Seminar Paper No. 489

OPTIMAL STRUCTURE OF THE FINANCIAL INTERMEDIATION INDUSTRY

by

Sonja Daltung

INSTITUTE FOR INTERNATIONAL ECONOMIC STUDIES
Stockholm University
Optimal Structure of the Financial Intermediation Industry

by

Sonja Daltung

Seminar Papers are preliminary material circulated to stimulate discussion and critical comment.

May 1991

Institute for International Economic Studies
S-106 91 Stockholm
Sweden
OPTIMAL STRUCTURE OF THE FINANCIAL INTERMEDIATION INDUSTRY

by

Sonja Daltung⁰

Institute for International Economic Studies
Stockholm University

May 1991
Revised October 1992

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.

⁰This is a revised version of SP No. 489, IIES. 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 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 the Jacob Wallenberg Foundation is gratefully acknowledged.
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 largely been due to the lack of theory explaining the functions of banks and other financial intermediaries. However, the development of informational economics has made it possible to rigorously analyze the raison d'être of financial intermediaries, and a great deal of effort has during the last decade 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 recent theoretical developments as a starting point to characterize the socially optimal structure of the financial intermediation industry, an issue which has so far been largely neglected.²

A recurring theme in the new 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 most countries do

¹See Gertler (1988) for a survey of this work and its implications for macroeconomic behaviour.

²As far as I know, there is no other study of the optimal structure of the financial intermediation industry that takes the role of the intermediary explicitly into account. There are some papers analyzing the market structure of endogeneous financial intermediation. These include Yanelle (1989) and Winton (1992). Yanelle (1989) shows how a coordination problem of entrepreneurs in their choice of bank can give rise to market equilibria with several active banks, although there are economies of scale in intermediation. Winton (1992) argues that there can be several active banks in the market equilibrium, because fixed own capital of the bank counteracts economies of scale in intermediation. Their relationship to my paper is discussed in the paper.
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 value of close relationships between banks and firms. As will be shown, a single bank might not be 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. But close relationships between banks and firms are also an important feature of actual banking. The most apparent examples are the banking systems of Germany and Japan, but there are close relations between banks and firms in other countries too, although less formal. For instance, for almost every Swedish commercial bank there is a business group that is associated with the bank. There are some empirical studies 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. For instance, Cable (1985) and Hoshi et al (1990) argue that the banking systems of Japan and Germany is one of the reasons for the relatively good performance of these economies.

The relationship analyzed in this paper is common ownership of firms and banks. The structure of the financial intermediation industry is examined in the context of the model of delegated monitoring, first developed by Diamond (1984), in a situation where banks are owned by entrepreneurs. The economy is similar to that analyzed by Williamson (1986), but with the important difference that the total number of agents is finite. 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

---

I will refer to this as "relationship banking", a term often used for banking systems which build upon closed relations between banks and firms.
action referred to as monitoring. The optimal contract between the entrepreneur and his investors is a so-called "standard debt contract", which minimizes the resources spent on monitoring by restricting monitoring to the default states.

The first question addressed in the paper is whether financial intermediation by a monopoly bank Pareto dominates direct lending. There are two ways in which this 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. However, 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 bank, otherwise its owner would always claim a low outcome and pay a small amount to investors. These costs decrease with the size of the bank in terms of the number of loans to entrepreneurs. It is shown that financial intermediation will dominate direct lending, if there are sufficiently many projects of sufficiently large scale in the economy.

The main question concerns the optimal structure of the financial intermediation industry. There are economies of scale in intermediation in the sense that the costs of delegated monitoring decrease with bank size. 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 to operate.

The optimal industrial structure depends on how costly a bank failure is to society compared with a business bankruptcy. If this cost difference is sufficiently large, the social planner will allow banks to charge the monopoly interest rates. The more profitable the bank is, the less likely it is to fail. But, although the safety of the bank industry as a whole
decreases with the number of banks, the planner might allow more than one entrepreneur set up a bank, while the society benefits from the close relationship between banks and firms with a common owner. The optimal number of banks depends on the parameters of the economy, especially the size of the economy. 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 economy is described in the next section. The case of direct lending is examined in section 3, while the monopoly bank is analyzed in section 4. The optimal structure of the financial intermediation industry is characterized in section 5, and a numerical example is provided in section 6. In section 7, the industrial structure in a free market is briefly discussed. 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. Finally, section 8 summarizes the results and discusses possibilities for further research.

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 $[0, x^m]$. 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 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. The entrepreneur must offer an incentive compatible contract to investors for them to be willing to make the loan. 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 Townsend (1979), Gale and Hellwig (1985), and Williamson (1986), where the optimal contract is shown to be a "standard debt contract"\(^4\): either the entrepreneur repays his investors a fixed amount (the amount borrowed plus interest), or the entrepreneur defaults, each investor monitors, and the investors split the realized output of the project.

\(^4\)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 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. Hence, the probability of bankruptcy and thereby the expected monitoring costs are minimized.

The standard debt contract is completely characterized by its fixed payment. Let \( r \) be the promised gross interest rate on the loan. Then the expected return to the investor from the direct lending contract is

\[
Y^d(r) = \int_{0}^{r} xf(x)dx + r(1 - F(r)) - cF(r).
\]

(1)

The first term is the expected return in the default regime, the second term is the expected return in the non-default regime, and the last term is the expected monitoring cost. Integrating the first term in (1) by parts gives

\[
Y^d(r) = r - \int_{0}^{r} F(x)dx - cF(r),
\]

(2)
where the integral term stands for the expected shortfalls from the stated rate on the debt.

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

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

which is negative for $r = x^m$, since $f(r) > 0$ for $r \in [0, x^m]$. Thus, $Y^d(r)$ reaches a maximum for $r < x^m$. 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. Furthermore, in order to avoid multiple equilibria, $Y^d$ is assumed to be concave, that is,

\begin{equation}
f(r) + cf(r) > 0, \quad \forall \ r \in [0, x^m].
\end{equation}

The expected return to the entrepreneur is

\begin{equation}
V(r) = k \int_{x-r}^{x} (x-r)f(x)dx,
\end{equation}

which is monotonically decreasing in $r$, but positive as long as there are some states in which the project return exceeds the promised payment.\footnote{This holds for the uniform distribution, for instance.}

The entrepreneur wants to carry out his project as long as he expects a positive return. Hence, he is willing to offer an interest rate as high as the largest outcome $x^m$, if \footnote{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 $V(r)$, net the expected bankruptcy cost, to be positive. The optimal contract is still a standard debt contract, as shown by Gale and Hellwig (1985).}
necessary, in order to attract funds. On the other hand, since investors are risk–neutral, they are willing to lend their endowments if they are offered a contract which gives an expected return of at least \( y \), the alternative rate of return. 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

\[
\hat{r} = \frac{\int xf(x)dx + \hat{r}(1 - F(\hat{r})) - cF(\hat{r})}{0} = y,
\]

where

\[
1 - F(\hat{r}) - cf(\hat{r}) \geq 0.7
\]

4. A Monopoly Bank

The information obtained through monitoring by one investor cannot be directly observed by other investors. Direct lending therefore involves duplication of effort: each entrepreneur borrows from \( k \) investors, each monitoring the entrepreneur when he defaults. Thus, 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, one way in which this incentive problem can be solved is that the monitoring agent becomes an

---

\( ^7 \)Since the left hand side in (6) is not monotonically increasing in \( r \), there may not exist a \( \hat{r} \) that solves equation (6). If a solution does not exist, there will be no direct lending at all. In the main text, I will focus on the case in which the monitoring cost is low enough for there to be a direct lending equilibrium.
intermediary, who finances his lending to entrepreneurs by issuing debt contracts to investors. In this section, I analyze under what conditions financial intermediation by a monopoly bank dominates direct lending in the sense that the bank provides intermediation on terms which make everybody at least as well off as in the direct lending equilibrium.

The bank is owned by an entrepreneur, who has obtained a charter to act as the sole intermediary. The entrepreneur acts as intermediary in addition to his entrepreneurial activity. As shown in the previous section, the entrepreneur bears the costs of asymmetric information, since investors have access to alternative investments. The entrepreneur therefore would like to commit to be truthful ex post. One way for the entrepreneur to commit is to borrow from a bank of his own, since he has no incentive to be dishonest with his own bank.

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. Thus, the bank writes the same loan contract with its owner as with all other entrepreneurs.\(^8\)

A 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 (5)) regardless of whether they write \(k\) debt contracts with individual investors, or one debt contract with the bank.

---

\(^8\)Winton (1992) 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. Winton shows that, because the own capital is given, increases in bank size decrease capitalization on a per depositor basis. This means that allowing for full integration between the entrepreneurial activity and the intermediary activity would strengthen the case for the optimality of several banks.
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 \( q \) projects, including its own,

\[
(8) \quad z = \frac{1}{q} \sum_{i=1}^{q} \min(r_{\ell} x_i),
\]

where \( r_{\ell} \) 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 \( G(z; r_{\ell}, q) \). 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, \( G(z; r_{\ell}, q) \) is non-increasing in the loan rate\(^9\), and the expected average payment is increasing in the loan rate. Denote the expectation of \( z \), given \( r_{\ell} \) by \( \bar{z}(r_{\ell}) \). We have that

\[
(9) \quad \bar{z}(r_{\ell}) = \int_{0}^{r_{\ell}} x f(x) dx + r_{\ell} (1 - F(r_{\ell})).
\]

Since projects are mutually stochastically independent, the distribution of \( z \) becomes more and more concentrated around \( \bar{z}(r_{\ell}) \) as \( q \) increases. According to the Strong Law of Large Numbers, \( z \) goes to \( \bar{z} \) with probability one as \( q \) goes to infinity. This is assumed to be a

---

\(^9\)Except at \( r_{\ell} \) where there is a jump of size \((1 - F(r_{\ell}))^q\), but this discontinuity does not affect the analysis since the deposit rate always is less than the loan rate.
smooth process so that, if \( z < \bar{z} \), then \( G(z; \bar{r}_\epsilon q_1) < G(z; \bar{r}_\epsilon q_2) \) for all \( q_1 > q_2 \).\(^{10}\)

Now, the expected return to the investor from the deposit contract is

\[
Y^i(r_d; r_\epsilon q) = \int_{0}^{r_d} zg(z; r_\epsilon q) dz + r_d(1 - G(r_d; r_\epsilon q)) - cG(r_d; r_\epsilon q),
\]

where \( r_d \) is the promised payment per deposit unit. The bank bears the costs of monitoring entrepreneurs independent of whether or not it goes bankrupt.\(^{11}\) Thus, the bank will default on its debt to depositors whenever \( z \leq r_d \). In case the bank defaults, each depositor monitors the bank. As before, the first term is the expected return in the default regime, the second term is the expected return in the non-default regime, and the last term is the expected monitoring cost.\(^{12}\) Integrating the first term of \( Y^i \) by parts gives:

\[
Y^i(r_d; r_\epsilon q) = r_d - \int_{0}^{r_d} G(z; r_\epsilon q) dz - cG(r_d; r_\epsilon q).
\]

Again, due to the monitoring cost, the expected return to depositors is not 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 \( r_\epsilon \). The probability of bank failure is smaller

---

\(^{10}\)This holds, for instance, for the uniform and the normal distributions.

\(^{11}\)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.

\(^{12}\)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, because the qualitative nature of the results does not depend on this monitoring cost, it is for simplicity set equal to \( c \).
for a higher loan rate. Finally, for a given loan rate and a given deposit rate less than \( \bar{r}(r_d) \), \( Y^i \) increases with the number of projects, since \( G \) decreases with \( q \). 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, \textit{the expected profit of the bank is}

\[
\Pi(r_d, r_c, q) = kq \int r_d \leq r \leq r_c \left( \frac{z - r_d}{z} \right) dG(z; r_c, q) dz - cF(r_c)(q - 1).
\]

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), which is always borne by the bank irrespective of whether or not it goes bankrupt.

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:

\[\text{The effect on the return of depositors from an increase in the loan rate is unambiguously positive, since the costs of the increased probability of default of entrepreneurs are borne by the intermediary.}\]
Let us first discuss existence of a solution, and then characterize the solution given existence. Note that there is always at least one choice of $r_d$, $r_\ell$ and $q$ that fulfils the "individual rationality" (IR) constraints (i) and (ii), and the feasibility constraints (iii) – (v), namely $r_d = r_\ell = \hat{r}$ 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 (13), since the entrepreneur always has the choice not to operate as an intermediary, and instead finance his own project by borrowing directly from investors. However, the question is whether there exists an intermediary solution to (13), that is whether there exists a solution when $q$ is restricted to the set $\{2,3,\ldots,m\}$. Since $Y^i$ is not monotonically increasing in the deposit rate, there does not always exist a $r_d$, for all feasible $q$ and $r_\ell$ such that $Y^i(r_d, r_\ell; q) \geq y$. We have the following result:

Lemma 1: There is always a finite size of the economy, defined as the number of entrepreneurs ($m$), for which there is an intermediary solution to (13).

Proof: As the bank grows ($q$ goes to infinity), the deposit rate required to fulfil constraint (i) converges to $y$, for $y \leq \bar{z}(r_\ell)$. The direct lending rate fulfils constraint (ii), and according to the direct lending equilibrium condition (6), $y$ is strictly less than $\bar{z}(\hat{r})$. Thus,
there is a finite \( q^C \) for which the IR constraints can be fulfilled. Then, for \( m \geq q^C \), there exists a solution to (13) when \( q \) is restricted to the set \( \{2, 3, \ldots, m\} \). Q.E.D. 14

I will assume that \( m \geq q^C \), and then characterize the optimal solution to (13) in order to examine whether the intermediary solution dominates direct lending.

The IR constraint for investors always binds regardless of whether the IR constraint for entrepreneurs binds or not. If (i) did not bind, it would be possible to decrease \( r_d \) somewhat at a constant \( r_{\ell} \) and \( q \). This would not affect constraint (ii), 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

\[
Y^i(r_d, r_{\ell}, q) = y. \tag{14}
\]

Combining (14) with (10), and substituting the IR constraint of investors into the profit function of the bank (12), gives

\[
\Pi^C(r_d, r_{\ell}, q) = kq[z(r_{\ell}) - (c/k)F(r_{\ell})] - kq[y + cG(r_d, r_{\ell}, q)] + cF(r_{\ell}). \tag{15}
\]

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

14 In the case where the monitoring cost is so high that the asymmetric information is an obstacle to direct lending, the bank is free to set the loan rate so as to maximize profits. Then, the ability of the bank to attract deposits also depends on the scale of the projects \( k \). A large scale implies that the monitoring cost per invested unit of the bank \((c/k)F(r^*_\ell)\) is low, and that the optimal loan rate \( r^*_\ell \) will be close to the maximum value of the project \( x^m \). Then the expected average payment is close to the expected return of the project, since \( \bar{z}(x^m) = \bar{x} \). As \( y < \bar{x} \), there always exists a scale of the projects for which a large enough bank is able to attract deposits.
monitoring costs. The last term is the benefit of the close relationship with the entrepreneur who owns the bank (the benefit of relationship banking).

The constrained profit function $\Pi^C$ decreases with $r_d$, since $G$ increases with $r_d$, given $r_\ell$ and $q$. $\Pi^C$ increases with $r_\ell$ for all $r_\ell \leq \hat{r}$. To see this, recall that the expression for the expected average payment is

$$\bar{z}(r_\ell) = \int_0^{r_\ell} xf(x)dx + r_\ell(1 - F(r_\ell)),$$

so that the derivative of the first term of $\Pi^C$ with respect to $r_\ell$ is

$$kq[1 - F(r_\ell) - (c/k)f(r_\ell)].$$

From (7) it follows that (16) is strictly positive for $r_\ell = \hat{r}$. 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 $\hat{r}$. From (4) then follows that (16) is also positive for all loan rates less than $\hat{r}$. Hence, the first term of $\Pi^C$ increases with the loan rate, for $r_\ell \leq \hat{r}$. The second term, the funding cost, is non-increasing in $r_\ell$ for all loan rates, as $G$ is non-increasing in $r_\ell$. Finally, the benefit of relationship banking increases with the loan rate. Hence, $\Pi^C$ increases with $r_\ell$ for all $r_\ell \leq \hat{r}$.

Denote the deposit rate required to fulfil the IR constraint of investors (14), given the loan rate and the number of projects, by $\bar{r}_d(r_\ell q)$. Note that $\bar{r}_d$ decreases both with the loan rate and with the number of projects. Then, noting that the IR constraint of entrepreneurs implies that the loan rate cannot be higher than the direct lending rate, and that the number of projects must be at least $q^C$ for intermediation to be feasible, the maximization problem of the bank can be restated as:
\[ \max_{\hat{r}_d, \hat{r}_q} \Pi^C(\hat{r}_d, \hat{r}_q) \]

s.t.

(i) \[ 0 \leq \hat{r}_d \leq \hat{r} \]

(ii) \[ q \in \{1, q^c, q^c+1, \ldots, m\} \]

It is now obvious that the IR constraint of entrepreneurs also binds. If this constraint did not bind, that is, if the loan rate was lower than the direct lending rate \( \hat{r} \), it would be possible to increase the loan rate somewhat at a constant \( q \). This would increase \( \Pi^C \) both directly, and indirectly via the induced decrease in the deposit rate. Hence, for any number of projects, the bank sets its loan rate equal to the direct lending rate \( \hat{r} \).

Given the optimal loan and deposit rates, the question is whether the entrepreneur with the bank charter chooses to act as an intermediary, or whether he prefers to borrow directly from investors, that is, whether or not intermediation is viable. For the entrepreneur as bank to choose \( q > 1 \), the expected gain from the loan and deposit contracts must cover the bank's expected monitoring costs. Substituting the direct lending equilibrium condition (6) into \( \Pi^C \) gives the following non-negative-profit condition:

\[ kq[cF(\hat{r}) - (c/k)F(\hat{r}) - cG(\hat{r}_d(\hat{r}, q); \hat{r}, q)] + cF(\hat{r}) \geq 0, \]

which can be rewritten as

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

To interpret condition (19) recall that both entrepreneurs and investors are exactly as well off as in the case of direct lending. Thus, the bank's profit fully arises from saved
monitoring costs. This means that, if the bank is viable, everybody is at least as well off as in the direct lending equilibrium, that is intermediation Pareto dominates direct lending. 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. We have the following result:

**Proposition 1:** There is always a finite size of the economy and a finite scale of the projects for which monopoly banking dominates direct lending.

Proof: Rewriting \((19)\) in per invested unit gives

\[
(20) \quad \left[ 1 - \frac{1}{k} \left( 1 - \frac{1}{q} \right) \right] F(\hat{r}) \geq G(\hat{r};q;\hat{r},q).
\]

A binding IR constraint for investors ((14) combined with (10)) implies that

\[
(21) \quad cG(\hat{r},q) = \int_{0}^{r_d} zg(z;\hat{r},q)dz + r_d(1 - G(r_d;\hat{r},q)) - y.
\]

Since \(z(\hat{r})\) is equal to the expectation of \(z\), it follows from the definition of expectation that

\[
(22) \quad \int_{0}^{r_d} zg(z;\hat{r},q)dz + r_d(1 - G(r_d;\hat{r},q)) < z(\hat{r}).
\]

Hence,
(23) \[ cG(r_d;\hat{r},q) < \tilde{z}(\hat{r}) - y. \]

From (6) we know that

(24) \[ cF(\hat{r}) = \tilde{z}(\hat{r}) - y. \]

Thus, the probability that the bank fails, G, must be strictly less than the probability that an individual entrepreneur defaults, F, for the IR constraint of investors to be fulfilled. Then, from (20) there is always a finite \( k \) for which intermediation is profitable, when the economy is large enough for there to exist an intermediary solution. Q.E.D.

It remains to determine the optimal choice of \( q \). Assume that \( k \) is large enough for intermediation to be profitable.\(^{15}\) Then, generally, the bank chooses to finance all \( m \) projects of the economy. As more thoroughly discussed in appendix A, there is only one 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. This case is ruled out by the assumption that the monopolist expects a positive return on each loan.

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

---

\(^{15}\)This does not need to be a restrictive assumption. It could very well be the case that intermediation is profitable for any scale of the projects, given that the economy is large enough for there to exist an intermediary solution. That is, it is possible that condition (20) is fulfilled for all \( k \geq 2 \), when \( q = q^c \).
Observation 1: A monopoly bank, which is owned by more than one entrepreneur, expects less 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 to monitor every non–performing loan of the bank in order to get his share of the portfolio return. Thus, a common ownership of the bank results in duplication of effort.

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.\footnote{Since $G$ is strictly decreasing in the number of projects, optimality requires that entrepreneurs are evenly distributed among banks. Here it is disregarded that a given number of entrepreneurs cannot be evenly distributed among just any number of banks. When the number of projects per bank is needed to be an integer, it is implicitly assumed that the remaining entrepreneurs borrow directly from investors. Since $m$ should be large compared to $n$, the deviation from the optimal asymmetric solution should not be too large.}
\[ \begin{align*}
(25) \quad \min_{r_d, r_L, \alpha} & \left[ cF(r_d)(m - n) + cG(r_d; r_L m/n)km \right] \\
\text{s.t.} & \\
(i) & \quad V^i(r_d; r_L m/n) \geq y \\
(ii) & \quad V(r_d) \geq V(\bar{x}) \\
(iii) & \quad \Pi(r_d; r_L m/n) \geq 0 \\
(iv) & \quad 0 \leq r_L \leq \alpha^m \\
(v) & \quad 0 \leq r_d \leq r_L \\
(vi) & \quad n \epsilon \{1, 2, \ldots, m\}.
\end{align*} \]

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 (i) and (ii) 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 (iii) is the IR constraint for intermediaries. The bank must make non-negative profits, otherwise no entrepreneur would be willing to act as an intermediary. The restrictions (iv) – (vi) are the feasibility constraints for the interest rates and the number of banks. Now, we have the following result:
Proposition 2: There is always a scale of the projects for 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 \( Y^1(\bar{r}_d; \bar{r}_m/n) = y \). If this constraint did not bind, it would be possible to decrease \( r_d \) somewhat at a constant \( r_m \) 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 constraint (i).

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 always exists a scale of the projects for 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. If the IR constraint for entrepreneurs (ii) did not bind, it would be possible to increase the loan rate somewhat at a constant number of banks, and thereby reduce the value of the objective function. This would not violate the IR constraint for intermediaries (iii), since both the increase in the loan rate and the induced decrease in the deposit rate

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

As shown by (18), 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 (18) by \( n \) and noting that \( q \) is equal to \( m/n \), gives the following expression for the total expected bank profits:

\[
(26) \quad cF(\bar{r})km - [cF(\bar{r})(m-n) + cG(\bar{r}, m/n; \bar{r}, m/n)km].
\]

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 the optimal loan rate is equal to \( \bar{r} \), 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:

$$(27) \quad km[cG(\hat{r}_d(n+1); \bar{r}, m/(n+1)) - cG(\hat{r}_d(n); \bar{r}, m/n)] \leq cF(\bar{r}).$$

The left hand side in (27) 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.

If the left hand side of (27) increases monotonically with $n$, there is a unique solution to the social optimization problem (25). This is what intuitively is expected. A small number of banks implies a large number 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 (27) 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, the left hand side in (27) is assumed to be monotonically increasing in $n$. Then the optimal number of banks is the largest number of $n$ for which condition (27) is fulfilled.
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 will depend 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, $x_\ell < x_h$, where the probability of the lower output is $p$. The alternative rate of return lies between these values, i.e., $x_\ell < y < x_h$. The monitoring cost, $c$, is small enough for the direct lending rate, given by

$$px_\ell + (1-p)\bar{r} - pc = y,$$

(28)

to be in the interval $(x_\ell, x_h)$. 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, $z = [sx_\ell + (q-s)\bar{r}]/q$. 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 $(x_\ell, x_h)$. 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.

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 $x_\ell = 1$, 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 (27) 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 0.000008, 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 0.000008 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. 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.

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.\(^{18}\)

\(^{18}\)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.
In the third and last stage, banks independently and simultaneously offer deposit contracts to investors to finance their loans to entrepreneurs.\textsuperscript{19}

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.\textsuperscript{20}

\textsuperscript{19}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.

\textsuperscript{20}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 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
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. 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.

her game is one active bank which offers 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.
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 act as intermediaries. The informational value of a close relationship between an investor and the bank would be that the investor does not have to monitor the bank in case it goes bankrupt, which 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, and each entrepreneur would have to monitor the projects of the other entrepreneurs 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 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, he 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 defense 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. 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. One could think of 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. 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.\(^{21}\) In such a situation a close relationship between the bank and the entrepreneur may be bad for society. Indeed, this is an explicit reason why authorities in some countries, for instance in the US, oppose close relations between banks and their borrowers. Against this stands the view that a close relationship between the

\(^{21}\)See for instance Brealey and Myers (1984).
firm and its bank reduce informational asymmetries that would otherwise lead to credit-rationing. Thus, fewer profitable investment opportunities are likely to be foregone with a banking system of the German or the Japanese type.

Perhaps the most limiting simplification of the model is to analyze relations in banking in a static environment. Especially, it is well-known that customer relationships are important in banking. The entrepreneur typically has private information about his project ex ante as well as ex post. A bank learns about its customers through the lending process. The firm continues to borrow from the same bank because it has better information about the firm's activities than other banks have, and therefore can offer loans on better terms. Customer relations are therefore likely to moderate the competition among banks, and should be of great importance to the structure of financial intermediation. The effect of customer relations on the behaviour and structure of financial intermediaries is an interesting area of future research.

Appendix A

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

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

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 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

---

$^{22}$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).
first brackets (the expected net profit from an external loan) is non-negative. Hence, for
the intermediary to choose to finance some (at least q̇), 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 ≥ 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 (6) implies that

\[ \tilde{z}(\hat{r}) - (c/k) F(\hat{r}) > y. \]

The number of projects required to fulfil the IR constraint of depositors, q̇, can very well
be larger than q'. In that case the intermediary chooses to provide all m entrepreneurs
with credit. However, if q̇ ≤ 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.
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 cF(\hat{r})/q. The total reduction of the
funding costs of a small change in q is, according to (A1), 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
F(\hat{r})/(kq^2), when q̇ ≤ q ≤ m, the bank will choose to finance some but not all projects. In
this case some entrepreneurs borrow directly from investors.

Appendix B

In this appendix, I discuss the conditions for the existence of a unique solution to
the social optimization problem in (25) in the case when the IR constraints of investors and
entrepreneurs bind. I show that, in relevant cases, these conditions are fulfilled for the
normal distribution.

Recall condition (27) for n+1 banks to better than n banks:
(B1) \[ \text{ckm}[G(\hat{r}_d(n+1); \hat{r}, m/(n+1)) - G(\hat{r}_d(n); \hat{r}, m/n)] - cF(\hat{r}) \leq 0. \]

If the left hand side in (B1) is monotonically increasing in \( n \), there is a unique solution to (25). In order to see how the left hand side behaves when the number of banks changes, treat \( n \) as a continuous variable and evaluate the derivative of the objective function in (25) with respect to \( n \) at the optimal interest rates:

(B2) \[ \text{ckm}(\frac{\partial G}{\partial \hat{r}_d} \frac{d\hat{r}_d}{dn} + \frac{\partial G}{\partial q} \frac{dq}{dn}) - cF(\hat{r}). \]

I will now determine how (B2) changes with \( n \). First, note that

\[ \frac{\partial G}{\partial \hat{r}_d} = g(\hat{r}_d, q), \quad \frac{d\hat{r}_d}{dn} = \frac{d\hat{r}_d}{dq} \frac{dq}{dn}, \quad \frac{\partial G}{\partial q} = \frac{\partial G}{\partial q} \frac{dq}{dn}, \]

where the optimal loan rate is suppressed and \( q = m/n \). Thus, (B2) can be rewritten as

(B3) \[ \text{ckm}(g(\hat{r}_d, q) \frac{d\hat{r}_d}{dq} + \frac{\partial G}{\partial q} \frac{dq}{dn}) - cF(\hat{r}). \]

When \( n \) increases, the number of projects per bank decreases, which increases the probability of bank failure both directly, and indirectly via the IR constraint of depositors. Now, the derivative of (B3) with respect to \( n \) is:

(B4) \[ \left[ \left( \frac{\partial G}{\partial \hat{r}_d} \frac{d\hat{r}_d}{dq} + \frac{\partial G}{\partial q} \frac{d\hat{r}_d}{dq} + g \frac{d^2\hat{r}_d}{dq^2} + \frac{\partial^2 G}{\partial q \partial \hat{r}_d} \frac{d\hat{r}_d}{dq} + \frac{\partial^2 G}{\partial q^2} \right) \left( \frac{dq}{dn} \right)^2 + \left[ g \frac{d\hat{r}_d}{dq} + \frac{\partial G}{\partial q} \right] \left( \frac{dq}{dn} \right)^2 \right] \]

For (B4) to be positive the effect on the distribution of a change in \( n \) (the first term) must be positive and large enough to dominate the effect on the number of projects per bank (the second effect), which is negative. I will derive a sufficient condition for this to be true for the normal distribution.

Assume that \( x_i \) is normally distributed with mean \( \bar{x} \) and variance \( \sigma^2 \). Then, the average, \( z = \Sigma x_i / q \), is normally distributed with mean \( \bar{x} \) and variance \( \sigma^2 / q \). Without loss
of generality, set $\bar{x} = 0$, and $\sigma = 1$. Then,

\[(B5) \quad g(z, q) = \sqrt{\frac{q}{2\pi}} e^{-z^2 q/2}, \quad G(\tilde{d}, q) = \sqrt{\frac{q}{2\pi}} \int_{-\infty}^{\tilde{d}} e^{-z^2 q/2} dz.\]

Consider first the effect on $(B3)$ of a change in $n$ when the deposit rate is constant. Then $(B4)$ is reduced to

\[(B6) \quad \frac{\partial^2 G}{\partial q^2} \left( \frac{dq}{dn} \right)^2 + \frac{\partial G}{\partial q} \left( \frac{d^2 q}{dn^2} \right).\]

We have that

\[(B7) \quad \frac{\partial G}{\partial q} = \frac{1}{2\sqrt{2\pi q}} \int_{-\infty}^{\tilde{d}} e^{-z^2 q/2} dz - \sqrt{\frac{q}{2\pi}} \int_{-\infty}^{\tilde{d}} \frac{z^2}{2} e^{-z^2 q/2} dz.\]

Integrating the second integral in $(B7)$ by parts gives

\[(B8) \quad \frac{\partial G}{\partial q} = \frac{1}{2\sqrt{2\pi q}} \tilde{d} e^{-\tilde{d}^2 q/2},\]

which is negative for $\tilde{d} < \bar{x}$. Differentiating $(B8)$ with respect to $q$ gives

\[(B9) \quad \frac{\partial^2 G}{\partial q^2} = -\frac{1}{2\sqrt{2\pi q}} \tilde{d} e^{-\tilde{d}^2 q/2} \left[ \frac{1}{2q} + \frac{\tilde{d}^2}{2} \right],\]

which is positive for $\tilde{d} < \bar{x}$. Now, substituting $(B8)$, $(B9)$, and the expressions for the derivatives of $q$ into $(B6)$ gives the following condition for $(B6)$ to be larger than zero

\[(B10) \quad \tilde{d}^2 q > 3.\]

This is true as long as $\tilde{d}$ is not too close to the mean, $\bar{x} = 0$, and $q$ is not too small. I will
show that (B10) is a sufficient condition for (B4) to be positive.

We have that

\[
\frac{\partial \bar{q}}{\partial \hat{r}_d} = -\hat{r}_d q \frac{2}{2\pi} e^{-\hat{r}_d q/2} > 0, \quad \text{for } \hat{r}_d < \bar{r},
\]

(B11)

\[
\frac{\partial^2 \bar{q}}{\partial q \partial \hat{r}_d} = \frac{1}{2\sqrt{2\pi}q} \hat{r}_d e^{-\hat{r}_d q/2} \left(1 - \frac{1}{\hat{r}_d q}\right),
\]

(B12)

which is negative whenever (B10) is fulfilled. Thus, as \( \frac{d\hat{r}_d}{dq} \) is negative, it remains to show that

\[
\frac{d^2 \hat{r}_d}{dq^2} \left(\frac{dq}{dn}\right)^2 + \frac{d\hat{r}_d}{dq} \frac{d^2 q}{dn^2} > 0,
\]

(B13)

whenever (B10) is fulfilled.

The optimal deposit rate for a given number of projects per bank, \( \hat{r}_d(q) \), is given by

\[
\hat{r}_d - \int_{-\infty}^{R} G(z,q)dz - cG(\hat{r}_d,q) - y = 0.
\]

(B14)

Define

\[
H(\hat{r}_d,q) = \hat{r}_d - \int_{-\infty}^{R} G(z,q)dz - cG(\hat{r}_d,q) - y.
\]

(B15)

Then, according to the Implicit Function Theorem,

\[
\frac{d\hat{r}_d}{dq} = -\frac{\partial H}{\partial q} / \frac{\partial H}{\partial \hat{r}_d}.
\]

(B16)

Differentiating (B16) with respect to \( q \) gives
\[
\frac{d^2i_d}{dq^2} = \left[ \frac{\partial^2 H}{\partial q^2} \frac{\partial H}{\partial r_d} - \frac{\partial H}{\partial q} \frac{\partial^2 H}{\partial r_d \partial q} \right] / \left( \frac{\partial H}{\partial r_d} \right)^2.
\]

Hence, condition (B13) can be written as

\[
\left[ \frac{\partial^2 H}{\partial q^2} \frac{\partial H}{\partial r_d} - \frac{\partial H}{\partial q} \frac{\partial^2 H}{\partial r_d \partial q} \right] / \frac{\partial H}{\partial q} \frac{\partial H}{\partial r_d} > 2/q.
\]

We have that

\[
\frac{\partial H}{\partial r_d} = 1 - G(i_d, q) - cG(i_d, q) \geq 0,
\]

\[
\frac{\partial^2 H}{\partial r_d \partial q} = - \frac{\partial G}{\partial q} - c \frac{\partial G}{\partial q} > 0
\]

Hence, the second term contributes positively to the left hand side of (B18). This means that it is enough to show that

\[
\frac{-\partial^2 H}{\partial q^2} / \frac{\partial H}{\partial q} > 2/q.
\]

This follows from

\[
\frac{-\partial^2 G}{\partial q^2} / \frac{\partial G}{\partial q} > 2/q,
\]

since

\[
\frac{\partial H}{\partial q} = - \int_{-\infty}^{r_d} \frac{\partial G}{\partial q} dz - c \frac{\partial G}{\partial q}.
\]

\[
\frac{\partial^2 H}{\partial q^2} = - \int_{-\infty}^{r_d} \frac{\partial^2 G}{\partial q^2} dz - c \frac{\partial^2 G}{\partial q^2}.
\]
Hence, for the normal distribution, the left hand side in (B1) is monotonically increasing in n whenever (B10) is fulfilled.

References


Yanelle, M O (1989), "Increasing Returns to Scale and Endogeneous Intermediation",
WWZ—Discussions Papers, No 8908.
Table 1.

<table>
<thead>
<tr>
<th></th>
<th>case a</th>
<th>case b</th>
<th>case c</th>
</tr>
</thead>
<tbody>
<tr>
<td>distribution of $x$</td>
<td>$x_l=1$, $x_h=3$</td>
<td></td>
<td>$x_l=0$, $x_h=4$</td>
</tr>
<tr>
<td>probability of $x_l$, $p$</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>certain rate of return, $y$</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monitoring cost, $c$</td>
<td>1</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>scale of projects, $k$</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of entrepreneurs, $m$</td>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>optimal deposit rate, $\hat{r}_d$</td>
<td>1.25009</td>
<td>1.25009</td>
<td>1.25002</td>
</tr>
<tr>
<td>optimal loan rate, $\hat{r}$</td>
<td>2.5</td>
<td>2.75</td>
<td>3.5</td>
</tr>
<tr>
<td>optimal number of banks, $n^*$</td>
<td>34</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>probability of bank failure, $G(\hat{r}_d;\hat{r},m/n^*)$</td>
<td>0.00008 -</td>
<td>0.00008</td>
<td>0.00002</td>
</tr>
<tr>
<td>bank profit, $\Pi$</td>
<td>133</td>
<td>128</td>
<td>900</td>
</tr>
<tr>
<td>cost savings compared to direct lending</td>
<td>4516</td>
<td>5652</td>
<td>4502</td>
</tr>
</tbody>
</table>