2}$$

The expected return of an entrepreneur borrowing k units is[8]

$$V(r) \equiv k \int_r^X (x - r)f(x)dx,$$

which can be rewritten as

$$V(r) = k[\bar{x} - r(1 - F(r)) - \int_0^r xf(x)dx]. \tag{3}$$

Substituting the IR condition (2) into the profit function in (3) gives the entrepreneur's equilibrium profit:

$$V(\hat{r}) = k[\bar{x} - y - cF(\hat{r})]. \tag{4}$$

From (4) follows that in a competitive capital market, the entrepreneur carries the monitoring costs. He therefore has incentive to search for alternative financing modes to minimize these costs.

4 A monopoly bank

Because the information obtained through monitoring by one investor cannot be directly observed by other investors, direct lending involves duplication of effort. This means that there is scope for delegating the monitoring of the projects to one agent. However, this agent must be provided with incentives to monitor and to truthfully report the outcome of the monitoring. As Diamond (1984) and Williamson (1986) show,

[8]For simplicity, the entrepreneur is assumed not to face any costs of bankruptcy. Imposing a constant bankruptcy cost on the entrepreneur would not change any of the results as long as the equilibrium interest rate is low enough for $V(\hat{r})$, net the expected bankruptcy cost, to be positive. The optimal contract is still a standard debt contract, as shown by Gale and Hellwig (1985).

one way to do this is to let the monitoring agent become an intermediary, who finances his lending to entrepreneurs by issuing debt contracts to investors.[9] The costs of providing the intermediary with incentives to be truthful is reduced by diversification of the loan portfolio, as diversification reduces the probability that the intermediary will have to declare bankruptcy. As a result, there is a size of the bank in terms of numbers of loans for which financial intermediation dominates direct lending.

Since delegation costs are reduced by diversification, and there is only a limited number of projects in the economy, it could be expected that it is optimal to allow for just one bank in the economy. In this section, I analyze the gain from monopoly banking. I will argue that the gain is larger if it is an entrepreneur who sets up the bank, than if the bank charter is given to an investor. Then, in the next section, I will show that the value of close relationships between entrepreneurs and banks can be reason for giving bank charter to more than one entrepreneur.

In order to focus on the value of close relationships between banks and firms, the entrepreneurial activity and the intermediary activity are considered as two businesses, which are managed independently of each other, although there is a common owner, and the bank writes the same loan contract with its owner as with all other entrepreneurs.[10]

I assume, for simplicity, that the bank has all the bargaining power in the negotiation with the entrepreneurs. As will be shown, this implies that bank profits will reflect the social surplus of financial intermediation.

The monopoly bank sets its loan and deposit rates so as to maximize expected profits, constrained by the need to ensure the participation of entrepreneurs and investors. In order to characterize the banker's optimal choice, the objective functions of investors, entrepreneurs, and the bank must be derived for intermediated lending.

Entrepreneurs have the same objective function given by (3) regardless of whether they write k debt contracts with individual investors, or one debt contract with the bank. The return to investors from the deposit contract will depend on the payments from entrepreneurs to the bank, because the bank will fail to pay its debt to depositors whenever the payment from entrepreneurs is less than the promised payment to depositors.

Let z be the average payment on the loans. Thus, given that the bank finances q

[9] I will assume that the bank uses debt contracts on both sides of its balance sheet. From the arguments for debt being the optimal contract between the entrepreneur and the investor, it follows that the contract between the bank and the investor should be a debt contract, but it is not obvious that the contract between the bank and the entrepreneur should be a debt contract, because the design of this contract affects the probability that the investor has to monitor the bank. Krasa and Villamil (1992b) show that for a sufficiently large bank, it is optimal with debt contracts both between the bank and the borrowers and between the bank and the investors. However, in my model, the bank may want to write another contract with its owner in order to minimize total monitoring costs, which I do not allow. This means that I may under-estimate the benefits of relationship banking.

[10] Winton (1995) considers banks owned by investors in a context similar to the current one. In his model, each bank has a fixed unit of own capital, the endowment of its owner. Winton shows that the own capital of the bank provides a cushion for depositors, which mitigates the incentive problem of the intermediary. This implies that, if the banking business and the entrepreneurial business were managed jointly, the project of the entrepreneur who owns the bank would be used as an equity stake in the banking business. Because the own capital is given, increases in bank size decrease capitalization on a per depositor basis, which means that allowing for full integration between the entrepreneurial activity and the intermediary activity would strenghten the case for the optimality of several banks.

projects,

$$z = \frac{1}{q}\sum_{i=1}^{q} \min(r_l, x_i), \tag{5}$$

where r_l is the loan rate. Either the entrepreneur pays back the loan with interest, or he defaults and the bank recovers whatever the project produced. The distribution of z, which depends on the distribution of x as well as on the number of projects and the loan rate, is denoted by $G(\cdot; r_l, q)$. An increase in the loan rate increases the payment in all states except those in which all entrepreneurs default, in which it is unaffected by the change in the loan rate. Hence, $G(z; r_l, q)$ is non-increasing in the loan rate,[11] and the expected average payment is increasing in the loan rate. Denote the expectation of z, given r_l, by $\bar{z}(r_l)$. We have that

$$\bar{z}(r_l) = r_l(1 - F(r_l)) + \int_0^{r_l} xf(x)dx.$$

Since projects are independent across entrepreneurs, the distribution of z becomes more and more concentrated around $\bar{z}(r_l)$ as q increases. According to the Strong Law of Large Numbers, z goes to $\bar{z}$ with probability one as q goes to infinity.

Now the expected return to the investor from the deposit contract can be written as

$$Y^i(r_d; r_l, q) \equiv r_d(1 - G(r_d; r_l, q)) + \int_0^{r_d} zg(z; r_l, q)dz - cG(r_d; r_l, q),$$

where r_d is the promised payment per deposit unit. The bank defaults on its debt to depositors whenever $z < r_d$.[12] The first term is the expected return in the non-default regime, the second term is the expected return in the default regime, and the last term is the expected monitoring cost.[13] Integrating the second term of Y^i by parts gives:

$$Y^i(r_d; r_l, q) = r_d - \int_0^{r_d} G(z; r_l, q)dz - cG(r_d; r_l, q), \tag{6}$$

where the integral term stands for the expected shortfalls from the stated rate on the debt. Again, due to the monitoring cost, the expected return to depositors is not

[11] Except at r_l where there is a jump of size $(1 - F(r_l))^q$, but this discontinuity does not affect the analysis since the deposit rate always is less than the loan rate.

[12] The bank bears the costs of monitoring entrepreneurs independent of whether or not it goes bankrupt. Assuming that the monitoring cost is an effort cost, which is borne by the monitoring agent, simplifies the analysis, since it allows circumvention of the limited liability restriction.

[13] The cost of the investor to monitor the bank is set equal to c for simplicity. There is no reason to believe that the cost of monitoring a bank is the same as the cost of monitoring an individual entrepreneur. More generally, the monitoring costs can be thought of as representing all kinds of bankruptcy costs such as, for example, the lost value due to early liquidation of illiquid assets, costs of collecting debt, and delayed payments, and there is no reason that these costs should be identical per invested unit for a business bankruptcy and a bank failure. However, the qualitative nature of the results does not depend on the exact form of the monitoring cost. For instance, even if we make the rather extreme assumption that each investor has to monitor each project when the bank goes bankrupt so that the monitoring cost of each investor equals qc, there is still a size of the economy for which financial intermediation dominates direct lending. See Krasa and Villamil (1992b) for an analysis of different monitoring cost structures.

monotonically increasing in the deposit rate. It is non-decreasing in the loan rate, for a given deposit rate, since the value of G is non-increasing in r_l, that is, the probability of bank failure is smaller for a higher loan rate. Finally, for a given loan rate and a given deposit rate less than $\bar{z}(r_l)$, Y^i increases with the number of projects, since G decreases with q, that is, the probability of a bank failure decreases with the diversification of the bank's portfolio.

Given the loan and deposit contracts, and the number of projects, the expected profit of the bank is

$$\Pi(r_d, r_l; q) \equiv qk \int_{r_d}^{r_l} (z - r_d) g(z; r_l, q) dz - (q-1)cF(r_l). \tag{7}$$

In case the average payment is less than the deposit payment, the bank defaults and the depositors split its assets. The first term of the profit function is the expected return of the bank in the non-default regime. The second term is the expected cost of monitoring the $(q-1)$ external projects (the bank never monitors the project of its owner).

The problem of the monopoly bank is to maximize profits under the constraints that investors should be willing to deposit their endowments at the bank, and entrepreneurs should be willing to borrow from the bank. There are many more investors than entrepreneurs, so that independent of which deposit rate the bank sets, there are investors who are willing to lend their endowments directly to entrepreneurs, given that they are promised an expected return of y. Thus, the deposit and loan contracts must be such that investors and entrepreneurs are at least as well off as they are from the direct lending equilibrium contract:

$$\max_{r_d, r_l, q} \quad \Pi(r_d, r_l; q) \tag{8}$$

s.t.

$$Y^i(r_d; r_l, q) \geq y, \tag{9}$$

$$V(r_l) \geq V(\hat{r}), \tag{10}$$

$$0 \leq r_l \leq X, \tag{11}$$

$$0 \leq r_d \leq r_l, \tag{12}$$

$$q \in \{1, 2, ...m\} \tag{13}$$

Note that there is always at least one choice of r_d, r_l, and q that fulfils the IR constraints (9) and (10), and the feasibility constraints (11)-(13), namely $r_d = r_l = \hat{r}$ and $q = 1$. This is equivalent to direct lending, since the bank finances only its owner's project. Hence, there always exists a solution to (8), since the entrepreneur always has the choice not to set up a bank, and instead finance his own project by borrowing directly from investors. The question is whether $q > 1$ in the optimal solution.

Since Y^i is not monotonically increasing in the deposit rate, there does not always exist a r_d, for feasible r_l and $q > 1$, such that $Y^i(r_d; r_l, q) \geq y$. However, there is a size of the economy in terms of the number of entrepreneurs, above which financial intermediation is feasible: if $r_d < \bar{z}(r_l)$, the probability of bank failure decreases with q, which in turn means that r_d fulfilling the IR constraint of investors approaches y, which we know is strictly less than $\bar{z}(\hat{r})$, so that for sufficiently large q, indeed $r_d < \bar{z}(\hat{r})$.[14]

Let us assume that the economy is sufficiently large for intermediation to be feasible, and characterize the solution to (8) in which the entrepreneur acts as an intermediary.

The IR constraint for investors always binds regardless of whether the IR constraint for entrepreneurs binds or not. If (9) did not bind, it would be possible to decrease r_d somewhat at a constant r_l and q. This would not affect constraint (10), and it would increase the bank's profit, which is monotonically decreasing in the deposit rate. Hence, given any loan rate and number of projects, the bank sets the deposit rate equal to the lowest value for which

$$Y^i(r_d; r_l, q) = y. \tag{14}$$

Denote the deposit rate required to fulfil the IR constraint of investors (14), given the loan rate and the number of projects, by $\tilde{r}_d(r_l, q)$. Note that $\tilde{r}_d$ decreases both with the loan rate and with the number of projects. Combining (14) with (6), and substituting the IR constraint of investors into the profit function of the bank (7), gives

$$\Pi^c(\tilde{r}_d(r_l, q), r_l, q) \equiv kq[\bar{z}(r_l) - \frac{c}{k}F(r_l)] - kq[y + cG(\tilde{r}_d(r_l, q); r_l, q)] + cF(r_l). \tag{15}$$

The term within the first brackets is the bank's expected return (per unit) from a loan contract. The term within the second brackets is the funding cost (per unit), which is equal to the alternative rate of return plus a compensation for depositors' expected monitoring costs. The last term is the benefit of the close relationship with the entrepreneur who owns the bank (the benefit of relationship banking).

Now we can conclude that also the IR constraint of the entrepreneur must bind, because the value of the constrained profit function Π^c is lower for all loan rates lower than $\hat{r}$. First, we know from the direct lending equilibrium that the first term in (15) is lower for all loan rates lower than $\hat{r}$; since the bank's expected monitoring cost per invested unit is lower than the expected monitoring cost of an individual investor, the expected return to the bank from the loan contract is strictly increasing in the loan rate at $\hat{r}$. The second term, the funding cost, is non-increasing in r_l for all loan rates, as the deposit rate is decreasing in the loan rate, and G is increasing in r_d and non-increasing in r_l. Finally, the benefit of relationship banking increases with the loan rate. Hence, for all loan rates less than $\hat{r}$ the value of Π^c would be lower than if the loan rate was equal to $\hat{r}$. Thus, for any number of projects, the bank sets its loan rate equal to the direct lending rate $\hat{r}$.

Given that the entrepreneur with the charter is going to pay the same interest rate on his loan independent of whether he is borrowing directly from investors or from

[14]For a formal proof, see Williamson (1996). Note that financial intermediation can be feasible although direct lending is not, if the scale of the projects k is sufficiently large. A large scale implies that the monitoring cost per invested unit of the bank ($\frac{c}{k}$) is low, and that the loan rate r^* which maximizes bank profits will be close to the maximum value of the project X. Then the expected average payment is close to the expected return of the project, since $\bar{z}(X) = \bar{x}$. As $y < \bar{x}$, there exists a size of the bank for which it is able to attract deposits.

his own bank, he will choose to set up the bank only if the intermediation business generates positive profits. Since $\hat{r}$ fulfills the direct lending equilibrium condition (2), we get the following non-negative-profit condition by substituting (2) into (15):

$$kq[cF(\hat{r}) - \frac{c}{k}F(\hat{r}) - cG(\tilde{r}_d(\hat{r},q);\hat{r},q)] + cF(\hat{r}) \geq 0, \tag{16}$$

which can be rewritten as

$$cF(\hat{r})\,[k + (k-1)(q-1)] \geq cG(\tilde{r}_d(\hat{r},q);\hat{r},q)kq. \tag{17}$$

Since both entrepreneurs and investors are exactly as well off as in the case of direct lending, the bank's profit comes only from saved monitoring costs. There are two ways in which the bank economizes on monitoring costs: nobody has to monitor the project of the bank's owner, and the other $(q-1)$ projects are monitored by one instead of k agents. Thus, the question is if these saved costs on monitoring entrepreneurs are large enough to cover the expected monitoring costs of depositors, for which they must be compensated.

Dividing (17) with ckq gives

$$\left[1 - \frac{1}{k} + \frac{1}{kq}\right] F(\hat{r}) \geq G(\tilde{r}_d(\hat{r},q);\hat{r},q). \tag{18}$$

Since the right hand side of (18) approaches zero as q goes to infinity while the left hand side approaches a positive number, there is a finite bank size at which the bank earns positive profits. Moreover, because both investors and entrepreneurs are as well off as in the direct lending equilibrium, monopoly banking Pareto dominates direct lending. Note also that the profitability of intermediation increases with the scale of the projects.

It remains to determine the optimal choice of q. Generally, the bank chooses to finance all m projects of the economy. As more thoroughly discussed in Appendix A, there is a special case in which the bank does not want to serve all entrepreneurs, and that is when the bank is close to making zero profits although it is almost safe. That is, when the gain from delegated monitoring is small. Then, the only profitable project is the project of the owner of the bank, and there is no point in financing another unprofitable project, as that would only lead to a very small reduction in the funding costs. I will rule out this case by assuming that the scale of the projects is sufficiently large for the monopoly bank to expect a positive return on every loan.

Before we turn to the question of the optimal structure of the financial intermediation industry, note the following:

Observation 1 *Total expected monitoring costs are lower if an entrepreneur is allowed to set up the bank than if the bank charter is given to an investor.*

If an investor owns the bank, this investor does not have to monitor the outcome of the bank's lending activities in case the bank is bankrupt. This gives rise to an informational gain. However, since the probability that the bank fails must be lower than the probability that an individual entrepreneur defaults for the bank to be viable, it follows that this gain is smaller than the gain that arises when an entrepreneur owns the bank.[15]

Observation 2 *A monopoly bank, which is owned by more than one entrepreneur, generates lower expected profits than does the bank with a single owner.*

To see this consider a monopoly bank jointly owned by two entrepreneurs. Since the outcome of monitoring is private information, each of these entrepreneurs has incentive to cheat the other and claim that the outcome of the projects that he has monitored is very low. This means that the owners have to check each other so that a common ownership of the bank will give rise to duplication of effort.[16]

5 Optimal industrial structure

The monitoring costs connected to lending in this economy is a dead weight loss to society. The total dead weight loss will depend on the structure of the financial market. Given that relationship banking is considered to be the best way to organize lending, the social optimization problem is to determine how many entrepreneurs that should be allowed to set up a bank, and what rates the banks should offer to entrepreneurs and depositors. On the one hand, there are economies of scale in intermediation; the expected monitoring costs of depositors decrease with the size of the bank. This points towards one bank. On the other hand, there is an informational gain of relationship banking, which points towards multiple banks.

Consider a social planner, who can control entry into intermediation and the interest rates offered by banks, but who cannot prohibit direct lending, or force entrepreneurs to operate as intermediaries. The planner does not care about income distribution, so he chooses the number of banks, n, and sets the rates so that investments are carried out at a minimum of expected monitoring in the economy as a whole:

$$\min_{r_d, r_l, n} \left[cF(r_l)(m-n) + cG(r_d; r_l, \tfrac{m}{n})km\right] \tag{19}$$

s.t.

$$Y^i(r_d; r_l, \tfrac{m}{n}) \geq y, \tag{20}$$

$$V(r_l) \geq V(\hat{r}), \tag{21}$$

$$\Pi(r_d, r_l, \tfrac{m}{n}) \geq 0, \tag{22}$$

[15] Of course, the total expected monitoring costs in the economy would be lower if the investor owning the bank held equity instead of debt. However, as already mentioned, the same would be true if the entrepreneur put his project as an equity stake in the bank. Equity in the hands of the bank owner will always reduce delegation costs. The point raised here is that there is a larger informational gain when the bank is owned by an entrepreneur.

[16] If a group of entrepreneurs for some reason trusted each others, these entrepreneurs could jointly own the bank, and the informational gain would be larger than if the bank was owned by only one entrepreneur. Such trust can be created if the entrepreneurs are involved with each other in more than one way so that the cost of cheating becomes too high. See Berglöf and Perotti (1994) for an analysis of how a financial network can work as a mechanism for effective corporate control.

$$0 \leq r_l \leq X, \quad (23)$$

$$0 \leq r_d \leq r_l, \quad (24)$$

$$n \, \epsilon \{1, 2, ..., m\}. \quad (25)$$

We see that if the planner did not recognize the benefit of relationship banking, the optimal number of banks would be one, since a single bank involves the lowest risk of bank failure. However, as a consequence of allowing one more entrepreneur to run a bank, the costs spent on monitoring entrepreneurs in the economy (the first term of the objective function) are reduced. This gain should be weighed against the increased costs spent on monitoring banks (the second term of the objective function), as a larger number of banks implies a smaller number of projects per bank and thereby an increased probability of bank failure.

The restrictions (20) and (21) are the IR constraints for investors and entrepreneurs, respectively. In the same way as the monopoly bank, the social planner is constrained by the alternatives of investors and entrepreneurs. They must not prefer the direct lending contract, as direct lending involves higher monitoring costs than intermediation. Restriction (22) is the IR constraint for intermediaries. The bank must make non-negative profits, otherwise no entrepreneur would be willing to set up a bank. The restrictions (23), (24), and (25) are the feasibility constraints for the interest rates and the number of banks. Now, we have the following result:

Proposition 1 *There is a scale of the projects above which minimizing expected monitoring costs is equivalent to maximizing total expected bank profits, subject to the IR constraints of depositors and entrepreneurs.*

Proof: Just as the monopoly bank, the planner sets the deposit rate equal to the lowest value for which $Y^i(r_d; r_l, \frac{m}{n}) = y$. If this constraint did not bind, it would be possible to decrease r_d somewhat at a constant r_l and n. This would not violate any of the other constraints, but it would reduce the expected monitoring costs, since the probability of bank failure would be reduced. Hence, for any choice of the loan rate and the number of banks, the optimal deposit rate is given by equation (14).

An increase in the loan rate has two counteracting effects on the objective function: it increases the probability that the entrepreneur defaults, while it reduces the probability of bank failure, both directly and indirectly via the IR constraint of depositors. Since the first term of the objective function is independent of the scale, while the second term increases with the scale, there exists a scale of the projects above which the second effect dominates and the objective function is decreasing in the loan rate. Then, the planner, in the same way as the monopoly bank, sets the loan rate equal to the direct lending rate.[17] If the IR constraint for entrepreneurs (21) did not bind, it would be possible to increase the loan rate somewhat at a constant number of banks,

[17] We know that the profit of the bank increases with the loan rate at $\hat{r}$. From (15) we can see that this does not necessarily imply that the social objective function decreases with the loan rate at $\hat{r}$, since $\bar{z}$ is strictly increasing in the loan rate, but the more restricted the profit maximum is by direct lending, the more likely it is that direct lending also will restrict the social optimum. From (15) we also can see that in case there is no direct lending equilibrium, the monopoly loan rate will be too high: then the monopoly bank will set the loan rate at a level where the social objective function is increasing in the loan rate.

and thereby reduce the value of the objective function. This would not violate the IR constraint for intermediaries (18), since both the increase in the loan rate and the induced decrease in the deposit rate increases the profit of the bank.

As shown by (16), when both investors and entrepreneurs just receive their reservation values, the profit of the bank is equal to the saved monitoring costs compared to direct lending. Multiplying the left hand side of (16) by n and noting that q is equal to $\frac{m}{n}$, gives the following expression for the total expected bank profits:

$$cF(\hat{r})km - [cF(\hat{r})(m-n) + cG(\tilde{r}_d(\hat{r}, \tfrac{m}{n}); \hat{r}, \tfrac{m}{n})km].$$

Hence, maximizing total expected profits of the banks is equivalent to minimizing expected monitoring costs. $Q.E.D.$

Assume that k is large enough for the IR constraint of entrepreneurs to bind. In this case, the IR constraint of banks does not bind. This follows from the assumption that a monopoly bank is expected to make strictly positive profits. Since the optimal structure must involve at least as high total expected bank profits as monopoly banking, each bank must make strictly positive profits in the social optimum. As the planner does not care about income distribution, he sees no point in giving depositors more than their reservation value, since that would increase the probability of bank failure. Neither does he see a point in hindering banks from exploiting entrepreneurs as much as possible. Since large-scale investments imply that the bank's monitoring cost per invested unit is low, while there are many depositors who all have to monitor the bank in case it defaults, the safety of the bank should have the highest priority. Intuitively, the more profitable a bank is expected to be, the less likely it is to fail.

It remains to determine how many banks that should be allowed to operate. Certainly, not every entrepreneur can have his own bank, since that is the same as direct lending, and monopoly banking is assumed to dominate direct lending. The question is whether the safety of the bank has such a high priority that the optimal number of banks is equal to one. Noting that independently of the optimal number of banks the optimal loan rate is equal to $\hat{r}$, and denoting the solution to the IR constraint of depositors for a given number of banks by $\hat{r}_d(n)$, the value of the objective function for $n+1$ banks can be compared to the value for n banks. This gives the following expression for $n+1$ banks to be better than n banks:

$$km[cG(\hat{r}_d(n+1); \hat{r}, \tfrac{m}{n+1}) - cG(\hat{r}_d(n); \hat{r}, \tfrac{m}{n})] \leq cF(\hat{r}). \qquad (26)$$

The left hand side in (26) represents the extra monitoring cost of depositors which is expected due to the existence of one more bank. The probability of bank failure increases directly as well as indirectly as the deposit rate must be raised to compensate depositors for their increased monitoring costs. The right hand side is the saved monitoring cost on the credit side.

There is a unique solution to the social optimization problem (19) when the left hand side of (26) increases monotonically with n. This is what we intuitively would expect. A small number of banks implies a large numbers of projects per bank. Then each bank is well diversified and the G-distribution rather concentrated around its mean. Since adding one more project to a distribution, which already is concentrated, should hardly affect the shape of the distribution, an increase in n should have a smaller effect on the left hand side of (26) at smaller values of n than at larger. However, there is a counteracting effect in that an increase in n at a small n involves a larger change in the number of projects per bank than at a large n. The formal analysis is presented in appendix B. There it is shown that in the case of the normal distribution the effect on the distribution dominates the effect on the number of projects, so that the left hand side does indeed increase in n. Then, according to the Central Limit Theorem, the same should be true for all G-distributions which are concentrated enough to behave as the normal distribution. Since G must be rather concentrated for the IR constraint of depositors to be fulfilled, I will treat the left hand side in (26) as monotonically increasing in n. Then, the optimal number of banks is the smallest number of n for which condition (26) is violated.

The optimal number of banks depends on how costly a bank failure is to society compared with a business bankruptcy. This cost difference will depend on the size of the economy and the scale of the projects. Consider the following comparative examples:

Increasing the scale, k, for a given number of projects, m.

Projects of a large scale imply large gains from delegated monitoring relative to the gains from relationship banking. The larger is the scale of the projects, the smaller is the optimal number of banks, given the number of projects.

Increasing the number of projects, m, for a given scale of the projects, k.

A large number of projects implies that there are many investors who are affected by increased monitoring costs when one more bank is established. Therefore, for a given scale of the projects, the bank industry should be safer in a larger economy than in a smaller one. However, this does not imply that the optimal number of banks is smaller in a larger economy, since the diversification possibilities is better in an economy with many projects.

Increasing the number of projects, m, for given total investments, km.

The more projects there are, for given total investments, the larger is the optimal number of banks. That is, an economy with many small projects should have more banks than an economy with few large projects.

In addition to the scale and number of projects, the optimal number of banks depends on the alternative rate of return, the size of the monitoring cost, and the distribution of projects. A higher alternative rate of return implies a higher direct lending equilibrium loan rate. This, in turn, implies larger bank profits, and a reduced probability of bank failure. It also implies a larger gain from relationship banking, because of a higher probability for the entrepreneur to default. Hence, the optimal number of banks increase with the direct lending rate. A higher monitoring cost also implies a higher direct lending equilibrium rate. Therefore, given that the monitoring cost per invested unit is the same for an individual loan as for a bank failure, a higher monitoring cost has the same effect as a higher alternative rate of return. The size of all these effects depends on the distribution of projects.

6 A numerical example

In this section, I illustrate the characterization of the optimal solution by a numerical example. Consider the special case in which the project outcome x is a discrete random variable with two possible outcomes, $x_l < x_h$, where the probability of the lower output is p. The alternative rate of return lies between these values, *i.e.*, $x_l < y < x_h$. The monitoring cost, c, is small enough for the direct lending rate, given by

$$px_l + (1-p)\hat{r} - pc = y,$$

to be in the interval (x_l, x_h). For a bank, which finances q projects, the average payment, z, is completely determined by the number of entrepreneurs who default and the loan rate. For instance, if s entrepreneurs default and $(q-s)$ repay their loans, $z = \frac{sx_l + (q-s)\hat{r}}{q}$. As the outcome of the project is independently distributed across entrepreneurs, the number of entrepreneurs who default is a binomial random variable with parameters (q, p).

Table 1 illustrates how the optimal number of banks depends on the parameters of this economy. First, we can conclude that the IR constraint of entrepreneurs binds for all $k \geq 2$. To see this, note that the probability that the entrepreneur defaults is constant for all loan rates in the interval (x_l, x_h). Then, the social planner wishes to set the loan rate as high as possible within this interval, since that minimizes the probability of a bank failure. Hence, independent of the size of the economy, the planner sets the loan rate equal to the direct lending rate.

Table 1.

exogenous variables	**case a**	**case b**	**case c**
distribution of x	$x_l = 1$, $x_h = 3$		$x_l = 0$, $x_h = 4$
probability of x_l, p	0.5		
alternative return, y	1.25		
monitoring cost, c	1	1.25	
scale of projects, k	10		
number of entrepreneurs, m	1000		
endogenous variables			
optimal deposit rate, $\hat{r}_d$	1.25009	1.25009	1.25002
optimal loan rate, $\hat{r}$	2.5	2.75	3.5
optimal numbers of banks, n^*	34	45	5
probability of bank failure,	0.00008	0.00008	0.00002
$G(\hat{r}_d, \hat{r}, \frac{m}{n^*})$			
bank profit, Π	133	128	900
cost savings	4516	5652	4502

Start with case (a) in the table, which will be used as a benchmark case. The project outcome takes with equal probability one of the two values $x_l = 1$, and $x_h = 3$, the alternative rate of return is equal to 1.25, and the monitoring cost is equal to 1. Then, the direct lending equilibrium rate is equal to 2.5. If it takes 10 units to finance one project and there are 1000 entrepreneurs, condition (26) says that the optimal number of banks is the one for which one more bank implies an increased probability of bank failure exceeding $0.5 \cdot 10^{-4}$. This allows for 34 banks, each with a probability of bankruptcy equal to 0.00008, to be compared with the default probability of an individual entrepreneur, which is 0.5.

Increasing the scale of the projects increases the bankruptcy costs of a bank failure relative to the costs of an individual default by an entrepreneur. Thus, the safety of banks should increase with the scale of projects. For instance, if $k = 100$, the planner allows for 26 banks, each with a probability of default equal to 0.000004 to be compared with 0.00008 when $k = 10$. The same is true for an increase in the number of entrepreneurs, but then more projects also make room for more banks. For instance, if the number of entrepreneurs is equal to 10 000, the optimal probability of bank failure is again 0.000004, but there is room for 10 times as many banks, that is, 260 banks. This illustrates what was said above about banks being safer in a larger economy.

Increasing the monitoring cost, or the alternative rate of return, should increase the optimal number of banks. In case (b) the monitoring cost is equal to 1.25, which implies a direct lending rate of 2.75. This increase in the monitoring cost aggravates the asymmetric information problem and increases the benefit of intermediation compared to direct lending, but also the relative benefit of relationship banking. A higher loan rate *ceteris paribus* increases the profitability of the bank and thereby reduces the probability of bank failure, which makes room for more banks. In this case, the optimal number of banks is equal to 45.

Increasing the risk of the project (while preserving the mean) also aggravates the asymmetric information problem, but it does not increase the advantage of intermediation compared to direct lending. On the contrary, if the outcome in the two states are changed but not the probability of the state occurring, as in case (c), the comparative advantage of intermediation is reduced. In this case, the social planner allows for only 5 banks.

Thus, the informational gain from relationship banking can be large enough for it to be optimal with several banks.[18] However, since a bank failure is more costly to society than a default by an individual entrepreneur, it should be less likely. The larger the cost difference, the safer banks should be compared to entrepreneurial firms.

7 Industrial structure in a free market

Having derived the optimal industrial structure, a natural question is what the structure would be in a free market. In this section, I will briefly argue that in general there is no reason to believe that a free market would provide the optimal structure of the financial intermediation industry. There are at least two factors that indicate that entry into an unregulated market would not lead to the optimal number of banks.

[18] This result is complementary to the result of Krasa and Villamil (1992a). They show that, if there also is aggregate risk, monitoring cost increasing in bank size can put a limit on the bank size. Thus, starting with a monopoly, there might be only benefits of increasing the number of banks. When reaching the optimal size with respect to monitoring cost, benefits of relationship banking suggests that one should increase the number of banks even further.

Firstly, the economies of scale in intermediation imply that banks have difficulties to share the market. If borrowers are sensitive to prices, each bank has incentive to undercut the offers of its competitors to increase the scale of lending. As a result of this undercutting incentive, price competition for loan customers among banks can result in only one active bank. To see this, consider the following three-stage game: In the first stage, entrepreneurs determine whether or not to establish a bank. Entrepreneurs make their decisions sequentially, and they face a small entry cost. Banks are numbered according to the order in which they enter the market. In the second stage, the established banks compete for loan customers. Banks independently and simultaneously set their loan rates. Entrepreneurs sign contracts with the bank with the lowest rate. If all banks offer the same loan rate, entrepreneurs go to the bank which was the first to enter the market.[19] In the third and last stage, banks independently and simultaneously offer deposit contracts to investors to finance their loans to entrepreneurs.[20]

The equilibrium of this game is that only the entrepreneur who is the first to make the entry decision, establishes a bank, and this bank behaves as the monopoly bank in section 4. Since there is a large enough amount of credit available to finance all projects, and investors behave competitively, there is no competition for deposits among banks in the last stage. Each bank offers the deposit contract which gives its depositors an expected return of y, and signs contracts with just enough depositors to finance its loans to entrepreneurs. Now, if there were more than one bank in the market, each bank would offer the loan rate at which it expects to make zero profits, given that it gets all entrepreneurs as loan customers, and every entrepreneur would borrow from the bank which was the first to enter the market. This is the lowest loan rate at which an entrepreneur is willing to act as intermediary. At every higher loan rate bank number two would have incentive to undercut in order to receive all entrepreneurs as loan customers, and thereby make positive profits. Then given that the first entrepreneur has established a bank, no other entrepreneur will do so, since they will not be able to recover the entry cost. Knowing this, the first entrepreneur will indeed establish a bank, if the entry cost is low enough.

Hence, if the optimal number of banks is larger than one, the market in this case provides too few banks.[21]

[19]This is also true for entrepreneurs who have established banks. Given that all other entrepreneurs behave in this way, the owner of a bank cannot benefit from deviating, although there is an informational gain if the entrepreneur borrows from his own bank. This behaviour of entrepreneurs gives the bank which enters first an advantage. We may think of this as a short-cut for an explicit dynamic model where incumbent banks have an advantage because borrowers face a small switching cost. In the symmetric Bertrand game, where entrepreneurs are distributed evenly among banks that charge the same loan rate, there is no Nash equilibrium in pure strategies.

[20]It is assumed that a bank that serves the whole market is able to attract deposits although it makes zero profits. That is, a monopoly bank is large enough for the zero-profit condition to be compatible with the IR constraint of investors.

[21]Yanelle (1989) analyzes a similar type of game, but she allows entrepreneurs to actively choose bank and focuses on the coordination problem of entrepreneurs in their choice of bank. She assumes that a bank which does not get enough loan customers to reach a scale at which intermediation is profitable at the promised loan and deposit rates abandons the market, leaving its entrepreneurs without finance. This creates an externality among entrepreneurs, because the probability that the entrepreneur actually gets the loan depends on how many other entrepreneurs that apply for credit

The other reason for free entry not to provide the optimal number of banks is the existence of an externality. As another bank enter the market and borrowers are distributed among more banks, the diversification of each bank, and thereby its safety, is reduced. The potential entrant is concerned with its own safety, since it affects his funding cost, but he does not consider the effect of his entry on the safety of the whole banking system. Hence, if banks cooperate in their rate setting and share the market, once they are established, there could be too many banks in the market equilibrium. To see this suppose that the banks cooperate on the best possible loan rate, the direct lending equilibrium rate. We know from section 4 that, when the bank sets the loan rate equal to the direct lending rate, the profits of the bank arise from saved monitoring costs. Free entry implies that there are so many banks in the market that a potential entrant cannot enter the market and make positive profits. This means that in this case only indivisibilities and entry costs hinder entry from eating up the total benefit of intermediation compared to direct lending, and if entry costs are low there will be too many banks in the market equilibrium.

To summarize, the threat of fierce competition can result in too few banks, while more cooperative behaviour might result in too much entry.

8 Heterogeneous entrepreneurs

In the analysis, I have assumed that all entrepreneurs are alike. In this section, I briefly discuss the case in which there are small and large entrepreneurs, where for simplicity the small are like the entrepreneurs in the previous analysis. Which type should get the bank charter? Intuitively, the informational gain from a close relationship with a bank is larger for a large entrepreneur than for a small one. However, this depends on the *bank's* expected cost of monitoring the entrepreneur being increasing with the size of the entrepreneur, which is not always the case.

Assume that it is the size of the project that varies between small and large entrepreneurs. The previous analysis shows that the gain from financial intermediation increases with the size of the projects, because with large projects there is more duplication of efforts in direct lending. However, this benefit is captured by the large entrepreneur borrowing from a bank. The extra informational gain from close relationships is larger for the large entrepreneur only if *the bank's cost of monitoring the project is increasing with the scale of the project.*

Assume instead that a large entrepreneur has access to more than one project. In this case the benefit of close relationships between the bank and a large entrepreneur does not only depend on the size of the bank's monitoring costs in case the entrepreneur defaults, but also on the correlation among the entrepreneur's projects. Take the case in which the bank's cost of monitoring a large entrepreneur is the same as the bank's cost of monitoring a small entrepreneur, given that the entrepreneur has defaulted. Then the benefit is larger for a *small* entrepreneur as long as the projects are not

at the same bank. Yanelle shows that due to this externality there exist equilibria with several active banks. However, she shows that, if one imposes the criterion of pay-off dominance, the unique equilibrium of her game is one active bank offering the loan rate at which it expects to make zero profits. In that equilibrium there is not only too few banks, there is also a too low loan rate.

perfectly correlated, in which case the benefit is the same, simply because the larger entrepreneur due to diversification is less likely to fail.[22] In the more general case, in which the bank's monitoring cost in case of default is increasing in the size of the entrepreneur, the benefit depends on whether the bank's *expected* monitoring cost is decreasing or increasing in the size of the entrepreneur. When the bank's monitoring cost is increasing with the size of the defaulting borrower, it is optimal to give the bank charter to an entrepreneur with many highly correlated projects. Because a diversified entrepreneur faces less of an incentive problem, the benefit of a close relationship with a bank is smaller for a well-diversified entrepreneur than for an undiversified entrepreneur.

9 Conclusions

In this paper, I analyze the optimal structure of the financial intermediation industry in a finite economy, where banks are owned by entrepreneurs. The role of banks is to reduce information costs in lending of resources from investors to entrepreneurs. There are two ways in which a bank owned by an entrepreneur economizes on monitoring costs. Firstly, delegating the monitoring task to the bank is a way of avoiding duplication of effort, since the scale of inputs for the project exceeds the wealth of any single investor. Secondly, since the entrepreneur has no incentive to be dishonest with his own bank, nobody has to monitor the project of the bank's owner. The gain from delegation increases with bank size, because the costs of providing the owner of the bank with incentive to be truthful to investors decrease with the diversification of the bank portfolio, while the informational gain from relationship banking is independent of the size of the bank.

In contrast to the previous literature, which has considered the incentives of investors to form and use intermediaries, I allow entrepreneurs to set up intermediaries. If an investor owns the bank, this investor does not have to monitor the bank in case it goes bankrupt. However, this is an unlikely event for a diversified bank. Therefore, in this model, it is better for the economy as a whole that banks are owned by entrepreneurs rather than depositors. This, however, does not imply that the costs of the asymmetric information problem would be minimized, if all entrepreneurs jointly owned one bank. A joint venture among entrepreneurs would create an incentive problem among the shareholders of the bank, who would have to check each others' projects, for each one to be sure to get his share of the portfolio return.

Given that relationship banking is the socially optimal way of organizing lending, the social optimization problem is to determine how many entrepreneurs that should be allowed to set up a bank, and what interest rates banks should charge. If a bank failure is much more costly to society than an individual business bankruptcy, banks should be allowed to charge monopoly rates. The optimal number of banks depends on the size of the economy as well as the scale of the projects. In a small economy with large projects there are small diversification possibilities. In this economy the optimal

[22]For this cost structure, the cost savings from financial intermediation decreases with the size of the entrepreneur. This follows from the analysis in Krasa Villamil (1992b). Intuitively, a perfectly diversified entrepreneur can borrow directly from investors without agency costs.

number of banks is small, perhaps equal to one. However, in a large economy with not too large projects there is room for several banks.

I have here argued that there is no reason to believe that a free market would provide the optimal structure of the financial intermediation industry. But, regulating price competition, without regulating entry at the same time, does not seem to be a good idea, since this might promote too much entry. Banks would then become too small and risky. The opposite policy, regulating entry but not interest rates, then seems better, especially if a bank failure is very costly.

In this model, the social planner does not worry about large spreads between deposit and loan rates for two reasons. Firstly, the planner does not care about income distribution. Secondly, a high loan rate does not affect investment decisions. Concerning the first assumption, one could argue that it seems reasonable in the present context; the analysis of income distribution objectives requires a more structural model of the social planner. In defence of the second assumption one may argue that, since the bank has full information about the projects, it should be able to use non-linear prices to extract consumer surplus, and it is well-known that a perfectly price discriminating monopolist does not distort resource allocation.

While the model seems to catch some important features of intermediation, it has some obvious limitations. Perhaps the most limiting simplification of the model is that it is based on one specific type of asymmetric information, which implies that the behaviour of the entrepreneur is not influenced by whether or not he owns the bank. There could be other information environments, where there would be an adverse effect on the behaviour, which then would reduce the benefit of relationship banking, or even turn it into a social loss. Indeed, this is one reason for some countries, for instance the USA, to forbid universal banking. For instance, from the theory of corporate finance it is known that an entrepreneur who is close to bankruptcy may have incentives to take too large risks, and even to undertake projects with negative expected returns, in an attempt to avoid bankruptcy.[23] The question, however, which asks for more research, is whether this incentive leads to larger costs if the entrepreneur has a close relationship with the bank.

A Appendix

In this appendix, I analyze the optimal choice of the number of loans by the monopoly bank. In order to analyze how Π^c behaves as the number of projects is changed, treat q as a continuous variable and differentiate (15) with respect to q[24] :

$$\frac{\partial \Pi^c}{\partial q} = k[\bar{z}(\hat{r}) - \frac{c}{k}F(\hat{r}) - y - cG(\tilde{r}_d(\hat{r},q);\hat{r},q)] - ckq\left[\frac{\partial G}{\partial \tilde{r}_d}\frac{\partial \tilde{r}_d}{\partial q} + \frac{\partial G}{\partial q}\right], \qquad (27)$$

where $\frac{\partial \Pi^c}{\partial q}$ is evaluated at the optimal interest rates. If the intermediary expects negative profits, he will not operate at all. On the other hand, if the intermediary expects

[23] See for instance Brealey and Myers (1984).

[24] G is differentiable in q according to the Convolution Theorem for Fourier transforms: the Fourier transform of the convolution is the product of the Fourier transforms of the individual distributions. See Feller (1971).

positive return on each external project, when $q = m$, he will choose to finance all m projects of the economy; since G is decreasing in q, $\frac{\partial \Pi^c}{\partial q}$ is positive for all q for which the term within the first brackets (the expected net profit from an external loan) is non-negative. Hence, for the intermediary to choose to finance some (at least q^c), but not all, projects of the economy, he must expect a negative return on each external project, but positive profits in total. We can note that if the first term is positive for some q', it is positive for all $q \geq q'$. Furthermore, there is always a finite q', given the scale of the projects, for which the expected return on each external loan is positive, as the direct lending equilibrium condition (2) implies that

$$\bar{z}(\hat{r}) - \frac{c}{k}F(\hat{r}) > y.$$

The number of projects required to fulfil the IR constraint of depositors, q^c, can very well be larger than q'. In that case the intermediary chooses to provide all m entrepreneurs with credit. However, if $q^c \leq m < q'$, it is possible that the intermediary expects positive profits in total, but a loss on each external project. Then, whether or not the bank will finance all m projects depends on whether or not the reduction in the funding cost of adding one project outweighs the cost of financing one more (a priori) unprofitable project. Since the bank is assumed to make positive profits, the total expected loss on the external projects cannot be larger than the benefit of relationship banking, which enters the profit function as a fixed revenue, $cF(\hat{r})$. Thus, the cost of adding one unprofitable project is less than $\frac{cF(\hat{r})}{q}$. The total reduction of the funding costs of a small change in q is $ckq\left[\frac{\partial G}{\partial \hat{r}_d}\frac{\partial \hat{r}_d}{\partial q} + \frac{\partial G}{\partial q}\right]$. Hence, only if the change in G due to an increase in the number of projects is smaller than $\frac{cF(\hat{r})}{kq^2}$, when $q^c \leq q \leq m$, the bank will choose to finance some but not all projects. In this case some entrepreneurs borrow directly from investors.

B Appendix

In this appendix, I discuss the condition for the existence of a unique solution to the social optimization problem in (19) in the case when the scale of the projects is sufficiently large for the IR constraints of investors and entrepreneurs to bind. I show that, in relevant cases, this condition is fulfilled for the normal distribution.

The derivative of the objective function with respect to n is

$$ckm\left(\frac{\partial G}{\partial \hat{r}_d}\frac{d\hat{r}_d}{dn} + \frac{\partial G}{\partial n}\right) - cF(\hat{r}), \tag{28}$$

where $\hat{r}_d(n)$ fulfils the IR condition of investors for $r_l = \hat{r}$. The term within parenthesis is positive, since a larger number of banks implies that each bank provides fewer loans, which increases the probability of bank failure both directly and indirectly through a higher deposit rate. If this term increases monotonically with n, there is a unique solution to the social optimization problem.

Consider the case in which x_i is normally distributed with mean $\bar{x}$ and variance σ^2. Then, the average, $z = \frac{n}{m}\sum x_i$, is normally distributed with mean $\bar{x}$, and variance

$\frac{n}{m}\sigma^2$. Assume for simplicity that $\sigma^2 = 1$. Then $G(z; \frac{m}{n}) = \Phi(\sqrt{\frac{m}{n}}(z-\bar{x}))$, and $g(z; \frac{m}{n}) = \varphi(\sqrt{\frac{m}{n}}(z-\bar{x}))\sqrt{\frac{m}{n}}$, where Φ is the cumulatitive distribution function and φ is the density function of a standard normal variable, and we can rewrite the term within the parenthesis in (28) as

$$g(r_d; \tfrac{m}{n})\left(\frac{d\hat{r}_d}{dn} - \frac{1}{2n}(r_d - \bar{x})\right). \tag{29}$$

The derivative of (29) w.r.t. n is equal to

$$\left(\frac{\partial g}{\partial \hat{r}_d}\frac{d\hat{r}_d}{dn} + \frac{\partial g}{\partial n}\right)\left(\frac{d\hat{r}_d}{dn} - \frac{1}{2n}(r_d - \bar{x})\right) +$$
$$g(r_d; \tfrac{m}{n})\left(\frac{d^2\hat{r}_d}{dnd\hat{r}_d}\frac{d\hat{r}_d}{dn} + \frac{d^2\hat{r}_d}{dn^2} + \frac{1}{2n^2}(r_d - \bar{x}) - \frac{1}{2n}\frac{d\hat{r}_d}{dn}\right),$$

which can be rewritten as

$$\begin{aligned}&\left(\frac{\partial g}{\partial \hat{r}_d}\frac{d\hat{r}_d}{dn} + \frac{\partial g}{\partial n} - g(r_d; \tfrac{m}{n})\frac{1}{2n}\right)\frac{d\hat{r}_d}{dn} - \\ &\left(\frac{\partial g}{\partial \hat{r}_d}\frac{d\hat{r}_d}{dn} + \frac{\partial g}{\partial n} - g(r_d; \tfrac{m}{n})\frac{1}{n}\right)\frac{1}{2n}(r_d - \bar{x}) + \\ &g(r_d; \tfrac{m}{n})\left(\frac{d^2\hat{r}_d}{dnd\hat{r}_d}\frac{d\hat{r}_d}{dn} + \frac{d^2\hat{r}_d}{dn^2}\right).\end{aligned} \tag{30}$$

I will show that a sufficient condition for (30) to be positive is that

$$\frac{m}{n}(r_d - \bar{x})^2 > 3, \tag{31}$$

which is to say that the probability of bank failure should be less than $\Phi(-\sqrt{3}) = 0.0418$. This should be true in all relevant cases as 0.0418 must be considered as a very high probability of bank failure. In the numerical examples given in section 6, the optimal probability of bank failure is far below this number.

We have that

$$\frac{\partial g(r_d; \frac{m}{n})}{\partial n} = -\varphi'(\sqrt{\tfrac{m}{n}}(r_d - \bar{x}))\frac{1}{2n}\sqrt{\frac{m}{n}}(r_d - \bar{x})\sqrt{\frac{m}{n}} - \varphi(\sqrt{\tfrac{m}{n}}(r_d - \bar{x}))\frac{1}{2n}\sqrt{\frac{m}{n}},$$

which can be rewritten as

$$\begin{aligned}\frac{\partial g(r_d; \frac{m}{n})}{\partial n} &= \varphi(\sqrt{\tfrac{m}{n}}(r_d - \bar{x}))\sqrt{\frac{m}{n}}\frac{1}{2n}\left(\left[\sqrt{\frac{m}{n}}(r_d - \bar{x})\right]^2 - 1\right) \\ &= g(r_d; \tfrac{m}{n})\frac{1}{2n}\left(\left[\sqrt{\frac{m}{n}}(r_d - \bar{x})\right]^2 - 1\right)\end{aligned} \tag{32}$$

since $\varphi'(x) = -x\varphi(x)$. Substituting (32) into (30) gives that the first two terms in (30) are positive when assumption (31) is fulfilled.

It remains to be shown that the last term is positive. Define

$$H(r_d; \tfrac{m}{n}) = r_d - \int_0^{r_d} G(z; \tfrac{m}{n})dz - cG(r_d; \tfrac{m}{n}) - y.$$

By the Implicit Function Theorem we have

$$\frac{d\hat{r}_d}{dn} = -\frac{\frac{\partial H}{\partial n}}{\frac{\partial H}{\partial \hat{r}_d}},$$

and

$$\begin{aligned} \frac{d^2\hat{r}_d}{dnd\hat{r}_d} &= -\frac{\left[\frac{\partial^2 H}{\partial n \partial r_d}\frac{\partial H}{\partial \hat{r}_d} - \frac{\partial^2 H}{\partial \hat{r}_d^2}\frac{\partial H}{\partial n}\right]}{\left(\frac{\partial H}{\partial \hat{r}_d}\right)^2}; \\ \frac{d^2\hat{r}_d}{dn^2} &= -\frac{\left[\frac{\partial^2 H}{\partial n^2}\frac{\partial H}{\partial \hat{r}_d} - \frac{\partial^2 H}{\partial \hat{r}_d \partial n}\frac{\partial H}{\partial n}\right]}{\left(\frac{\partial H}{\partial \hat{r}_d}\right)^2}, \end{aligned} \tag{33}$$

where

$$\begin{aligned} \frac{\partial H}{\partial n} &= \int_0^{r_d} g(z; \tfrac{m}{n})\frac{1}{2n}(z - \bar{x})dz + cg(r_d; \tfrac{m}{n})\frac{1}{2n}(r_d - \bar{x}) < 0; \\ \frac{\partial H}{\partial \hat{r}_d} &= 1 - G(r_d; \tfrac{m}{n}) - cg(r_d; \tfrac{m}{n}) > 0; \\ \frac{\partial^2 H}{\partial n \partial r_d} &= \frac{1}{2n}g(r_d; \tfrac{m}{n})(r_d - \bar{x}) + \frac{1}{2n}cg(r_d; \tfrac{m}{n}) + \frac{1}{2n}c\frac{dg}{dr_d}(r_d - \bar{x}); \\ \frac{\partial^2 H}{\partial \hat{r}_d^2} &= g(r_d; \tfrac{m}{n}) - c\frac{dg}{dr_d}; \\ \frac{\partial^2 H}{\partial n^2} &= \int_0^{r_d}\left[\frac{dg(z; \frac{m}{n})}{dn}\frac{1}{2n}(z - \bar{x}) - \frac{1}{2n^2}g(z; \tfrac{m}{n})(z - \bar{x})\right]dz + \\ & \quad c\left[\frac{dg(r_d; \frac{m}{n})}{dn}\frac{1}{2n}(r_d - \bar{x}) - \frac{1}{2n^2}g(z; \tfrac{m}{n})(r_d - \bar{x})\right], \end{aligned}$$

and $\frac{\partial^2 H}{\partial r_d \partial n} = \frac{\partial^2 H}{\partial n \partial r_d}$. We have that

$$\frac{\partial g(r_d; \frac{m}{n})}{\partial r_d} = \varphi'(\sqrt{\tfrac{m}{n}}(r_d - \bar{x}))\frac{m}{n}, \tag{34}$$

which can be rewritten as

$$= -\varphi(\sqrt{\tfrac{m}{n}}(r_d - \bar{x}))\sqrt{\frac{m}{n}}(r_d - \bar{x})\frac{m}{n} = -g(r_d; \frac{m}{n})(r_d - \bar{x})\frac{m}{n}.$$

Substituting (34) into $\frac{\partial^2 H}{\partial n \partial r_d}$ and $\frac{\partial^2 H}{\partial \hat{r}_d^2}$, gives that these derivative are negative when (31) is fulfilled. Substituting (32) into$\frac{\partial^2 H}{\partial n^2}$ gives that also this derivative is negative. This gives that the last term in (33) is positive when (31) is fulfilled. Q.E.D.

References

[1] Bhattacharya, S, and A V Thakor (1993), "Contemporary Banking Theory", *Journal of Financial Intermediation*, 3, 2-50.

[2] Benston, G J (1994), "Universal Banking", *Journal of Economic Perspectives*, 8, 121-143.

[3] Berglöf, E, and E Perotti (1994), "The Governance Structure of the Japanese Financial Keiretsu", *Journal of Financial Economics*, 36, 259-284.

[4] Boyd, J, and E Prescott (1986), "Financial Intermediary-Coalitions", *Journal of Economic Theory*, 38, 211-232.

[5] Brealey, R, and S Myers (1984), *Principles of Corporate Finance*, sec ed, McGraw-Hill Book Company.

[6] Cable, J (1985), "Capital Market Information and Industrial Performance: The Role of West German Banks", *The Economic Journal*, 95, 118-132.

[7] Diamond, D (1984), "Financial Intermediation and Delegated Monitoring", *Review of Economic Studies*, 51, 393-414.

[8] Feller, W (1971), *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd Ed., John Wiley & Sons, New York.

[9] Gale, D, and M Hellwig (1985), "Incentive-Compatible Debt Contracts I: The one-Period Problem", *Review of Economic Studies*, 52, 647-664.

[10] Gertler, M (1988), "Financial Structure and Aggregate Economic Activity: An Overview", *Journal of Money Credit and Banking*, 20, 559-588.

[11] Hoshi, T, A K Kashyap, and D S Scharfstein (1990), "The Role of Banks in Reducing the Costs of Financial Distress in Japan", *Journal of Financial Economics*, 27, 67-88.

[12] Hoshi, T, A K Kashyap, and D S Scharfstein (1991), "Corporate Structure, Liquidity, and Investment: Evidence from Japanese Industrial Groups, *Quarterly Journal of Economics*, 106, 33-60.

[13] Krasa, S and A P Villamil (1992a), "A Theory of Optimal Bank Size", *Oxford Economic Papers*, 44, 725-749.

[14] Krasa, S and A P Villamil (1992b), "Monitoring the Monitor: An Incentive Structure for a Financial Intermediary", *Journal of Economic Theory*, 57, 197-221.

[15] Lindgren, H (1990), "Long-Term Contracts in Financial Markets: Bank-Industry Connections in Sweden, Illustrated by the Operations of Stockholms Enskilda Bank, 1900-70", in M Aoki, B Gustafsson, O E Williamson (eds.), *The Firm as a Nexus of Treaties*, Swedish Collegium for Advanced Study in the Social Sciences series, London.

[16] Mayer, C (1988), "New Issues in Corporate Finance", *European Economic Review,* 32, 1167-1189.

[17] Sundin, A and S I Sundqvist (1995), *Owners and Power in Sweden's Listed Companies,* Dagens Nyheters Press.

[18] Townsend, R (1979), "Optimal Contracts and Competitive Markets with Costly State Verification, *Journal of Economic Theory,* 21, 265-293.

[19] Williamson, S D (1986), "Costly Monitoring, Financial Intermediation, and Equilibrium Credit Rationing", *Journal of Monetary Economics,* 18, 159-179.

[20] Winton, A J (1993), "Competition Among Financial Intermediaries when Diversification Matters", Banking Research Center working paper 185, Northwestern University.

[21] Winton, A J (1995), "Banks as Delegated Monitors: Industrial Structure in a Finite Economy", *Journal of Financial Intermediation,* 4, 158-187.

[22] Yanelle, M O (1995), "Banking Competition and Market Efficiency", WP, Delta.