Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 452

September 1993

LONG-TERM BANKING RELATIONSHIPS IN GENERAL EQUILIBRIUM

Michael S. Gibson

ABSTRACT

I examine the relationship between a financial intermediary ("bank") and a borrowing firm in a three-period overlapping generations model. The model can accomodate two financing arrangements between the bank and the firm: one requires commitment to a long-term contract, the other does not. Which arrangement is chosen depends on whether such a commitment can be credibly made. After defining the two arrangements, I compare their features with real-world financial dealings. Once the form of the long-term relationship between the bank and the firm is set, investment and output of the economy can be determined. Disruptions in financial markets can affect real investment and output by disrupting established long-term relationships.

Long-term Banking Relationships in General Equilibrium

Michael S. Gibson*

1 Long-term banking relationships

1.1 Introduction

When a small business needs to borrow money to finance an investment, it turns to a bank. A bank can provide funds to a firm that cannot issue public securities because investors lack public information on the firm. A small business looking for credit could be offered credit by many banks, but it must choose to deal with one. Once a relationship is established and the bank acquires information about the small business by lending it money, the bank will be able to provide future loans at a lower cost than an outside bank. A long-term relationship between a bank and a firm takes advantage of that lower cost. The primary focus of this paper is how the contracts between the bank and the firm are written and how these contracts affect capital accumulation.

I embed the contracting decision of the bank and the firm in a three-period overlapping generations model, with an investment technology borrowed from Bernanke and Gertler (1989). Both consumers and firms live for three periods. Because a firm lives for three

*The author is a staff economist in the Division of International Finance. This paper represents the views of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or other members of its staff. I would like to thank Roland Benabou, Olivier Blanchard, Gary Engelhardt, Jon Faust, John Montgomery, Don Morgan and seminar participants at the Federal Reserve Board for helpful comments on an earlier draft. I am responsible for any errors.

periods, it makes two investment decisions: one in the first period of its life and one in the second. Firms must borrow to finance investment, and a long-term relationship in the model simply means the firm funds both investments by borrowing from the same bank. The dealings between the bank and the firm take one of two forms, depending on whether the bank is able to commit to obeying a long-term contract. If it can commit, the two parties will agree in advance to the terms of both loans. If the bank cannot commit, they will agree on the terms of the first loan knowing the bank will use its position as insider to make profits on the second loan. Deriving and discussing these two arrangements is the heart of the paper.

The terms of the loan contract determine which firms can invest. Investment determines the next period's capital stock, which determines output and wages. Consumers receive their income, consume and save, yielding a supply of funds for next period's investment. In this manner the general equilibrium of the model is established. Once long-term relationships between banks and firms have been formalized, the importance of bank-specific shocks like bank failure and liquidity crises can be discussed.

The model in this paper establishes that long-term banking relationships are important, but their importance varies across firms. Very productive firms can bypass banks; only middle-productivity firms benefit from the lower borrowing cost a long-term relationship provides. I compare the loan contracts derived in the model with actual borrowing practices and argue that the model's results are consistent with financial arrangements observed in practice.

1.2 Related literature

Bernanke and Gertler (1989) provide a general equilibrium model where a firm's ability to get credit depends on its net worth as well as its productivity. Because theirs is a two-period overlapping-generations model, they do not consider long-term relationships. They focus on the role firms' net worth plays in economic fluctuations. Net worth plays a role because a firm with higher net worth can supply more collateral and lower the agency costs of external finance. Their model, like mine, features costly state verification. I borrow the investment technology from Bernanke and Gertler but eliminate net worth to focus on long-term relationships.

Gertler (1992) considers a problem similar to mine, where a firm borrows to finance investment and project outcomes are private information, but without costly state verification. He solves for the optimal long-term contract, which is contingent on all publicly observable variables. His contract emphasizes the role of collateralizable future profits, which make a long-term contract better than a sequence of short-term contracts by expanding the group of borrowers who invest. In my model firms provide no internal finance, so collateral is not an issue. My paper considers both types of long-term arrangements between the bank and the firm: a long-term contract, like Gertler, and a sequence of short-term contracts.

A long-term relationship is justified in this paper because it reduces agency costs. This is somewhat different from the approach others have taken. Sharpe (1990) assumes a bank learns about the productivity of the firms it deals with. An insider bank knows more about the productivity of the firm than an outsider bank, and the insider bank can expropriate some of the surplus from the relationship because it has a better idea of what the outcome of the firm's investment project will be. Haubrich (1989) claims that a long-term relationship can eliminate monitoring costs altogether; repeated dealings allow the bank to apply something like statistical process control techniques to the firm's output to see whether it is cheating. In an infinite horizon, by a Law of Large Numbers argument, this strategy will costlessly prevent cheating by firms. Haubrich's model features moral hazard rather than costly state verification as the problem banks must deal with. One similarity is that in both models a firm builds up a reputation with its bank that cannot be transferred to another bank; I assume, but do not derive, a similar condition.

Several papers by Douglas Diamond address topics in financial intermediation related to long-term relationships. Diamond (1991a) is the most relevant here; in that paper firms choose whether to finance a long-term investment with long-term debt or rolled-over short-term debt. The firm's decision depends on a tradeoff between the liquidity risk associated with short-term debt (its lender may not agree to roll over its short-term debt) and the lower cost of refinancing short-term debt if favorable information on the firm's type becomes available to the lender. Neither of these issues is present in my model because the firm's type is public knowledge. Diamond's model does not allow for relationships between borrowers and lenders; the effects of such relationships are the focus of my paper.

I explicitly model the long-term financing relationship between a bank and a firm and put it into a general equilibrium model where the effects of financial arrangements on capital accumulation are clear. Some of the authors mentioned above have allowed for long-term relationships and others have looked at the effects of bank lending on capital accumulation, but none has considered the combination. I consider a wide range of firms: those who deal with banks along with those too productive to need banks and those not productive enough to earn bank credit. Banks face stiff competition for the business of high-productivity firms, who do not need banks' monitoring capabilities and can go to decentralized capital markets to finance their investments. The distinction between bank-financed firms and capital market-financed firms is an important one in my model.[1]

2 Setup of the model

2.1 Agents

The model economy in this paper is populated by overlapping generations of agents, who each live for three periods. Each generation has mass one; there is no population growth. All agents have a labor endowment of one in the first period of life, they care only about consumption in the third and final period of life, and they are risk neutral. The decisions of agents in this model are extremely simple: because there is no disutility of work they work when young, they save everything until their last period of life and then they consume all they have.

[1]Diamond (1991b) also makes this distinction. In his model bank finance is less costly than capital market finance for some firms because bank finance mitigates moral hazard.

2.2 Production

The output and investment technology in the model economy is taken from Bernanke and Gertler (1989). Output of the single good is produced using capital and labor according to

$$y_t = f(k_t)$$

where y_t and k_t are output and capital per worker and $f(\cdot)$ is the production function. Output that is not consumed is either used in an investment project (described below) or stored. Storage yields a fixed gross return $R \geq 1$. Initially I assume there is always storage in equilibrium, so the equilibrium interest rate is R. At the end of the paper I will consider an equilibrium where this is not so.

2.3 Firms

Along with each new generation of agents, an equally-sized cohort of firms is born. Firms also live for three periods and are also risk-neutral. Ownership of a cohort of firms is spread evenly across the contemporaneous generation, and shares in the firms are not traded.[2] A "firm" consists of the opportunity to undertake an investment project twice, once when young and once when middle-aged. An investment project converts savings in period t into a random amount of capital in period $t+1$. The amount of savings required to fund a project varies across firms, but the distribution of the random amount of capital produced is the same for all firms and does not depend on firm effort. This implies there will be neither moral hazard nor adverse selection in the model.

An investment project requires input $x(\omega)$ and yields a random amount of capital $\tilde{K}$, distributed continuously on $[\underline{K}, \overline{K}]$ with density function g. The distribution of $\tilde{K}$ is common knowledge and exhibits increasing hazard.[3] Firms are indexed by the parameter ω, uniform

[2]Ownership in this model economy simply means the right to a fraction of the firm's profits, if any. It does not confer any right of control or any liability for covering the firm's losses.

[3]The hazard function $h(K)$ is defined as $h(K) \equiv g(K)/(1 - G(K))$. Increasing hazard ($h'(K) > 0$) is a common assumption in the financial contracting literature and most common distributions (e.g. normal,

on $[0,1]$; a firm with productivity ω requires $x(\omega)$ units of savings to undertake its project. Both ω and $x(\cdot)$ are public information. I assume $x' > 0$, so a firm with higher ω requires more savings to fund its project. Since all projects have the same probability distribution g (although not the same outcome!), a firm with higher ω is less efficient, *ex ante*. A firm with high ω will be referred to as a low productivity firm (and vice versa) throughout the paper. All investment projects are independent draws from the distribution g, so there is no aggregate uncertainty about the total amount of capital produced by any (large) number of investment projects. (It the mean of $\tilde{K}$ times the number of projects.) A firm will operate its project if it can borrow $x(\omega)$ and if by operating it would earn positive expected profits. Any profit realized from the firm's investment is immediately distributed among the agents in the contemporaneous generation, so a firm cannot retain profits to finance future investment.[4]

2.4 Information

The model economy contains an asymmetry of information, because the outcome of a project is known only to the firm undertaking the project. Other agents in the economy are able to verify the outcome of the project, but only by incurring a cost. The information structure in the economy is one of "costly state verification" (Townsend 1979).

If an agent monitors a firm's project outcome, the outcome is known only to that agent.[5] If many agents want to know the result of a single firm's investment project, they must each pay the verification cost. To avoid duplication of monitoring, zero-profit intermediaries ("banks") will emerge to do all the lending and monitoring, as in Diamond (1984). These intermediaries can guarantee a fixed return to depositors while investing in risky projects because they diversify their lending across a large number of firms. Unlike agents and firms, banks will be long-lived. In the period after the loan is made, the capital produced by the investment project is divided between the bank and the firm according to the terms of their

uniform) exhibit increasing hazard. See, for example, Morgan (1993).

[4]Assuming away retained earnings greatly simplifies the exposition of the model. Section 6.2 relaxes the assumption; the fundamental results do not change.

[5]This is in contrast to Bernanke and Gertler (1989), where the results of monitoring are made public.

loan contract.

2.5 Monitoring costs

To provide a motivation for long-term interaction, I assume that monitoring the outcome of a firm's investment project is more costly for a first-time lender than for a repeat lender. Monitoring costs $\gamma_1 + \gamma_2$ the first time a bank lends to a firm, but only γ_1 the second time, because the bank learns something about the firm by dealing with it for a period. Even if no monitoring occurs in the first period of the relationship, the firm still benefits from lower monitoring costs in the second period because the bank has learned something about the firm in dealing with it the first time.[6] To reduce monitoring costs, a firm investing for the second time will borrow from the same bank it borrowed from the first time.

3 Contracts between banks and firms

A firm goes to a bank to get the credit it needs to invest. To each firm that asks for credit, a bank will offer a loan contract, consisting of a loan amount and a repayment schedule. In the most general contract, the firm could repay in either or both periods of its life and its repayment could depend on the outcomes of its projects when young and old. In any single period, a firm will borrow from only one bank; to borrow from more than one would be more expensive because it would require duplication of monitoring. In the remainder of this section, I discuss the features of several contracts, including which firms get credit and at what terms. Then, I compare the results under the various contracts.

[6] One way to justify this structure of monitoring costs, based on Fama (1985), would be to assume firms maintain deposit balances during the first period of a relationship. Holding a firm's deposits gives a bank valuable information that reduces monitoring costs in the future.

3.1 Perfect information

As a benchmark, consider the firm's investment decision under perfect information. This case would apply if the outcome of the firm's investment project were public information, if state verification were costless or if the firm could invest with internal funds. Under perfect information, the firms whose investment projects have positive expected value, net of opportunity cost, would operate. Other firms would not operate and relationships between banks and firms would not matter. The expected amount of capital produced by an investment project is the expected value of $\tilde{K}$, denoted hereafter as K. If the expected price of capital next period is q, the expected value of the project's output is qK. The opportunity cost of undertaking the investment is just the interest lost on the savings invested in the project: $Rx(\omega)$. The surplus is $qK - Rx(\omega)$. I assume the most productive firm's investment is always profitable under perfect information,

$$qK - Rx(0) > 0,$$

and the least productive firm's investment is never profitable under perfect information,

$$qK - Rx(1) < 0,$$

for all relevant q, to rule out a corner solution. All projects with $qK - Rx(\omega) > 0$ would be undertaken under perfect information. If $\breve{\omega}$ is defined by $qK - Rx(\breve{\omega}) = 0$, firms with $\omega \leq \breve{\omega}$ would be able to invest under perfect information and firms with $\omega > \breve{\omega}$ would not.

3.2 An assortment of contracts

I will consider three types of interaction between the bank and the firm in what follows. All three consist of two one-period debt contracts. They differ in the link between the terms of the first and second loans. The first case to consider is the unrealistic one of no link, and is described solely as a benchmark. The second is a long-term *contract*: terms of both loans are fixed before the first-period loan is made. The third is a long-term *relationship*: the firm deals with the same bank when young and when old, but the terms of its second-period loan are not determined until the second loan is made. Before proceeding I summarize the one-period debt contract, the building block upon which all three contracts are based.[7]

Under a debt contract, the firm repays a constant amount F to the bank, unless its project has a bad outcome and yields less than F, in which case the bank pays to monitor the firm's project outcome and takes everything the firm produced.[8] A firm is only monitored if its investment turns out badly. The bank's expected return under a standard debt contract is

$$ER(F;q) \equiv q\left[\int_{\underline{K}}^{F}(K-\gamma)g(K)dK + \int_{F}^{\overline{K}} F\,g(K)dK\right] \tag{1}$$

where q is the price of capital next period, and γ is the monitoring cost. For the bank to earn zero expected profit net of opportunity cost, the bank's expected return must equal its opportunity cost:

$$ER(F;q) = Rx(\omega) \tag{2}$$

where R is the interest rate and $x(\omega)$ is the amount a firm of productivity ω must borrow to operate its project. The bank's expected repayment increases with F, the contractual repayment, up to an interior maximum at $\overline{F} \in (\underline{K}, \overline{K})$. Above $\overline{F}$, the bank's expected repayment falls, so no contract will call for a repayment greater than $\overline{F}$. $\overline{F}$ is defined by the first-order condition that results from maximizing $ER(F;q)$ with respect to F:

$$-\gamma g(\overline{F}) + (1 - g(\overline{F})) = 0$$

[7]This is the contract first derived by Townsend (1979) which has since become a standard in the financial intermediation literature. I focus on the case of non-random monitoring. See Townsend (1979) and Mookherjee and Png (1989) for discussion of the merits of random monitoring.

[8]The contractual interest rate i is related to the contractual repayment F by

$$(1+i)\,x(\omega) = F.$$

or

$$\frac{g(\overline{F})}{1-G(\overline{F})}=\frac{1}{\gamma} \tag{3}$$

Because $\overline{F}$ is a function of γ, it will be written as $\overline{F}(\gamma)$ from now on. A firm with high enough ω will not be granted a loan because the opportunity cost of funds exceeds the maximum return the bank can earn under a debt contract ($Rx(\omega) > ER(\overline{F};q)$). For future reference, notice that the expected repayment and the cutoff $\overline{F}$ both are lower when the monitoring cost γ is higher.

3.3 Myopic long-term relationship

As a benchmark against which to measure the contracts described in the next two sections of the paper, this section describes a myopic long-term relationship. The adjective "myopic" applies because the parties make a loan arrangement when the firm is young without looking ahead to the next period and realizing the firm will also want credit when old. In this scenario, in both the first and second periods of its life a firm obtains a bank loan at the repayment that earns the bank zero profits in that period.

The repayments F^1 and F^2 are determined by

$$q_{t+1}\left[\int_{\underline{K}}^{F^1}(K-(\gamma_1+\gamma_2))g(K)dK+\int_{F^1}^{\overline{K}}F^1\,g(K)dK\right]=Rx(\omega) \tag{4}$$

and

$$q_{t+2}\left[\int_{\underline{K}}^{F^2}(K-\gamma_1)g(K)dK+\int_{F^2}^{\overline{K}}F^2\,g(K)dK\right]=Rx(\omega), \tag{5}$$

which are just the one-period zero profit condition (2) rewritten for the two periods in which a firm born at time t is alive. Neither party takes account of the long-run nature of their interaction.

The contract will be feasible for firms with ω low enough so that in both periods a repayment exists that earns the bank zero profit ($F^1 \leq \overline{F}(\gamma_1+\gamma_2)$ and $F^2 \leq \overline{F}(\gamma_1)$). Since monitoring costs are higher in the first period and $ER(\overline{F}(\gamma_1+\gamma_2)) < ER(\overline{F}(\gamma_1)$, the first period's loan will break down first.[9] All firms with $\omega < \omega^*$ will get credit, where ω^* is defined by

$$q_{t+1}\left[\int_{\underline{K}}^{\overline{F}(\gamma_1+\gamma_2)}(K-(\gamma_1+\gamma_2))g(K)dK+\int_{\overline{F}(\gamma_1+\gamma_2)}^{\overline{K}}\overline{F}(\gamma_1+\gamma_2)\,g(K)dK\right]$$
$$= Rx(\omega^*)$$

The cutoff ω^* marks the firms that would receive credit and invest if banks ignored the fact that a firm lives for two periods.

3.4 Long-term contract

3.4.1 What is a long-term contract?

The second of the three contracts to be analyzed is a long-term contract consisting of two one-period debt contracts whose repayment amounts are agreed upon in advance. A sequence of short-term contracts where the terms are agreed upon in advance is one type of long-term contract, but it will not necessarily be the optimal long-term contract. The optimal long-term contract could take a more complex form.

Research by Chang (1990) and Webb (1992) describes the optimal long-term contract in a two-period model with costly state verification. Though neither paper's result is directly applicable here, by using their results I can describe the form of the optimal contract. The second period of the contract would be a standard debt contract, with a fixed repayment amount and monitoring for outcomes below that amount. The fixed repayment amount would vary inversely with the reported first-period outcome, providing an incentive for a good first-period outcome to be reported truthfully. The range of outcomes where monitoring occurs can be reduced or eliminated. Such a contract is quite complicated to write down and solve; its complex structure makes it hard to compare with the long-term relationship defined in section 3.5 below.

Also, because payments in non-monitoring states are state-contingent in the first period,

[9] If $q_{t+1} = q_{t+2}$, as it will in the stationary equilibrium.

this arrangement looks more like a closed-end equity fund than bank debt. In this paper I want to focus on long-term *bank* relationships. If I restrict contracts to have fixed payments in states without monitoring, ruling out all equity-like financing arrangements, the long-term contract analyzed below is the optimal long-term contract.[10] For these reasons I leave the question of the optimal long-term contract and focus on the long-term contract made up of two one-period debt contracts whose repayments are both agreed upon in advance.

3.4.2 Terms of the contract

We can single out the most productive firms, those who can repay what they borrow in every state of nature. Such a firm will never be monitored, so the lower second-period monitoring cost is immaterial. Since lending is channeled through banks to avoid duplication of monitoring, if monitoring is certain not to occur firms can borrow directly from a group of savers, bypassing banks altogether. The repayment paid by the firm satisfies $q_{t+1}F^1 = q_{t+2}F^2 = Rx(\omega)$, which is just the lender's zero-profit condition. These firms will have $\omega \leq \underline{\omega}$, where $\underline{\omega}$ is defined by $q\underline{K} = Rx(\underline{\omega})$. In what follows the loan contracts of the remaining firms, with $\omega > \underline{\omega}$, are considered.

If, when they initiate a two-period relationship, the bank and firm agree to repayments and loan amounts for both periods of the relationship, the following contract would emerge. Because banking is competitive, the bank earns no expected profits over the life of the relationship, though it could earn positive profit in one period, balanced by a loss in the other period. The best long-term contract would choose F^1 and F^2 to minimize monitoring costs,

$$q_{t+1}(\gamma_1 + \gamma_2)G(F^1) + q_{t+2}\,\gamma_1\,G(F^2) \tag{6}$$

while earning zero profits for the bank:

$$q_{t+1}\left[\int_{\underline{K}}^{F^1}(K - (\gamma_1 + \gamma_2))g(K)dK + \int_{F^1}^{\overline{K}} F^1 g(K)dK\right] \tag{7}$$

[10]This result is a straightforward application of section II of Webb (1992).

$$+ q_{t+2}\left[\int_{\underline{K}}^{F^2}(K-\gamma_1)g(K)dK + \int_{F^2}^{\overline{K}} F^2 g(K)dK\right] = 2\,Rx(\omega).$$

To minimize (6) subject to (7), form a Lagrangian and take first-order conditions.[11] The first-order conditions are

$$(\gamma_1+\gamma_2)g(F^1)(1-\lambda)+\lambda(1-G(F^1))=0 \tag{8}$$

$$\gamma_1\, g(F^2)(1-\lambda)+\lambda(1-G(F^2))=0 \tag{9}$$

where λ is the Lagrange multiplier on (7). Combining (8) and (9) gives

$$\frac{h(F^1)}{h(F^2)}=\frac{\gamma_1}{\gamma_1+\gamma_2} \tag{10}$$

where h is the hazard function associated with the density g. Since h is increasing by assumption, $F^1 < F^2$. The bank charges the firm less in the first period of the relationship than in the second because monitoring costs are higher in the first period, and the bank and firm commit to work together to reduce monitoring costs. This motivation for a low introductory rate is reasonable, yet it is very different from what emerges in the next section. Combining (10) with the bank's zero profit condition yields the solutions for F^1 and F^2. See Figure 1, where the first-order condition (10) is labeled "FOC" and the bank's outside return constraint (7) is labelled "ORC." The intersection, point B, reveals the optimal choice of F^1 and F^2. (Ignore points A, C1, and C2 for now.)

The repayments F^1 and F^2 depend on ω, as we see from (7), the bank's zero profit condition. As ω rises, the expected repayment to the bank increases to cover the opportunity cost of more borrowed funds. In Figure 1, ORC shifts away from the origin as ω increases, and F^1 and F^2 will both increase. At some level of ω, it becomes impossible for the bank to recoup its investment; once $F^1=\overline{F}(\gamma_1+\gamma_2)$ and $F^2=\overline{F}(\gamma_1)$ the expected repayment to

[11] A complete statement of the minimization would include two inequality constraints on F^1 and F^2, that they both be greater than or equal to $\underline{K}$. These are omitted in the text for expositional simplicity. The absence of discounting in (6) and (7) makes no difference for the first-order conditions (8) and (9).

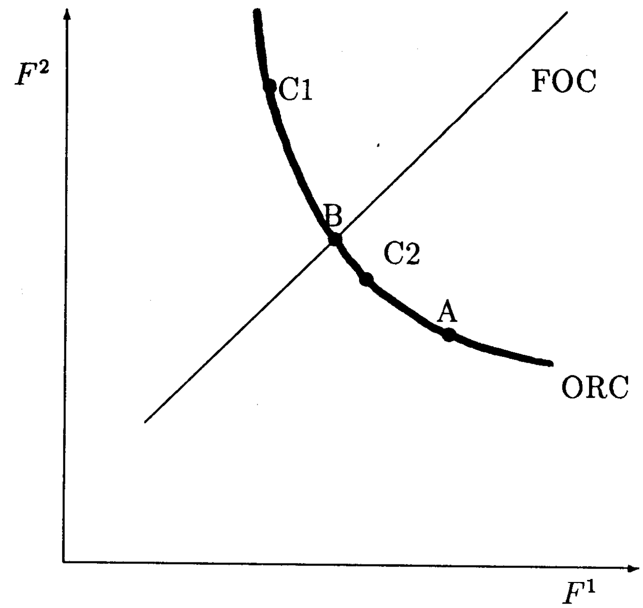

Figure 1: Optimal repayments in the three contracts

the bank is at its maximum.[12] Above this level of ω, the contract breaks down. Call this level $\hat{\omega}$.

Firms with $\omega \leq \hat{\omega}$ will receive an offer of credit from a bank. Which firms accept the offer and choose to operate? Any firm with positive expected profit will do so. Let

$$\mathcal{F}(F^1, F^2) = q_{t+1}\left[\int_{F^1}^{\overline{K}}(K - F^1)g(K)dK\right] + q_{t+2}\left[\int_{F^2}^{\overline{K}}(K - F^2)g(K)dK\right]$$

be the expected profit of the firm. Firm profits are always positive for any firm that operates its investment project, because the firm risks none of its own funds and the repayments F^1 and F^2 will always be strictly less than $\overline{K}$. Note that firm profits fall as ω rises because F^1 and F^2 both increase. Still, $\mathcal{F} > 0$, so every firm with $\omega \leq \hat{\omega}$ will ask for and receive credit from a bank in both periods of its life.

When operating under a long-term contract, the bank and firm agree up front what the present and future repayments should be. However, if no binding commitment mechanism exists, the bank could not commit to a second period repayment. Once the second period of

[12]To see that $F^1 = \overline{F}(\gamma_1 + \gamma_2)$ and $F^2 = \overline{F}(\gamma_1)$ simultaneously, combine (10) with (3).

the relationship begins, the bank has some market power that it can exploit, and it will no longer be optimal for the bank to go along with the agreement it made.[13] I now turn to the contract that would emerge in that case.

3.5 Long-term relationship

A different contract will be written if the second period loan is not contractible in the first period of the relationship. The bank that supplied credit to the firm in the first period (the "insider" bank) will be able to supply it with credit in the second period at a lower cost, because the information it gathered by dealing with the firm in the first period allows it to monitor the firm more cheaply, should monitoring be necessary. In the second period of the relationship, the bank will exploit the market power its informational advantage over the other banks gives it to earn some profit. In the first period, the bank and the young firm will negotiate a loan, both aware the bank will have an advantage over other banks once the relationship has been established. The bank is unable to commit not to exercise its market power, as it had to be able to do to implement the long-term contract.

To find the details of the sequence of short-term contracts, I work backwards from the second loan, as is usual in such problems. First, I find the profit-maximizing repayment charged by the insider bank in the second period. The bank will make a profit off the second loan. Going back to the first period of the relationship, the bank will offer a repayment at which it loses enough money on the first loan that the total profit from the relationship, first and second loans together, is zero, as it must be since banking is competitive.

3.5.1 Terms of the second loan

Recall that firms with $\omega \leq \underline{\omega}$ are never monitored because they borrow so little they are able to repay in every state of nature. Monitoring costs are irrelevant for these firms, and the insider bank has no advantage over other banks when competing for these firms' second-period business.

[13] I ignore reputation effects, which could be important here. For a discussion of reputation, see footnote 19.

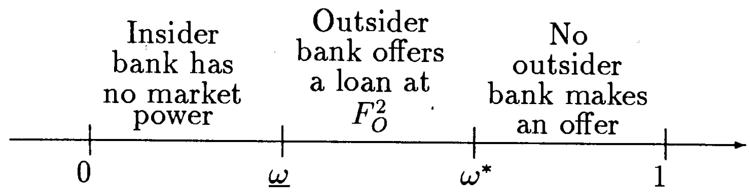

Figure 2: Second period offers

For firms with $\omega > \underline{\omega}$, expected monitoring costs are positive so the cost advantage of the insider bank is relevant. In the second period of the relationship, an outsider bank offers to lend $x(\omega)$ to the firm for one period with a repayment that yields it zero profit. Call this repayment F_O^2; it satisfies

$$q_{t+2}\left[\int_{\underline{K}}^{F_O^2}(K-(\gamma_1+\gamma_2))g(K)dK + \int_{F_O^2}^{\overline{K}} F_O^2\, g(K)dK\right] = Rx(\omega). \tag{11}$$

For ω above some cutoff ω^*, this equation has no solution (because at ω^*, $F_O^2 = \overline{F}(\gamma_1+\gamma_2)$). Firms less productive than ω^* are not able to get a second-period loan from an outsider bank because the outsider bank can not charge a repayment high enough to cover the cost of making the loan.

Firms with $\omega > \underline{\omega}$ can be divided into two groups: those who receive an offer from an outsider bank ($\omega \in (\underline{\omega}, \omega^*]$) and those who do not ($\omega > \omega^*$). See Figure 2. The insider bank treats the two groups differently. For firms with an outside offer, the insider bank can retain the customer by charging an repayment less than or equal to F_O^2. Because it maximizes profits, it always charges exactly F_O^2. The insider bank limit prices against the outsider bank's offer. Let $\mathcal{F}_2$ denote the firm's expected profits in the second period. Those profits are

$$\begin{aligned}\mathcal{F}_2(F_O^2) &= q_{t+2}\int_{F_O^2}^{\overline{K}}(K - F_O^2)g(K)dK \\ &= q_{t+2}K - Rx(\omega) - q_{t+2}(\gamma_1+\gamma_2)G(F_O^2).\end{aligned}$$

The bank's expected second-period profits ($\mathcal{B}_2$) are

$$\mathcal{B}_2(F_O^2) = q_{t+2}\,\gamma_2\,G(F_O^2).$$

Looking back to (11), note that $dF_O^2/d\omega > 0$, which implies $d\mathcal{B}_2/d\omega > 0$ and $d\mathcal{F}_2/d\omega < 0$.

For firms without an outside offer, the insider bank acts as a monopolist and charges its revenue-maximizing repayment $\overline{F}(\gamma_1)$. The bank is the residual claimant in this region. The firm is left with expected profits of

$$\mathcal{F}_2(\overline{F}(\gamma_1)) = q_{t+2}\int_{\overline{F}(\gamma_1)}^{\overline{K}} (K - \overline{F}(\gamma_1))g(K)dK$$

while the bank, in expectation, earns

$$\mathcal{B}_2(\overline{F}(\gamma_1)) = q_{t+2}K - Rx(\omega) - q_{t+2}\,\gamma_1\,G(\overline{F}(\gamma_1)) - \mathcal{F}_2(\overline{F}(\gamma_1)).$$

In this region, $d\mathcal{B}_2/d\omega < 0$ and $d\mathcal{F}_2/d\omega = 0$. For ω large enough, the bank's second-period expected profit becomes negative and the bank does not make a second-period loan. Call this cutoff $\overline{\omega}$.

To summarize, in the second period of a relationship, four things could happen (see Figure 3):

$\omega \leq \underline{\omega}$ The firm borrows so little that it repays the loan in every state of nature. Monitoring never occurs so the insider bank has no advantage over other banks. The insider bank can make no second-period profits: $\mathcal{B}_2 = 0$. The repayment F^2 satisfies a zero-profit condition for the bank: $qF^2 = Rx(\omega)$.

$\underline{\omega} < \omega \leq \omega^*$ The firm borrows enough that expected monitoring costs are positive, but not so much that no outsider bank makes it an offer. An outsider bank offers the firm a loan that leaves the outsider bank with zero profits and the firm with positive profits. The insider bank limit prices against the outsider bank's offer, giving the firm the same profit and keeping the savings in reduced monitoring costs for itself. $\mathcal{B}_2 > 0$. $F^2 = F_O^2$.

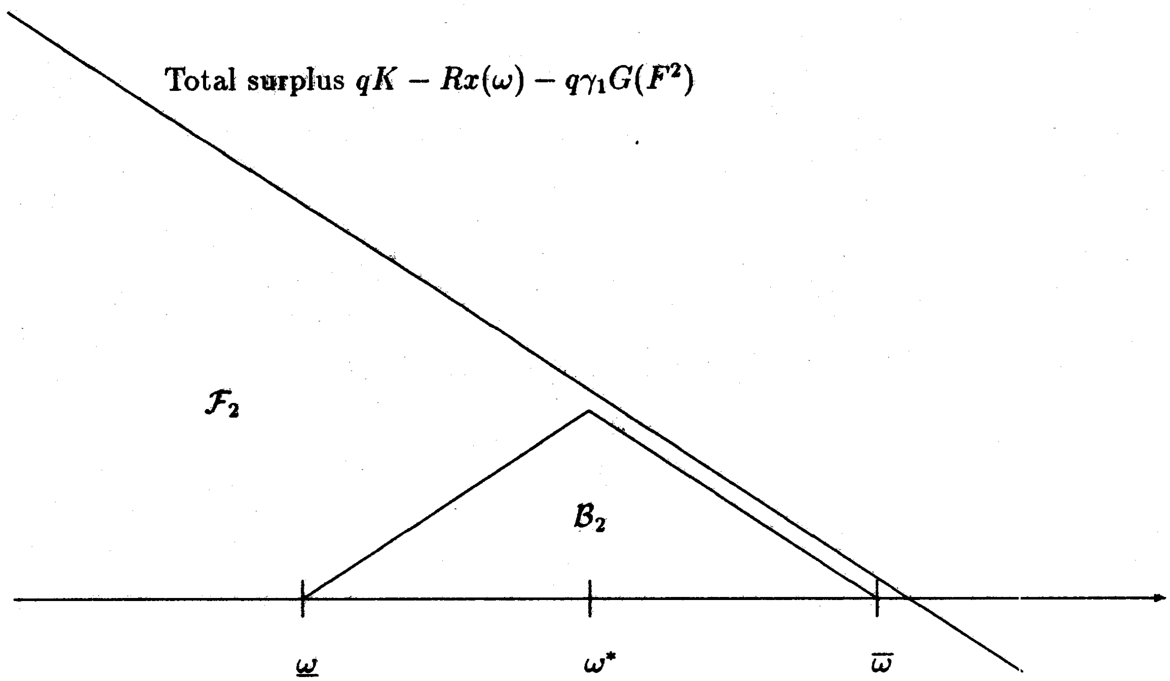

Figure 3: Second period profits

$\omega^* < \omega \leq \overline{\omega}$ The firm borrows so much that expected monitoring costs are high enough to make a potential relationship with an outsider bank unprofitable for the bank. The insider bank is not constrained by outside competition and extracts all the surplus it can from the relationship. The firm is left with some profit since $\overline{F}(\gamma_1) < \overline{K}$. $\mathcal{B}_2 > 0$. $F^2 = \overline{F}(\gamma_1)$.

$\omega > \overline{\omega}$ The firm needs to borrow so much that even a loan from an insider bank would be unprofitable. The firm cannot get credit and does not operate in the second period. $\mathcal{B}_2 = 0$.

I have characterized the second-period loan contract between a firm of any productivity and its insider bank. Now I turn to the first period of the relationship.

3.5.2 Terms of the first-period loan

When a bank and a young firm meet to negotiate a loan, both know what will happen when the firm is old and a "captive customer" of the bank. In other words, $\mathcal{B}_2$ is known. It only depends on ω, which is public information. The bank offers the young firm a repayment F^1 that ensures the bank earns zero profit over the course of the relationship ($\mathcal{B}_1 + \mathcal{B}_2 = 0$).[14] If $\mathcal{B}_2 > 0$, this implies $\mathcal{B}_1 < 0$, so the bank makes a loss on its first period loan. F^1 is determined by

$$\mathcal{B}_1 + \mathcal{B}_2 = 0$$

or

$$q_{t+1}\left[\int_{\underline{K}}^{F^1}(K-(\gamma_1+\gamma_2))g(K)dK + \int_{F^1}^{\overline{K}} F^1 g(K)dK\right] - Rx(\omega) + \mathcal{B}_2 = 0 \tag{12}$$

As ω increases, $x(\omega)$ increases and $\mathcal{B}_2$ changes, first increasing because outside competition becomes weaker and then decreasing because the bank is taking almost all of the surplus from the relationship and the surplus is falling (see Figure 3). At some level of ω between ω^* and $\overline{\omega}$, the bank becomes unable to earn zero profit on its funds; (12) will have no solution because $F^1 > \overline{F}(\gamma_1 + \gamma_2)$. That level of ω will be $\hat{\omega}$. We know that $\hat{\omega}$ lies between ω^* and $\overline{\omega}$ because at ω^* the bank would earn zero first-period profits by charging $\overline{F}(\gamma_1 + \gamma_2)$, so it can easily earn $-\mathcal{B}_2$ by charging less. At $\underline{\omega}$, the bank loses money in the first period even if it charges $\overline{F}(\gamma_1 + \gamma_2)$, and it has no second period profits to offset the loss. Somewhere in between lies the ω where $\mathcal{B}_1(\overline{F}(\gamma_1 + \gamma_2)) + \mathcal{B}_2(\overline{F}(\gamma_1)) = 0$.

A firm with $\omega \leq \hat{\omega}$ will invest in both periods of its life. This cutoff is identical to that in the long-term contract, so the same group of firms receives credit in both financial arrangements. The marginal firm ($\omega = \hat{\omega}$) must be charged $F^1 = \overline{F}(\gamma_1 + \gamma_2)$ and $F^2 = \overline{F}(\gamma_1)$ for the bank's outside return constraint to be satisfied, regardless of whether a long-term relationship or a long-term contract is in force. Comparing these two arrangements with the perfect information case, fewer firms get credit ($\hat{\omega} < \check{\omega}$) because the costs of monitoring make some firms unprofitable.

Figure 3 reveals that middle-productivity firms gain the most from a long-term relationship. Firms with productivity near ω^* yield the most second-period profit to the bank, and all second-period bank profits are transferred to the firm via a low first-period repayment.

[14] For simplicity, I again assume neither the firm nor the bank discounts the future.

Middle-productivity firms need a bank because some single agent must be prepared to monitor their investments on behalf of savers. Banks play that role here. Very productive firms ($\omega < \underline{\omega}$) gain nothing from a long-term relationship because they can finance their investments without agency costs. The publicly available information on these firms (ω and $\underline{K}$) is sufficient to guarantee lenders a competitive return with no chance of monitoring, eliminating the need for intermediation entirely. Very unproductive firms ($\omega > \overline{\omega}$) gain nothing because even though their investments may have positive social value ignoring monitoring costs (if $\omega \in (\overline{\omega}, \check{\omega})$), the monitoring is essential to produce enough information about the investment outcome to elicit the necessary credit. Even the lower monitoring cost of a two-period relationship cannot make these firms' investments profitable. Internal finance, through firm equity, would help all firms and would expand the range of firms that can operate by reducing the need for external funds which are accompanied by a risk of monitoring.[15]

3.6 Interpretation : Compare and contrast the three

I have discussed three possible arrangements between a bank and a firm in this model economy: a myopic long-term relationship, a long-term contract, and a long-term relationship. Each would provide a firm with a different pair of repayments and a different level of profit. The same group of firms (those with $\omega \leq \hat{\omega}$) will get credit under the last two arrangements; fewer firms would get credit if the parties were myopic.[16]

We can first compare the myopic long-term relationship with the long-term contract. In Figure 1, point A represents the repayment pair of the former and point B the latter. Both lie on the ORC curve, since the bank makes zero profit in both cases. The long-term contract conserves on monitoring costs by concentrating more of the monitoring in the second period, where it is less costly. The myopic long-term relationship does the opposite because the bank breaks even in each individual period. In a one-period debt contract, higher monitoring costs require a higher repayment. Monitoring costs more in the first period, so

[15] This is the main point of Bernanke and Gertler (1989). See also section 6.2 below.

[16] If firms could finance their investments with inside funds, as they do in Bernanke and Gertler (1989), the set of firms that get credit under the long-term relationship and the long-term contract could be different.

the first-period repayment exceeds the second-period repayment under a myopic long-term relationship. More monitoring occurs in the first period, where it is relatively expensive. Not only is monitoring more expensive, making investment less productive, but fewer firms are able to invest because the cutoff for getting credit is lower ($\tilde{\omega} < \hat{\omega}$) under a myopic long-term relationship. More capital is produced and firm profits are larger in an economy with long-term contracts.

The more important comparison is between the long-term contract and the long-term relationship. In Figure 1, these are points B and C1 or C2. Points C1 and C2, either of which can represent the long-term relationship, must lie above the 45 degree line, since the bank opportunistically charges a high repayment in the second period. However, point C can lie on either side of the FOC line (represented by C1 and C2). For low levels of ω, the bank will not have much market power in the second period and F^1 and F^2 will be nearly equal and close to $\underline{K}$. F^1 will always be lower than F^2 (if $\omega > \underline{\omega}$) so the bank can make a first-period loss to counter its second-period profit. The amount of capital wasted by monitoring will be greater at C than at B, simply because B minimizes monitoring costs. Firm profits will be higher and more capital will be produced under a long-term contract than under a long-term relationship despite the fact that the same group of firms operate in both cases. Financial intermediation is more efficient if banks and firms write long-term contracts; long-term relationships result in more resources being devoted to monitoring, because the interests of the bank diverge from the interests of the firm and the bank cannot commit not to act in its own interest.

Whether the long-term contract or the long-term relationship emerges depends on whether the second-period repayment is contractible in the first period the bank and the firm deal with each other. In the very simple model economy described so far, nothing prevents the bank and the firm from contracting on the second period repayment. There is no aggregate uncertainty in the model, so the parties can compute their expected return from any future repayment. If the model economy were made more complex, the second-period interest rate could become not contractible. In the contracts literature, one reason for a variable to be considered not contractible is if economic conditions are expected to change in such an unpredictable way that agents cannot form an expectation with which to compute their payoffs. Or, it could be too costly to compute such an expectation. In the present model, the uncertainty of the real world has been vastly simplified. If more real-world complexity were present in the model, the second-period interest rate might be not contractible. The model in this paper implies the long-term contract will be chosen if uncertainty is well-defined and well-understood; a more complex model might lead to the opposite deduction.

In a more realistic model, some firms could be offered a long-term contract while other firms are not. Some firms might have investment projects whose risk characteristics are well-known or easily understood with a minimum of specialized knowledge; these firms could be offered a long-term contract. Firms with innovative, uncommon, difficult to understand investment projects might find that the costs of writing a long-term contract exceed the benefits; these firms might obtain financing through a long-term relationship.

A second way to establish which is the more relevant contract is to look at real-world financial arrangements to see whether they more closely resemble the long-term contract or the long-term relationship. Banks lend money primarily for short maturities. According to the Survey of Terms of Bank Lending conducted by the Board of Governors of the Federal Reserve System, the weighted average maturity of commercial and industrial loans made by commercial banks during the week August 3–7, 1992 was 185 days, or just over six months.[17] Banks also lend primarily under loan commitments. According to the same survey, commercial banks made 66 percent of short term and 65 percent of long term commercial and industrial loans under commitment. A firm that expects to be in business for ten years and requires bank credit will borrow what it needs and roll over its loan every six months. If the firm and its bank agree on the terms of future loans at the time of the first loan, perhaps by signing a ten-year loan commitment, they are effectively mimicking the long-term contract of the model. If they have no loan commitment, or the loan commitment they have lasts for less than the life of the firm, they are closer to the long-term relationship. The latter seems

[17] *Federal Reserve Bulletin*, November 1992, p. A76.

closer to the popular conception of how U.S. commercial lending works, though insufficient data on the terms of real-world lending arrangements to be able to say for sure.

The long-term contract and the long-term relationship both feature a low first-period repayment, but for different reasons. The bank charges a low repayment when initiating a long-term contract because monitoring is more expensive in the first period and both parties strive to minimize monitoring costs. When a bank begins a long-term relationship, it charges a low repayment to capture the firm as a customer; it knows it will obtain some market power and can earn some rents in the future.[18] Both the long-term contract and the long-term relationship contrast with the myopic long-term relationship, where the repayment is high in the first period and low subsequently.

The choice between a long-term contract and a long-term relationship comes about because of a conflict of interest between the bank and the firm. If the bank acts solely according to self-interest, it will charge as high a repayment as it can in the second period of the relationship. If the bank and the firm work together to maximize joint profit, the long-term contract would be feasible.[19] This distinction suggests an international comparison of banking practices. In Anglo-Saxon countries (the United States and United Kingdom), ownership of banks and firms is separated by law. In Europe and Japan, banks have close ties to the firms they lend to, often including an equity stake.[20] The latter arrangement, called universal banking, helps ensure the bank will not act out of pure self-interest, because it reaps some of the benefits of its customers' profits. The financial intermediation sector in universal banking countries may be closer to long-term contracts, while intermediation in Anglo-Saxon countries may be closer to long-term relationships.[21] Of course, as discussed above, this has a

[18] This is the motivation for a low introductory price typically found in models of repeated interaction with relationship-specific investment, such as Farrell and Shapiro (1989) and Sharpe (1990).

[19] The bank's eagerness to gouge the firm in the second period obviously ignores considerations of reputation. Sharpe (1990) discusses the effects of reputation. He finds that reputation can mitigate the problems of bank opportunism in the second period of a relationship. In his model with reputation, the outcome under a long-term relationship can duplicate the outcome under a long-term contract if certain technical conditions hold.

[20] See Bisignano (1990) for a comparison of the two systems.

[21] If the bank and the firm were one and the same, the need for costly state verification would be eliminated as the bank could provide costless "internal" funds to finance the firm's investment.

direct implication for the efficiency of financial intermediation in the two groups of countries. Universal banking would be more efficient, if the conclusions of the model could be applied this way, because there would be less dead-weight loss from the monitoring of borrowers by lenders.

4 General equilibrium

Given an agreement between banks and firms on the terms of financing, investment and capital accumulation can be determined. The next step is to calculate the equilibrium over time, as successive cohorts of firms and agents are born, live, and die. The general equilibrium of the model is determined exactly as in Bernanke and Gertler (1989). At the beginning of period t, the capital stock k_t is predetermined (by last period's investment decisions). The capital stock determines output and the wage. Firms that invested in $t-1$ sell whatever capital their investments netted them and redistribute profits to their owners. Saving is equal to the wages of the young plus the saving of the middle aged. I assume for now that saving is always adequate to fund all projects with positive social value, so rationing never occurs. Investment remains to be determined.

Under either long-term contracts or long-term relationships, every firm with $\omega < \hat{\omega}$ will invest; the cutoff level $\hat{\omega}$ depends positively on q_{t+1} and q_{t+2}, so $\hat{\omega}$ can vary with t (and will be written as $\hat{\omega}_t$ where necessary to avoid confusion). Firms born in t with $\omega \leq \hat{\omega}_t$ will invest in period t; so will firms born in $t-1$ with $\omega \leq \hat{\omega}_{t-1}$. Supply of next period's capital stock k_{t+1} is equal to the sum of capital produced by old firms and capital produced by young firms.

The investments of young firms in period t yield

$$\int_0^{\hat{\omega}_t} (K - (\gamma_1 + \gamma_2)G(F^1(\omega)))\, d\omega$$

units of capital in $t+1$; those of old firms yield

$$\int_0^{\hat{\omega}_{t-1}} (K - \gamma_1 G(F^2(\omega)))\, d\omega$$

and the repayments F^1 and F^2 have been written explicitly as functions of ω as a reminder that less productive firms pay higher repayments, and so are more likely to be monitored. The repayments F^1 and F^2 will of course depend on the type of interaction between the bank and the firm: long-term contract or long-term relationship.

Per capita capital formation is

$$k_{t+1} = K\hat{\omega}_t - \int_0^{\hat{\omega}_t} (\gamma_1 + \gamma_2) G(F^1(\omega))\, d\omega \tag{13}$$

$$+K\hat{\omega}_{t-1} - \int_0^{\hat{\omega}_{t-1}} \gamma_1\, G(F^2(\omega))\, d\omega;$$

(13) is the capital supply function. Four of the variables on the right-hand side vary with q_{t+1}: $\hat{\omega}_t$ and $\hat{\omega}_{t-1}$ increase with q_{t+1}, because a higher price for capital will make more firms creditworthy; F^1 and F^2 decrease with q_{t+1}, because a bank needs to charge less if the capital with which it is repaid is worth more. Differentiate (13) with respect to q_{t+1} to see that $dk_{t+1}/dq_{t+1} > 0$, as expected for a supply curve.

Demand for next period's capital stock is determined by setting its price equal to its marginal product:

$$q_{t+1} = f'(k_{t+1}) \tag{14}$$

Combining (13) and (14) yields the within-period equilibrium.

Within a period, equilibrium is determined by the intersection of capital supply and demand. However, the position of the capital supply curve depends on q_{t+2}, because the number of young firms able to obtain credit will change with the price of the capital they produce when old. In this sense the equilibrium of the model is forward-looking. The equilibrium is also backward-looking, as the number of old firms investing in period t will depend on how many old firms got credit in period $t-1$, which depended on q_t. The equilibrium path of q_t over time is the solution to the second-order non-linear difference equation produced by substituting (13) into (14):

$$q_{t+1} = f'\left(K\hat{\omega}_t - \int_0^{\hat{\omega}_t} (\gamma_1 + \gamma_2) G(F^1(\omega))\, d\omega \right. \tag{15}$$

$$\left. + K\hat{\omega}_{t-1} - \int_0^{\hat{\omega}_{t-1}} \gamma_1\, G(F^2(\omega))\, d\omega \right),$$

where $\hat{\omega}_t$, $\hat{\omega}_{t-1}$, F^1 and F^2 are implicitly functions of q at different times.

One solution to (15) is a stationary solution where the price of capital and the capital stock are both constant over time. To see that a stationary solution exists and is unique, note that $dk/dq > 0$ along the capital supply curve and $dk/dq < 0$ along the capital demand curve.[22] In the stationary solution, a constant fraction of each generation of firms gets credit and invests. The variables q, k and $\hat{\omega}$ do not vary over time; neither do the repayments $F^1(\omega)$ and $F^2(\omega)$.

Any additional equilibrium paths of q_t will have the following feature: a time period with expensive, scarce capital will be followed by a period with cheap, abundant capital. To see this, look at Figure 4. The central capital supply curve S_2 represents the stationary equilibrium where $q_{t+2} = q_{t+1}$. Along the upper capital supply curve (S_1), the same price of capital in period $t+1$ brings forth less investment than in the stationary equilibrium; for this to be the case, q_{t+2} must be lower than in the stationary equilibrium, meaning fewer firms can finance investment. By the same logic, along S_3 q_{t+2} must be higher than in the stationary equilibrium. Any equilibrium path of q_t that differs from the stationary equilibrium must have q_t alternating between high and low values. Anytime that $q_{t+1} > q$ (such as at the intersection of S_1 and D), q_{t+2} must be less than q and vice versa. Such oscillating equilibria cannot be ruled out, though they do not have to be present.

[22] To prove $dk/dq > 0$, we must deal with the fact that dF^1/dq_{t+2} will be positive for firms with ω just below ω^* because one effect of higher q is tougher outside competition for the bank in the second period, lower second period profits, and a higher first period repayment. For other levels of ω this effect will be dominated by the usual effect: higher q means a lower repayment is needed for the bank to cover its cost of funds. It is reasonable to assume the sum over ω of all these effects is negative so that $dk/dq > 0$.

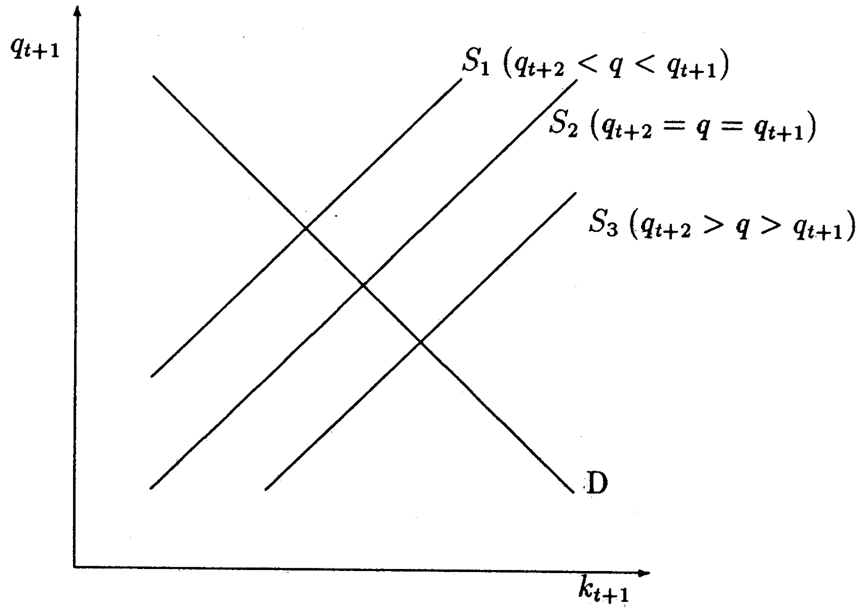

Figure 4: Oscillating equilibrium

5 Applications of the model

5.1 Limited availability of savings

Two interesting applications of the model emerge if I relax the assumption that the supply of savings is ample enough to fund every profitable investment project. I originally made the assumption to ensure that storage was always positive, tying down the equilibrium interest rate at R and making it easier to analyze the bank's lending decision. If storage were zero and banks could not lend to all the firms whose projects are creditworthy, the banks would have to ration the available credit among firms. Once rationing occurs, in exchange for the added complexity of worrying about the equilibrium repayment, I can tell a richer story of credit allocation that includes determining how banks allocate a limited supply of credit between prospective borrowers.

Consider first the case where a single bank has difficulty raising funds to lend to its customers. In the model economy as stated so far, there would be no reason for this to happen since all banks provide identical riskless opportunities for savers. Two dimensions along which banks with identical loan portfolios could differ, in reality, are capital adequacy and the honesty of their management. Banks with low capital and banks which are suspected of having poor quality management often have trouble raising funds in both wholesale and retail markets. A bank could have trouble raising funds in two ways: be forced to pay a risk premium or simply be quantity-rationed. I will analyze the second type below, but the first type would have broadly similar effects.

If a bank had only a limited supply of funds available and was forced to choose what loans to make, it will always serve its old customers first, under both a long-term contract and a long-term relationship. Under a long-term contract, the bank will fulfill its obligations under existing contracts by supplying funds to old customers before it takes on new obligations by signing contracts with new customers. Under a long-term relationship, old customers yield positive profits for the bank while new customers yield no profit; banks will always lend to old customers first.

If a bank operating with long-term relationships had to sever ties with some old borrowers, it would prefer to keep the firms whose loans provide it with the largest second-period profit. These will not be the same firms the bank would lend to if its aim were to promote social welfare. The most profitable loans are those to firms with average productivity (near ω^*), as Figure 3 makes clear. The least profitable, and therefore first to be denied credit, would be the high productivity firms on whose loans the bank earns no profit (firms with $\omega \leq \underline{\omega}$). These firms will get credit somewhere else.[23] The bank will first allocate funds to firms near ω^*, working its way down both sides of the triangle in Figure 3, lending to some firms with $\omega > \omega^*$ and some with $\omega < \omega^*$. A social planner would restrict the bank to lend first to its customers with $\omega > \omega^*$, since those firms cannot get credit from another bank.

As this last argument suggests, a liquidity crisis in a single bank will cause a drop in aggregate investment, as some customers of the affected bank either pay more for credit at another bank or shut down due to lack of credit. The bank's customers with $\omega > \omega^*$ will not

[23]Blackwell and Santomero (1982) came to the same conclusion for the same reasons, though in a very different model.

be able to get credit elsewhere. They will have to shut down, causing aggregate investment to fall. Customers with $\omega \in (\underline{\omega}, \omega^*]$ will get credit from another bank, but the outsider bank will insist on a higher repayment to cover its higher monitoring cost. These firms will still produce some capital, but less (in expectation) than if they had been able to continue the relationship with the insider bank. Firms with $\omega \leq \underline{\omega}$ are unaffected.

Now consider the case where the model economy shifts from having ample savings to having insufficient savings to fund all profitable projects. All banks will have to bid for funds, raising the equilibrium interest rate above R. Under these conditions, a new stable equilibrium will emerge with a smaller group of firms receiving credit to fund investment. The most productive firms will make up this group, since they can afford to pay more for funds. Along the transition to the new equilibrium, the disruption of relationships described above will temporarily reduce the capital stock.

5.2 Bank failure

Bank failure is another way established relationships can be disrupted. The possibility of bank failure will affect the terms of loan contracts, and when a bank failure occurs aggregate investment will fall. Return to the version of the model with ample savings and suppose that in each period, with some probability a single, randomly chosen bank fails because of fraud. Assume that when a bank fails, its borrowers still repay their loans, but depositors receive nothing. The terms of all loan contracts will change slightly, to reflect the chance that the lending bank will fail. Some savers lose their savings, but that has no effect on investment as credit is not rationed. More importantly, some firms will see their relationship with their bank disappear. These firms will have to turn to another bank for credit to invest in the second period of their life. Some of the firms that lost their banking relationship will be able to get a one-period loan as a new borrower; firms with $\omega \leq \omega^*$ will be able to get credit for just one period. However, less productive firms with $\omega \in (\omega^*, \hat{\omega})$ will not be able to replace the credit they were expecting from their failed lender, because their investments are not profitable enough to cover the higher cost of monitoring. They were only able to get credit in the first period because the bank took a loss, expecting to earn profits off the relationship in the second period. Aggregate investment will fall, temporarily, following a bank failure. It will return to its previous level once all firms who were affected by the failure have died.

6 Extensions and conclusions

6.1 More than three periods

If the firms and agents in this economy lived for more than three periods, the repayments charged under a long-term relationship would come closer to those charged under a long-term contract. A longer horizon implies outsider banks provide stiffer competition for insider banks during most of the relationship, because outsider banks could reap profits from dealing with the firm for many future periods instead of just one. In the final period of the firm's life, the situation would be exactly as described in section 3.5 for the second period of the two-period relationship. In the next to last period of the firm's life, an outsider bank will offer to initiate a two-period relationship (with certain firms) and the insider bank will limit price against the outsider bank's offer. This will continue back to the first period of the firm's life, where banks will, as before, offer it a low introductory rate to capture it as a customer. The range of firms that can get credit expands as the number of periods grows, and the range of firms who are charged the monopoly repayment in the last period of life will shrink. This latter effect makes the distance between the two arrangements diminish but not disappear.

6.2 Retained earnings

An obvious extension of the model would allow a firm to retain the earnings from its first investment to help finance its second investment. If no monitoring occurs in the first period, the firm will be left with $K_1 - F^1$ where K_1 denotes the outcome of the firm's first investment. Up until now, the firm has returned all profits to its owners as a dividend. If the firm retains these funds, it could borrow less in the second period, which would reduce the need for second-period monitoring, reducing expected monitoring costs. Allowing retained earnings to partially or fully finance investment will reduce the deadweight loss associated with borrowing, as numerous authors have pointed out.[24] In the present model, retained earnings' main effect would be to create a link between a firm's two investment projects, a link that is otherwise absent. While such a link may be an empirically appealing feature, it adds complexity without providing any substantial improvements in understanding of the contracts governing long-term dealings. For this reason, I describe (but do not derive) the changes that would occur in the long-term contract and long-term relationship if firms were allowed to retain earnings.

The long-term contract would change to make the terms of the second loan depend on the outcome of the first investment project.[25] Still considering only non-equity contracts (cf. section 3.4.1), the optimal contract is a sequence of one-period debt contracts where the terms of the second contract depend on the first period outcome. A good first-period outcome leaves the firm with more retained earnings, so it borrows less for its second investment. Because the firm borrows less, the bank can charge a lower repayment with lower deadweight loss. By reducing deadweight loss, allowing retained earnings increases firm profits. In effect, the optimal contract consists of a loan amount and repayment for the first loan, and a menu of loan amount-repayment pairs for the second loan. The bank must be able to commit to the second period terms in advance for the long-term contract to be feasible.

The long-term relationship will be essentially the same with or without retained earnings financing the second period investment. The bank will still offer a low first-period repayment that exactly offsets its expected profit when it exercises its market power in the second period of the relationship. The only difference arises because the bank's profits in the second period of the relationship – which help determine the terms of the first-period loan – depend on the amount borrowed in the second period, which depends on the result of the firm's first-period investment. The expression for the bank's expected second-period profits becomes much

[24]This is the point of Bernanke and Gertler (1989), as well as the "pecking-order" theory of corporate finance (Myers and Majluf 1984).

[25]The contract for this case is described exactly in Webb (1992).

more complicated. The terms of the first-period loan are determined as before, by ensuring that the bank earns zero expected profits over the course of the relationship.

The essential conclusions of the model remain unchanged. The long-term contract is still more efficient than the long-term relationship, as it always must be since it expands the set of permissible contracts. Middle-productivity old customers will still mean the most to the bank. The range of firms that can get credit will expand somewhat. Firms that before were just barely not given credit can now get credit because of the chance they might get a good outcome in the first period and be able to contribute some retained earnings to finance a second period investment. Adding retained earnings makes the long-term contract more complex to calculate and therefore slightly less likely to occur. The more complex the long-term contract, the more costly it is for agents to write and enforce it.

6.3 Conclusion

The model in this paper improves upon previous research into the theoretical nature of bank lending because it explicitly considers the long-term interaction between borrowers and lenders and embeds it in a general equilibrium framework. Long-term relationships are an important feature of financial markets; a good model should feature them.

The nature of the interaction between a bank and a firm depends on what commitment technology is available. If they can commit to a long-term contract, such a contract will be preferred because it eliminates the potential for the bank to extract rents from the firm in future periods of the relationship. If they cannot commit, their initial dealings will reflect their awareness of the bank's informational advantage once the relationship has begun. Commitment will always be better, from a social welfare point of view, because it expands the set of feasible contracts.

Firms are affected differently by long-term banking relationships. Very productive firms are unaffected, because publicly-available information is sufficient to induce savers to lend to them. Such lending would not have to go through a bank; borrowing by these firms is akin to a commercial paper market where a firm's creditworthiness alone backs up its debts.

The least productive firms are not helped by a long-term banking relationship because not even a long-term relationship can reduce agency costs enough to allow them to get external credit.

The middle-productivity firms are most affected by long-term banking relationships. For these firms to get credit, more information is required than what is publicly available. Banks produce that information cheaply by lending money under a debt contract. More of these firms get credit once a long-term relationship is taken into account by the lender than when the potential for long-term dealings is ignored. Identifying which firms are helped by long-term banking relationships is a main contribution of the model in this paper.

Another main contribution of the paper is to identify two situations where long-term relationships could be very important for aggregate investment: when a bank only has limited funds to lend and when a bank fails. When faced with a limited amount of funds, the bank will keep its most profitable customers and drop the rest. Traditional theory suggests this is innocuous, as the dropped customers would simply get credit elsewhere. But dropped customers who depended on the lower financing cost from their bank relationship for access to credit will *not* be able to get credit elsewhere, and the real economy will suffer as a result. The bank's profit-maximizing choice will not be socially optimal. When a bank fails, again, some firms will not be able to simply move to another lender. Real investment will fall in both cases.

The results of this paper complement the results of Bernanke and Gertler (1989) by expanding the range of financial shocks that can affect the real economy. They showed how a shock to a firm's balance sheet could reduce its investment by restricting its access to credit. I show how financial market turmoil, specifically a bank liquidity crisis or a bank failure, can reduce investment by disrupting established financing relationships.

References

Bernanke, Ben and Mark Gertler, "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review* 79 (March 1989), 14–31.

Bisignano, Joseph, "Structures of Financial Intermediation, Corporate Finance and Central Banking," Bank for International Settlements mimeo, December 1990.

Blackwell, Norman R. and Anthony M. Santomero, "Bank Credit Rationing and the Customer Relation," *Journal of Monetary Economics* 9 (January 1982), 121–129.

Chang, Chun, "The Dynamic Structure of Optimal Debt Contracts," *Journal of Economic Theory* 52 (1990), 68–86.

Diamond, Douglas W., "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies* 51 (1984), 393–414.

Diamond, Douglas W., "Debt Maturity Structure and Liquidity Risk," *Quarterly Journal of Economics* 106 (1991a), 709–737.

Diamond, Douglas W., "Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt," *Journal of Political Economy* 99 (1991b), 689–721.

Fama, Eugene F., "What's Different about Banks?" *Journal of Monetary Economics* 15 (1985), 29–39.

Farrell, Joseph and Carl Shapiro, "Optimal Contracts with Lock-In," *American Economic Review* 79 (March 1989), 51–68.

Gertler, Mark, "Financial Capacity and Output Fluctuations in an Economy with Multi-Period Financial Relationships," *Review of Economic Studies* 59 (1992), 455–472.

Haubrich, Joseph G., "Financial Intermediation: Delegated Monitoring and Long-Term Relationships," *Journal of Banking and Finance* 13 (1989), 9–20.

Mookherjee, Dilip and Ivan Png, "Optimal Auditing, Insurance, and Redistribution," *Quarterly Journal of Economics* 104 (1989), 399–415.

Morgan, Donald P., "Financial Contracts When Costs and Returns Are Private," *Journal of Monetary Economics* 31 (1993), 129–146..

Myers, Stewart C. and Nicholas Majluf, "Corporate Financing and Investment Decisions When Firms Have Information that Investors Do Not Have," *Journal of Financial Economics* 13 (1984), 187–221.

Sharpe, Steven A., "Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships," *Journal of Finance* 45 (September 1990), 1069–1087.

Townsend, Robert M., "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory* 21 (1979), 265–293.

Webb, David C., "Two-Period Financial Contracts with Private Information and Costly State Verification," *Quarterly Journal of Economics* 107 (1992), 1113–1124.

International Finance Discussion Papers

Please address requests for copies to International Finance Discussion Papers, Division of International Finance, Stop 24, Board of Governors of the Federal Reserve System, Washington, D.C. 20551.

International Finance Discussion Papers

IFDP Number	Titles	Author(s)
	1992	
439	Life Expectancy of International Cartels: An Empirical Analysis	Jaime Marquez
438	Daily Bundesbank and Federal Reserve Intervention and the Conditional Variance Tale in DM/$-Returns	Geert J. Almekinders Sylvester C.W. Eijffinger
437	War and Peace: Recovering the Market's Probability Distribution of Crude Oil Futures Prices During the Gulf Crisis	William R. Melick Charles P. Thomas
436	Growth, Political Instability, and the Defense Burden	Stephen Brock Blomberg
435	Foreign Exchange Policy, Monetary Policy, and Capital Market Liberalization in Korea	Deborah J. Lindner
434	The Political Economy of the Won: U.S.-Korean Bilateral Negotiations on Exchange Rates	Deborah J. Lindner
433	Import Demand and Supply with Relatively Few Theoretical or Empirical Puzzles	Andrew M. Warner
432	The Liquidity Premium in Average Interest Rates	Wilbur John Coleman II Christian Gilles Pamela Labadie
431	The Power of Cointegration Tests	Jeroen J.M. Kremers Neil R. Ericsson Juan J. Dolado
430	The Adequacy of the Data on U.S. International Financial Transactions: A Federal Reserve Perspective	Lois E. Stekler Edwin M. Truman
429	Whom can we trust to run the Fed? Theoretical support for the founders views	Jon Faust
428	Stochastic Behavior of the World Economy under Alternative Policy Regimes	Joseph E. Gagnon Ralph W. Tryon
427	Real Exchange Rates: Measurement and Implications for Predicting U.S. External Imbalances	Jaime Marquez
426	Central Banks' Use in East Asia of Money Market Instruments in the Conduct of Monetary Policy	Robert F. Emery
425	Purchasing Power Parity and Uncovered Interest Rate Parity: The United States 1974 - 1990	Hali J. Edison William R. Melick