Intermediary Financing without Commitment

Yunzhi Hu and Felipe Varas*

April 26, 2024

Abstract

Intermediaries reduce agency problems through monitoring, but credible monitoring requires sufficient retention until the loan matures. We study credit markets when intermediaries cannot commit to retention. Two structures are examined: certification (investors lend alongside intermediaries) and intermediation (investors lend through intermediaries via short-term debt). With a commitment to retention, they are equivalent. Without commitment, certification leads to loan sales and reduced monitoring over time. Intermediation, through short-term debt, encourages the intermediary to retain loans and incentivizes monitoring. Our analysis explains intermediaries' reliance on short-term debt - it enables the intermediary to internalize monitoring externalities.

Keywords: commitment; durable-goods monopoly; financial intermediaries; monitoring; dynamic games; optimal control in stratified domains;.

*Hu: Kenan-Flagler Business School, UNC Chapel Hill; Yunzhi_Hu@kenan-flagler.unc.edu. Varas: Universidad de los Andes, Chile; lvaras@uandes.cl. The authors would like to thank Douglas Diamond and Raghuram Rajan for very detailed feedback, comments and suggestions. The authors are also grateful to comments from Philip Bond, Brendan Daley, Peter DeMarzo, Darrell Duffie, Michael Fishman, William Fuchs, Thomas Geelen, Brett Green, Denis Gromb, Zhiguo He, Andrey Malenko, Daniel Neuhann, Christine Parlour, Stefano Pegoraro, Adriano Rampini, Michael Sockin, Anjan Thakor, Sheridan Titman, Jessie Wang, Victoria Vanasco, Vish Viswanathan, Shengxing Zhang, Jing Zeng, as well as participants at UNC, Chicago Booth, Duke, Kellogg, NUS, Cambridge, Virtual Finance Seminar Series, UT Austin, AFBC, FTWebinar, MoFiR, MFA, Minnesota Carlson, the Third PHBS Workshop in Macroeconomics and Finance, the FTG at University of Washington, Barcelona GSE Summer Forum, CICF, and Cambridge Corporate Finance Theory Symposium.

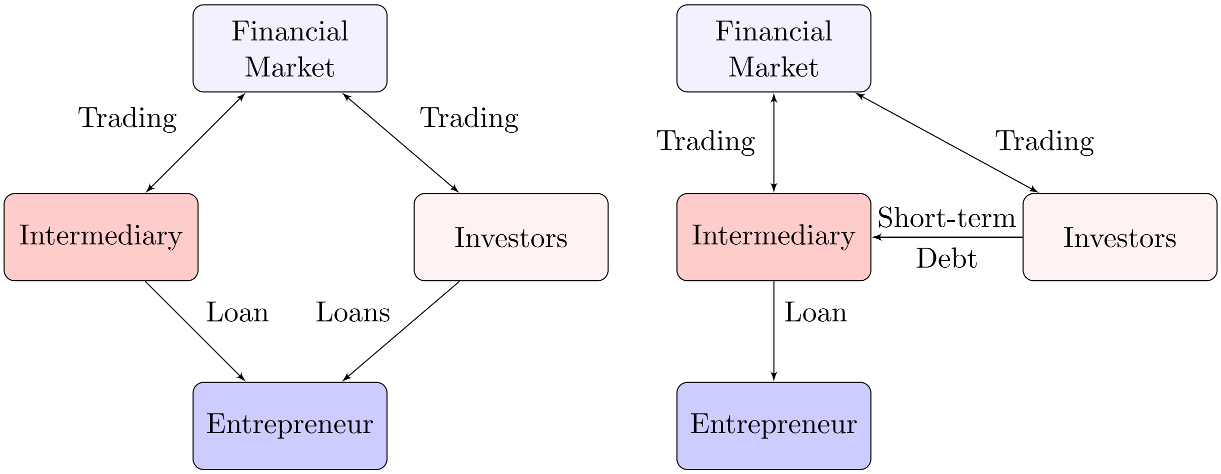
1 Introduction

Financial intermediaries conduct valuable services and, therefore, benefit the real economy. To appropriately align intermediaries' incentives, the optimal financing arrangement typically involves them retaining a fraction of loans as the skin in the game; otherwise, the incentives can be misaligned. With the development of the secondary loan market, intermediaries' commitment to loan retentions is limited (Drucker and Puri, 2009). This paper studies the equilibrium dynamics in loan sales and monitoring when intermediaries *cannot* commit to their retentions. We show that short-term debt can resolve the bank's lack of commitment problem and, therefore, align the incentives for monitoring and loan sales.

We build on the classic model of Holmstrom and Tirole (1997) in which banks can monitor to increase an entrepreneur's borrowing capacity. To maintain incentives for monitoring, the bank is required to retain a sufficient fraction of loans on its balance sheet. We assume the bank has no commitment to hold the loan, but rather has the option to sell loans. Given this, we study two types of intermediary structures: certification and intermediation. As shown in Figure 1, in certification, the bank lends alongside other investors. In intermediation, the bank borrows short-term debt from investors and then lends a collection of its own funds and those from investors. The certifying bank does not internalize the positive effect that its monitoring has on other investors. Therefore, the lack of commitment induces the bank to sell its loans gradually, and the bank's monitoring intensity declines over time. By contrast, the intermediating bank does not sell ex-post, because selling would immediately increase the interest rate charged on short-term debt. Hence, the short-term debt leads the bank to internalize the full value of its monitoring.

More specifically, we model an entrepreneur endowed with an investment opportunity, which requires a fixed-size of investment and pays off some final cash flows at a random time in the future. She has limited personal wealth and needs to borrow to make up the investment shortfall. Due to moral hazard in effort choices, she can only pledge a fraction of the final output to outside creditors, including banks and investors. Banks have a higher cost of capital, but only they can monitor to reduce the entrepreneur's private benefits. Although monitoring increases the project's pledgeable income and enables the entrepreneur to borrow more, it also entails a physical cost. Therefore, a credible monitor needs to retain a sufficient fraction of loans as its skin in the game.

We depart from Holmstrom and Tirole (1997) by introducing a competitive financial market, in which the bank is allowed to trade its loans and issue short-term debt against the loans. The bank's loan retention will be the state variable in our model. Loans are rationally priced, and therefore, the prices depend on the bank's incentives to monitor both contemporaneously and in the future. If the bank has sold or is expected to sell a large fraction of the loans, it will monitor less, and consequently, the price of loans will fall. This price impact deters the bank from selling the loans



(a) Certification (b) Intermediation

Figure 1: Certification vs. Intermediation

too aggressively.

Although certification and intermediation are equivalent in a model under the bank's commitment to its retention, they lead to different dynamics in loan sales and monitoring if there is no such commitment. A certifying bank has incentives to sell loans because it faces a higher cost of capital than investors, making it the inefficient loan holder. The price of the loan will drop when the bank tries to sell because investors anticipate reduced retention and monitoring. However, the price impact is insufficient to deter loan sales, because the bank has multiple future selling opportunities and cannot commit to not exploiting them. Since selling a small loan fraction only slightly reduces monitoring, the initial price impact will be minimal, and the bank will be incentivized to sell a positive amount. However, after this initial sale, the bank faces increased incentives to sell more loans. This is because the price impact from diminished monitoring now applies to only the remaining loan fraction retained by the bank, which is smaller. In equilibrium, the bank trades off the immediate trading gains versus the drop in its future payoff, including the drop in the loan's valuation as well as the decline in the expected payments it can collect upon the project maturing. The price of the loan is such that the certifying bank does not benefit from its ability to trade loans at all. The driving force is the time-inconsistency problem - the bank cannot credibly commit to maintaining the loan's value through monitoring in the future with the opportunities to sell. Each loan sale makes the bank's remaining stake smaller, reducing its incentive to internalize monitoring externalities. This dynamic leads to a gradual loan sales and less monitoring over time.

In contrast, an intermediating bank that issues short-term debt does not have incentives to sell its loans. This is because if the bank tries to sell, it would not only depress the loan's price, but more importantly, it would increase the interest rate on the short-term debt that was used to finance the loan, making it more costly for the bank. Since the bank constantly needs to roll over or refinance its short-term debt, an elevated interest rate acts as a punishment against loan sales. Consequently, despite having no explicit contractual commitment to retain loans, the bank finds it optimal to keep the entire loan portfolio on its balance sheet. This is due to the disciplining effect of having to repeatedly roll over short-term debt at potentially higher rates if it sells loans.

The certification role vs intermediation role relates to a monitoring externality. Monitoring benefits investors, but banks bear the costs, leading to sub-optimal monitoring (the free-rider problem). For certification, this externality, combined with a lack of commitment to retention, causes banks to sell loans and reduce monitoring probability over time. However, for intermediation, the externality is internalized by short-term debt. Short-term debt pricing immediately reflects any reduction in monitoring by the bank attempting to sell loans. This forces the bank to maintain monitoring incentives as both it and investors share the costs and benefits.

Our paper provides an explanation for why intermediaries rely on short-term debt. We show that an intermediating bank issuing long-term debt behaves like a certifying bank, selling off loans over time. Crucially, our mechanism requires that the intermediary's short-term debt is exposed to credit risk and this risk is fairly priced by markets. If, instead, the bank issues riskless short-term debt, or if its risky short-term debt is not fairly priced, then the intermediary exhibits the same loan sale behavior as a certifying bank.

Finally, we turn to the entrepreneur's initial choice of the intermediary structure. Given that an intermediating bank does not sell the loan, the intermediation structure naturally has a higher borrowing capacity. Therefore, an entrepreneur with a low net worth (high borrowing needs) may only be able to borrow enough under intermediation. However, we show that if the entrepreneur's net worth is high enough, she may end up choosing certification.

2 A Two-period Example

Before proceeding to the main model, we present a simplified example to illustrate the basic economic tradeoff. The model behind this example is studied in Appendix D.

There are three dates, and the model has one entrepreneur, one bank, and a competitive set of investors. All parties are risk-neutral and are protected by limited liabilities. We assume investors do not discount the future, whereas the bank and the entrepreneur discount the payoff one-period ahead by $\Phi < 1$.

Figure 2 describes the timing. The entrepreneur has a project that requires her effort at $t = 3$, after which the outcome is realized. The project produces R with probability $p_H = 1$ if the entrepreneur works but with probability $p_L = 0$ if the entrepreneur shirks. There are two shirking options: the high option brings private benefits $B > 0$, whereas the low brings $b = 0$.[1] At $t = 3$ and before the entrepreneur chooses her effort, the bank can monitor and eliminate the high shirking option by paying a non-pecuniary cost $\tilde{\kappa}$. This cost follows the uniform distribution on $[0, \bar{\kappa}]$ and is realized at the beginning of $t = 3$. For illustrative purposes, we assume that the entrepreneur does not have any skin in the game and has pledged all the cash flows R as a loan to the bank and investors. The financial market is potentially open at $t = 1$ and $t = 2$, in which the bank can trade the loan and issue debt. We normalize the total share of the loan as one. Let θ_0 be the share of the loan retained by the bank at the beginning of $t = 1$ and $1 - \theta_0$ be that retained by investors.[2] After trading, the bank retains a share θ_1 and θ_2 of the loan at the beginning of $t = 2$ and $t = 3$. We assume the bank issues one-period debt D_1 and D_2 at $t = 1$ and $t = 2$.

Given a realization of the monitoring cost κ, the bank chooses to monitor iff $\kappa \leq \theta_2 R - D_2$. For the remainder of this section, we sometimes refer to $\theta_2 R - D_2$ as the bank's *net exposure*.

$t = 1$	$t = 2$	$t = 3$
- Bank sells loan $\theta_0 - \theta_1$	- Bank repays debt D_1	- Monitoring cost $\tilde{\kappa}$ is realized.
- Bank issues debt D_1	- Banks sells loan $\theta_1 - \theta_2$	- Bank decides whether to monitor.
	- Bank issues debt D_2	- The entrepreneur chooses to work or shirk.
		- Cash flows are realized. Payments are settled.

Figure 2: Timing in the two-period model

Without the Financial Market. If the financial market is never open, the bank and investors always retain their shares until $t = 3$. Therefore, $\theta_2 = \theta_0$, and $D_2 = 0$. The bank monitors if and only if the realized monitoring cost κ is below $\theta_0 R$.

Financial market is open once. Now we analyze the model in which the financial market is open once at $t = 2$. Therefore, the bank's retention before the market opens is $\theta_1 = \theta_0$. In

[1] Therefore, the entrepreneur is indifferent between working and the low shirking option. In this case, we assume that she chooses to work. The low shirking option is redundant in this illustrative example, but not in the main model with $b \geq 0$.

[2] In the appendix, we further enrich the model by adding a date $t = 0$ at which θ_0 is determined as part of the optimal financing arrangement.

the financial market, the bank can either sell loan $\theta_1 - \theta_2$ or issue debt D_2 against it. For any $(\theta_1, \theta_2, D_2)$, the bank's payoff at $t = 2$ is

$$\hat{V}_2(\theta_2, D_2, \theta_1) = \Phi\left(\underbrace{p_2(\theta_2, D_2)(\theta_2 R - D_2)}_{\text{loan payment} - \text{debt repayment}} - \underbrace{\int_0^{(\theta_2 R_o - D_2)} \frac{\kappa}{\bar{\kappa}} d\kappa}_{\text{monitoring cost}}\right) + \underbrace{p_2(\theta_2, D_2) D_2}_{\text{debt issuance}} + \underbrace{q_2(\theta_2, D_2)(\theta_1 - \theta_2)}_{\text{loan sales}},$$

where $p_2(\theta_2, D_2) = \frac{\theta_2 R - D_2}{\bar{\kappa}}$ is the probability that the bank monitors. Given $p_H = 1$ and $p_L = 0$, $p_2(\theta_2, D_2)$ is the probability that the project generates R. Because investors do not discount the future, $p_2(\theta_2, D_2)$ is also the price of debt D_2, and $q_2(\theta_2, D_2) = p_2(\theta_2, D_2)R$ is the price per share of the loan. The bank chooses loan sales $\theta_1 - \theta_2$ and debt issuance D_2 to maximize $\hat{V}_2(\theta_2, D_2, \theta_1)$, subject to the constraint that $D_2 \leq \theta_2 R$. Given θ_1, its maximized continuation value is

$$\begin{aligned} V_2(\theta_1) = & \max_{\theta_2 \in [0,1], D_2 \geq 0} \hat{V}_2(\theta_2, D_2, \theta_1) \\ & \text{s.t.} \quad D_2 \leq \theta_2 R. \end{aligned}$$

This issuance constraint $D_2 \leq \theta_2 R$ arises because if D_2 exceeds $\theta_2 R$, the bank will always default: it can never pay off the debt at $t = 3$. It is easily derived that the first-order condition on θ_2 and D_2 are linearly dependent, so that the choice between θ_2 and D_2 is undetermined.[3] This result implies that *if the financial market is open only once, loan sales and debt issuance are equivalent to the bank*. Intuitively, both are equivalent approaches for the bank to reduce its net exposure or, equivalently, its skin in the game.

Financial market is open twice. Now, we turn to the model in which the financial market is open at both $t = 1$ and $t = 2$. The analysis at $t = 2$ stays unchanged from the one above, with the only distinction that the bank starts at $t = 2$ with retention θ_1 that can differ from θ_0 due to the possibility of trading at $t = 1$. Recall that $V_2(\theta_1)$ is the bank's maximized continuation payoff at $t = 2$. Given any θ_0, the bank's payoff from loan sales and issuing one-period debt at $t = 1$ is

$$\hat{V}_1(\theta_1, D_1, \theta_0) = \underbrace{D_1}_{\text{debt issuance}} + \underbrace{q_1(\theta_1, D_1)(\theta_0 - \theta_1)}_{\text{loan sales}} + \underbrace{\Phi\left(V_2(\theta_1) - D_1\right)}_{\text{payoff at } t=2}.$$

[3] That is, $\frac{\partial \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial \theta_2} = -R \frac{\partial \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial D_2}$.

Note that the one-period debt D_1 will be paid with certainty at $t = 2$ and is therefore riskless.[4] Moreover, when investors buy the loan at $t = 1$, they anticipate the bank to also sell loans and/or issue debt at $t = 2$, and such anticipation is reflected in the loan price q_1 so that $q_1(\theta_1, D_1) = q_2(\theta_2(\theta_1), D_2(\theta_1))$. Note that even though the optimal solution $\{\theta_2, D_2\}$ for a given θ_1 is not unique, any solution will generate the same default probability and the same loan price $q_2(\theta_2, D_2)$.

The bank's problem at $t = 1$ is to choose loan sales $\theta_0 - \theta_1$ and debt issuance D_1 to maximize $\hat{V}_1(\theta_1, D_1, \theta_0)$, subject to the issuance constraint $D_1 \leq V_2(\theta_1)$. We show in the appendix that the solutions are

$$\begin{aligned}\theta_1 &= \theta_0 \\ D_1 &= V_2(\theta_1).\end{aligned}$$

In contrast to the results at $t = 2$, loan sales and debt issuance are no longer equivalent at $t = 1$. In fact, the bank prefers debt issuance strictly over loan sales: it does not sell any loan but issues the maximum amount of debt against it. Intuitively, this result holds because the $t = 1$ buyers of the loan are concerned with the bank's decisions of loan sales and debt issuance at $t = 2$. The more the bank sells the loan and/or issues debt at $t = 2$, the less likely that it will monitor at $t = 3$, and the loan buyers at $t = 1$ anticipate receiving lower expected payments when the loan finally matures. By contrast, the creditors at $t = 1$ are not concerned with the bank's decisions at $t = 2$, because they will get fully repaid before the bank sells loans and issues debt then.

Financial market without debt issuance. Finally, if the financial market is open on both dates but debt issuance is not allowed, then the bank sells loans on both dates so that $\theta_1 \leq \theta_0$ and $\theta_2 \leq \theta_1$ hold. The formal proof is available in the appendix.

The next proposition summarizes the results in this section.

Proposition 1. *The equilibrium in the model with two periods is as follows.*

1. *At $t = 2$, loan sales and debt issuance are equivalent to the bank; that is, given any θ_1, there is a continuum of θ_2 and D_2 that maximize the bank's payoff at $t = 2$.*

2. *At $t = 1$, the bank strictly prefers issuing debt to loan sales; that is, given any θ_0, $\theta_1 = \theta_0$ and $D_1 = V_2(\theta_1)$.*

3. *If debt issuance is not allowed, i.e., $D_1 = D_2 = 0$, then the bank sells loans on both dates, $\theta_1 \leq \theta_0$ and $\theta_2 \leq \theta_1$.*

[4] In this example, D_1 is always risk-free, driven by the particular timing assumption that the project never matures before $t = 3$. Later in the continuous-time model, we assume that the project matures after a random shock so that short-term debt is no longer riskless.

It is intractable to study a full-dynamic model in discrete time that allows the financial market to open more than twice. The dynamics of loan sales and the welfare implications are, therefore, difficult to characterize. Formulating the problem in continuous time allows for a clean characterization of the equilibrium and associated trading dynamics.

Commitment Solution. The bank's net exposure at maturity, given by $\theta_2 R - D_2$, is the key determinant of its monitoring incentives. To achieve the desired net exposure, the bank has two equivalent strategies: selling loans or issuing debt. With commitment, the price of the loan q_1 depends directly on the bank's committed net exposure $\theta_2 R - D_2$ at the beginning of $t = 3$, regardless of the specific strategy employed. For this reason, when the bank can commit to its net exposure, there is no distinction between loan sales and debt issuances because both allow for achieving the optimal net exposure, with the corresponding monitoring probability $p_2(\theta_2, D_2)$.

3 The Model

3.1 Agents and Technology

Time is continuous and goes to infinity: $t \in [0, \infty)$. There are three groups of agents: one entrepreneur (she) – the borrower; competitive intermediaries – banks; and investors. All agents are risk-neutral and have limited liability. The entrepreneur starts out with cash level A, whereas banks and investors have deep pockets. We assume investors do not discount future cash flows, whereas the entrepreneur and intermediaries discount the future at a rate $\rho > 0$. Investors in our model should be interpreted as institutional investors such as sovereign wealth funds, hedge funds, insurance companies, and cash-rich companies.

At time 0, the entrepreneur has access to a project that requires a fixed investment size $I > A$. Thus, she needs to borrow at least $I - A$. The project matures at a random time τ, which arrives upon a Poisson event with intensity $\phi > 0$. Define $\Phi = \frac{\phi}{\rho+\phi}$ as the effective time discount the entrepreneur and banks apply to the project's final cash flows. At τ, the project generates the final cash flows R in the case of success and 0 in the case of failure. The probability of success is p_H if the entrepreneur works at τ, and $p_L = p_H - \Delta$ if she shirks. Two options of shirking are available: the high option brings private benefit B, which exceeds b, the private benefit of the low option. We assume the project's expected payoff is always higher if the entrepreneur works; that is, $p_H R > p_L R + B$.

3.2 Monitoring, Financial Structures, and Contracts

At $t = 0$, a competitive set of banks is present, and the entrepreneur enters into a contract with one of them. Banks in the model should be broadly interpreted as any lender that is capable of costly reducing the agency frictions. Note we do not allow for multiple banking relationships to avoid duplication of monitoring efforts (Diamond, 1984). At τ, the project matures, and the bank can monitor to eliminate the high shirking option. To do so, it needs to pay private monitoring cost $\tilde{\kappa} > 0$, where $\tilde{\kappa} \in [0, \bar{\kappa}]$ has a distribution with $F(\cdot)$ and $f(\cdot)$ being the cumulative distribution function (CDF) and probability density function (PDF), respectively. The stochastic-cost assumption smooths the bank's equilibrium monitoring decisions, which become a continuous function of its loan retention. Stochastic costs can be interpreted as variations in legal and enforcement costs, or simply fluctuations in the costs of hiring loan officers.[5] Figure 3 describes the timing.

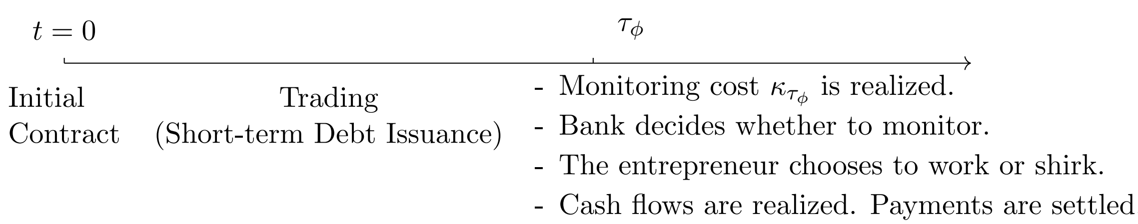


Figure 3: Timing

We study two types of financial structures: certification and intermediation. In certification, the bank puts its own funds in the entrepreneur's venture to certify it will monitor, which then attracts investors to invest in the venture as well. One example of this type of bank is the lead investment bank in loan syndication. Under limited liability, no agent receives anything if the project fails. If the project succeeds, let R_f be the cash flows retained by the entrepreneur, and $R_o = R - R_f$ be the scheduled payments to outside creditors, namely, the bank and investors. For the remainder of this paper, we also refer to R_o as *loans*.

In intermediation, investors do not directly invest in the entrepreneur's project. Instead, they lend to the bank, which then lends to the entrepreneur a collection of its own funds and the money from investors.[6] One example of this type of bank is a shadow bank, which borrows short-term

[5]Two alternative formulations will generate results equivalent to that of a stochastic monitoring cost: The first is to introduce a continuous distribution of private benefits, and the second one is to assume that monitoring is only effective with some probability. This probability is assumed to vary continuously with the bank's monitoring effort.

[6]The money from investors should not be interpreted as FDIC-insured deposits. Instead, it comes from investors who are sensitive to new information regarding the bank's underlying business. These investors put their money in the

debt from lenders. The bank offers short-term debt contracts $\{D_t, y_t\}$ to investors over time, where D_t is the amount of debt and y_t is the associated interest rate. We model short-term debt as one with instant maturity. Debt with instant maturity is the continuous time analogous to one-period debt in discrete time.

Given that the entrepreneur has (weakly) the highest cost of capital among all the agents, she should retain as little stake as possible. Therefore, in both certification and intermediation, it is optimal to let the entrepreneur retain $R_f = b/\Delta$, which guarantees she will work if the bank monitors. Therefore, $R_o = R - b/\Delta$.[7]

3.3 Trading and Pricing in the Financial Market

A competitive financial market is open after time 0 in which loans can be traded.[8] We normalize the total share of loans outstanding to one and use θ_t to denote the bank's retention at time t. In our model, θ_t will be the payoff-relevant state variable in the Markov perfect equilibrium introduced later. Sometimes θ_t is also referred to as the bank's skin in the game, which is publicly observable. Before trading starts, the bank's initial retention $\theta_0 \in [0, 1]$ is optimally chosen by the entrepreneur in certification, and $\theta_0 = 1$ holds by definition in intermediation. We consider trading strategies that admit both smooth and atomistic trading, as well as mixed strategies over the time of atomistic trades. A Markov trading strategy is defined as $(\theta_t)_{t\geq 0}$ being a Markov process.

The price of loans depends on whether the entrepreneur works or shirks, which in turn depends on the probability of bank monitoring and, therefore, its net exposure. Conditional on the project maturing, let $p(\theta, D)$ be the equilibrium probability of success; investors of the loan receive

$$d(\theta, D) = p(\theta, D)\, R_o \tag{1}$$

per share. Let $q(\theta, D)$ be the price of the loan per share when $\theta_t = \theta$. In a competitive financial market, the price is given by the expected present value:

$$q(\theta, D) = \mathbb{E}\left[d(\theta_\tau, D_\tau) \middle| \theta_t = \theta \right], \tag{2}$$

bank in the form of short-term debt. In practice, one can interpret short-term debt as brokered deposits, repurchase agreements, wholesale lending, and commercial papers. More broadly, they can be interpreted as "money-market preferred stock," which carries a floating dividend rate that is reset periodically to maintain the stock's market value at par. The distinction between equity and debt is unimportant in our setup when cash flows are binary, with one realization being zero. As we show below, the crucial feature is that current information about the bank promptly becomes impounded into the rate it pays to capital suppliers.

[7] The entrepreneur can also borrow directly from investors, but the borrowing capacity is (weakly) lower.

[8] We assume the entrepreneur's retention R_f is not tradable, or equivalently, the entrepreneur can commit to holding onto R_f on the balance sheet.

where the expectation operator is taken with respect to the equilibrium path of $\{\theta_s\}_{t\leq s\leq\tau}$.

The probability of success $p(\theta, D)$ will differ in certification and intermediation. Let κ be the realization of the stochastic monitoring cost $\tilde{\kappa}$ at τ. In certification, $D = 0$, and the bank with retention θ chooses to monitor if and only if

$$p_H\theta R_o - \kappa \geq p_L\theta R_o \Rightarrow \kappa \leq \kappa_c \coloneqq \Delta R_o\theta, \tag{3}$$

where $\Delta \coloneqq p_H - p_L$. In intermediation, a bank with retention θ and short-term debt D monitors if and only if

$$p_H(\theta R_o - D) - \kappa \geq p_L(\theta R_o - D) \Rightarrow \kappa \leq \kappa_i \coloneqq \Delta\left(R_o\theta - D\right). \tag{4}$$

From now on, we use subscripts c and i to differentiate certification and intermediation.

Assumption 1 restricts the (expected) monitoring cost to be sufficiently low.

Assumption 1.

$$\int_0^{\Delta R-b} \kappa dF(\kappa) \leq \Phi F(\Delta R - b)(\Delta R - b) - (1-\Phi)\, p_L(\Delta R - b).$$

This assumption leads to the following result. If the bank always retains the entire loan (i.e., $\theta_t \equiv 1,\ \forall t \leq \tau$), the bank's payoff exceeds that if it immediately sells the entire loan and never monitors. If this assumption is violated, bank monitoring is never needed in equilibrium.

We also impose Assumption 2, such that the intermediation bank's payoff (defined later) is concave in the amount of short-term debt issuance. This assumption is satisfied by many commonly used distribution functions, such as uniform distribution.

Assumption 2. *The distribution function $f(\tilde{\kappa})$ satisfies that for $\forall\kappa \in [0, \bar{\kappa}]$,*

$$\begin{aligned}
&-\Phi\left[\hat{p}_i(\theta, D) f(\kappa) + f^2(\kappa)\Delta^2 D\right] < \Delta f'(\kappa) D < f(\kappa),\forall D \in [0, p_H R_o\theta],\ \forall\theta \in [0,1] \\
&\frac{f'}{f^2} < \frac{\Phi}{1-\Phi}\Delta.
\end{aligned}$$

4 Equilibrium

This section analyzes the equilibrium in certification and intermediation structures. As depicted in Figure 1, there are two key distinctions between these structures. First, an intermediating bank can issue short-term debt, whereas a certifying bank cannot. Second, the intermediating bank is the only lender, while the certifying bank is one of many lenders. Due to this second difference, the bank's initial retention θ_0 is endogenously determined at $t = 0$ in the certification structure. In this case, competition among banks at $t = 0$ leads them to offer a θ_0 that maximizes the entrepreneur's payoff. By contrast, in the intermediation structure, the bank's initial retention satisfies $\theta_0 = 1$, given that it is the only lender. In subsection 6.1, we consider an extension where the intermediating bank is also one of many lenders and demonstrate that our main results remain robust. Therefore, the key distinguishing factor between the two structures is the intermediating bank's ability to issue short-term debt.

4.1 Certification Equilibrium

In certification, $D \equiv 0$, so that we write all relevant valuation and prices as functions of θ. If the project matures at time t, the bank receives loan payments net of the monitoring cost

$$\pi_c(\theta) = p_c(\theta) R_o \theta - \int_0^{\kappa_c} \kappa dF(\kappa), \tag{5}$$

where the project succeeds with probability

$$p_c(\theta) := p_L + F(\kappa_c)\Delta, \tag{6}$$

upon which the bank receives $R_o\theta$. Following the Envelope Theorem, it is easily shown that

$$\pi_c'(\theta) = p_c(\theta)R_o + \underbrace{\frac{\partial p_c(\theta)}{\partial \theta} R_o \theta - \kappa_c f(\kappa_c) \frac{\partial \kappa_c}{\partial \theta}}_{=0} = d_c(\theta). \tag{7}$$

Let $G(\theta)$ be the bank's cumulative trading gains. In the case of atomistic trading, the bank's holding jumps from θ to θ^+ and the associated trading gain is $dG(\theta) = q(\theta^+)(\theta - \theta^+)$. Note that trading is settled at a price $q(\theta^+)$ to reflect the price impact. In the case of continuous trading, $dG(\theta) = -q(\theta)\dot{\theta}dt$. The bank maximizes the sum of its payoff upon the project's maturation $e^{-\rho(\tau-t)}\pi_c(\theta_\tau)$ and the cumulative trading gains $\int_0^\tau e^{-\rho(s-t)} dG(\theta_s)$. Because τ follows the exponential distribution, the bank's problem can be equivalently written as

$$\max_{\{\theta_t\}_{t\geq 0}} \mathbb{E}\left[\int_0^\infty e^{-(\rho+\phi)t}\Big(\phi\pi_c(\theta_t)\,dt + dG(\theta_t)\Big)\right], \tag{8}$$

where the expectation operator also allows for mixed strategies in $\{\theta_t\}$. Let V_c be the entrepreneur's expected payoff:

$$V_c = \mathbb{E}\left[\int_0^\infty e^{-(\rho+\phi)t}\phi\left\{\mathbb{1}_{\{\kappa\leq\kappa_c\}}p_H R_f + \mathbb{1}_{\{\kappa>\kappa_c\}}(p_L R_f + B)\right\}dt\right]. \tag{9}$$

If the realized monitoring cost is lower than the threshold κ_c defined in (3), the bank monitors, and the entrepreneur receives $p_H R_f$ in expectation. Otherwise, the bank chooses not to monitor, and the entrepreneur receives the expected return $p_L R_f$ together with the private benefits B.

We consider a Markov perfect equilibrium in which the state variable is the bank's retention θ, henceforth, the certification equilibrium.[9]

Definition 1. *A **certification equilibrium** is a Markov perfect equilibrium consisting of a price function $q\colon [0,1] \to \mathbb{R}_+$ and a trading strategy $(\theta_t)_{t\geq 0}$ that satisfy the following:*

1. *Given $\theta_0 \in [0,1]$, $(\theta_t)_{t\geq 0}$ is a Markov trading strategy that maximizes (8).*

2. *For all $\theta \in [0,1]$, the price $q(\theta)$ satisfies the break-even condition (2).*

In general, the bank can trade loans smoothly or atomistically. We show both types of trading can occur in equilibrium. Let $\Pi_c(\theta)$ be the bank's value function with retention θ.[10] In the smooth-trading region, $\Pi_c(\theta)$ satisfies the following Hamilton-Jacobi-Bellman (HJB) equation:

$$\rho\Pi_c(\theta) = \max_{\dot{\theta}} \ \phi\Big[\pi_c(\theta) - \Pi_c(\theta)\Big] + \dot{\theta}\Big[\Pi_c'(\theta) - q_c(\theta)\Big]. \tag{10}$$

Whereas the left-hand side stands for the bank's required return, the first term on the right-hand side represents the event of the project maturing, in which case the bank receives $\pi_c(\theta)$ defined in (5). The second term captures the overall benefit of trading, which includes the change to the bank's continuation value and the trading gain. A necessary condition for smooth trading is

$$\Pi_c'(\theta) = q_c(\theta), \tag{11}$$

so that the bank is indifferent between trading or not: the per-share trading gain $q_c(\theta)$ is offset by the drop in the bank's continuation value $\Pi_c'(\theta)$. Substituting the indifference condition (11) into (10), we get that in the region of smooth trading,

$$\Pi_c(\theta) = \frac{\phi}{\rho + \phi}\pi_c(\theta) = \Phi\pi_c(\theta). \tag{12}$$

[9] If $\{\theta_t\}$ is not restricted to the class of Markov processes, one may construct equilibria that are close to the commitment solution. In the context of a durable-goods monopoly, Ausubel and Deneckere (1989) show that in the no-gap case, non-Markov equilibria exist in which the seller can achieve payoffs close to the commitment solution. The logic behind the construction is similar to the one in the folk theorem for repeated games. The no-gap case corresponds to the version of our model under $p_L = 0$. At $\theta = 0$, the marginal valuation of investors coincides with the bank's under $p_L = 0$.

[10] A certifying bank does not issue debt, so there is no distinction between the bank value and the equity value.

Given that $\Phi\pi_c(\theta)$ is also the bank's continuation value if it never sells any loan until the project matures, (12) implies that in equilibrium, the bank does not benefit from its ability to trade these loans in the financial market at all. The observation that lack of commitment fully offsets the trading gains has been noted in previous models on bargaining (Fuchs and Skrzypacz, 2010; Daley and Green, 2020) and in other corporate finance settings (DeMarzo and Urošević (2006) in trading by a large shareholder, and DeMarzo and He (2021) in leverage dynamics).

Even though the bank's equilibrium payoff is identical to if it always retained the entire loan, this does not imply the bank will never trade loans on the equilibrium path. In fact, the price of loans if the bank followed a strict no-trade strategy would be too high, incentivizing the bank to sell. The bank sells loans for two reasons: First, due to its higher cost of capital, the bank's marginal valuation is below investors', making it the inefficient holder of the loan. Selling a small loan fraction only slightly reduces monitoring incentives while the initial price impact is minimal, incentivizing some sales. Second, after the initial sale, the bank is willing to sell more because the price impact only accrues to its remaining retention stake, which gets smaller. We characterize the bank's equilibrium trading strategy, which comes from the determination of the equilibrium loan prices. Because investors do not discount future cash flows, $q_c(\theta)$ must satisfy the investors' zero-profit equation whenever the bank trades smoothly:

$$0 = \phi\left[d_c(\theta) - q_c(\theta)\right] + \dot{\theta} q_c'(\theta). \tag{13}$$

The left-hand side is the investors' required rate return, which is zero. On the right-hand side, $\phi\left[d_c(\theta) - q_c(\theta)\right]$ stands for the event of project maturing, in which case investors receive $d_c(\theta)$ per share, whereas the last term is the change in the loan price due to trading.[11] Combining (11), (12), and (13), and using the relation $d_c(\theta) = \pi_c'(\theta)$, one can derive the following equilibrium trading strategies:

$$\dot{\theta} = -\phi \frac{(1-\Phi)\,\pi_c'(\theta)}{\Phi \pi_c''(\theta)} < 0. \tag{14}$$

In the smooth-trading region, the bank sells loans over time and its retention declines continuously. Intuitively, the equilibrium loan price is forward-looking, and therefore takes into account the bank's monitoring decisions in the future. To satisfy the bank's indifference condition, the equilibrium price of the loan cannot be too high. The only trading strategy consistent with this price requires the bank to sell its loans over time.

So far, we have only considered the case of smooth trading. Meanwhile, the bank also has the

[11] The terms without subscripts c have been defined in equations (1) and (2).

option to sell an atom of loans. In general, the bank can sell either a fraction or all the remaining loans. Lemma 4 in the appendix proves the bank will never sell a fraction. This result follows the intuition in standard Coasian dynamic models. Atomistic trading arises whenever the bank has strict incentives to sell. If so, it prefers to sell as fast as possible. Given this result, we are left to check when the bank decides to sell off all the remaining loans at a price $q_c(0)$, where $q_c(0) = p_L R_o$ is the per-share loan price without monitoring. We show that a unique θ_* exists below which the bank finds it optimal to sell off all the remaining loans; that is, $\theta_* q_c(0) = \Phi\pi_c(\theta_*)$, and $\theta q_c(0) > \Phi\pi_c(\theta)$ if and only if $\theta < \theta_*$.

Next, we derive the trading strategy at $\theta = \theta_*$. On the one hand, the bank cannot hold onto the remaining loans forever because the resulting loan price will be too high to induce the bank to sell. On the other hand, the bank cannot sell smoothly either; if so, shortly afterward, θ will drop below θ_*, and the bank will have strict incentives to sell off the rest of the loan. Furthermore, it cannot be that the bank sells off all the remaining loan immediately after θ_t reaches θ_*, because if so, the price of the loan will experience a deterministic downward jump, inconsistent with equation (2). The only (stationary) trading strategy at θ_* consistent with (2) is for the bank to adopt a mixed strategy:[12] the bank sells off all its remaining loans at time ν, which arrives upon a Poisson event at intensity λ that satisfies

$$q_c(\theta_*) = \mathbb{E}\left[d_c(\theta_\tau)|\theta_t = \theta_*\right] = \frac{\lambda}{\phi+\lambda} d_c(0) + \frac{\phi}{\phi+\lambda} d_c(\theta_*). \tag{15}$$

Finally, at $t = 0$, the entrepreneur chooses θ_0 to maximize $V_c(\theta_0) + L_c(\theta_0)$, subject to the borrowing constraint

$$L_c(\theta_0) = \underbrace{\Pi_c(\theta_0)}_{\text{bank lending}} + \underbrace{(1-\theta_0)q_c(\theta_0)}_{\text{investors' lending}} \geq I - A.$$

We show in the appendix that $L_c(\theta_0)$ increases monotonically so that $\theta_0 = 1$ maximizes the entrepreneur's initial borrowing. Intuitively, under this contract, the bank needs to spend the longest period to fully sell its loans. However, $V_c(\theta_0) + L_c(\theta_0)$ reaches its maximum at an interior point. Later in subsection 5.3, we present a linear-quadratic example with closed-form solutions, in which we show that an arrangement with $\theta_0 = 1$ might lead to excessive monitoring from the perspective of the entrepreneur.

Proposition 2 summarizes the previous discussion and describes the equilibrium outcome. The formal proof requires verification that the bank's trading strategy is optimal, which is supplemented

[12]The delay can also be deterministic, but the equilibrium is no longer within the class of a Markov perfect equilibrium. The equilibrium would also depend on the time since the bank's retention reached θ_*. The price q_t would not be stationary, and it would depend on the trading history before time t.

in the appendix using results from the theory of optimal control in stratified domains.[13]

Proposition 2 (Certification Equilibrium)**.** *A unique* ***certification equilibrium*** *exists. Given the bank's initial retention θ_0, the bank sells its loans smoothly at a rate given by equation (14) until $T_* := \inf\{t > 0 : \theta_t = \theta_*\}$, after which it sells off its remaining loans at some Poisson rate λ that satisfies (15). The equilibrium loan price is*

$$q_c(\theta) = \begin{cases} \Phi\Big(p_L + F\left(\Delta R_o \theta\right)\Delta\Big) R_o & \theta > \theta_* \\ \Big(p_L + \frac{\phi}{\lambda+\phi} F\left(\Delta R_o \theta_*\right)\Delta\Big) R_o & \theta = \theta_* \\ p_L R_o & \theta < \theta_*. \end{cases} \tag{16}$$

If $L_c(1) < I - A$, the entrepreneur cannot borrow enough to invest at $t = 0$. If $L_c(1) > I - A$, then the optimal choice satisfies $\theta_0 < 1$.

4.2 Intermediation Equilibrium

Let $D_t = D$ be the outstanding debt of an intermediation bank at time t. If the project matures, the bank's equity holders receive loan payments net of debt repayments and monitoring cost:

$$\hat{\pi}_i\left(\theta, D\right) = \hat{p}_i\left(\theta, D\right)\left(\theta R_o - D\right) - \int_0^{\kappa_i} \kappa dF\left(\kappa\right), \tag{17}$$

where the project succeeds with probability

$$\hat{p}_i\left(\theta, D\right) := p_L + F\left(\kappa_i\right)\Delta, \tag{18}$$

upon which the bank's equity holder receives $\max\{\theta R_o - D, 0\}$. Besides trading gains, an intermediating bank also receives income from issuing short-term debt. In particular, the bank's net income from debt issuance at time t is $dD_t - y_t D_t dt$, where

$$y_t = \hat{y}\left(\theta, D\right) = \phi\left(1 - \hat{p}_i(\theta, D)\right) \tag{19}$$

[13] Due to the discontinuity in the price function $q_c(\theta)$, the HJB equation (10) is discontinuous at θ_*. This technical problem can be sidestepped using (discontinuous) viscosity solution methods.

compensates the default risk borne by creditors.[14] In intermediation, the bank trades loans and issues short-term debt to maximize the expected payoff upon the project maturing, together with the cash flows from short-term debt issuance and trading gains $dG(\theta_t)$; that is,

$$E(\theta, D) = \max_{\{\theta_t,\, D_t\}} \mathbb{E}\left[\int_0^\infty e^{-(\rho+\phi)t} \Big(\phi\hat{\pi}_i\left(\theta_t, D_t\right) dt + \left[dD_t - \hat{y}\left(\theta_t, D_t\right) D_t dt\right] + dG\left(\theta_t\right) \Big) \right]. \tag{20}$$

The choice of D_t in (20) is restricted by the bank's limited liabilities, which imposes an issuance constraint as illustrated below in (21). Lemma 1 shows that we can solve short-term debt issuance and loan trading separately.

Lemma 1. *The maximization problem* (20) *is equivalent to solving*

$$\phi\pi_i(\theta) := \max_{D \leq \Pi_i(\theta)} \left\{ \phi \left[\hat{p}_i(\theta, D)\theta R_o - \int_0^{\kappa_i} \kappa dF\left(\kappa\right) \right] + \rho D \right\}, \tag{21}$$

where

$$\Pi_i\left(\theta\right) = E(\theta, D) + D = \max_{(\theta_t)_{t\geq 0}} \int_0^\infty \mathbb{E}\left[e^{-(\rho+\phi)t} \Big(\phi\pi_i(\theta_t)dt + dG\left(\theta_t\right) \Big) | \theta_0 = \theta, D_0 = D \right]. \tag{22}$$

Given this result, we can suppress the problem's dependence on D_t and use θ_t as the state variable. Note that in (21), debt issuance is bounded by the endogenous constraint $D \leq \Pi_i\left(\theta\right)$, which arises from the bank's limited liabilities.[15] In (22), the left-hand side $\Pi_i\left(\theta\right)$ includes $E\left(\theta, D\right)$, the value to the bank's equity holders, and D, the value of short-term debt. One implication of (22) is that even though the bank's equity holders decide its trading strategy, maximizing the bank's equity value is equivalent to maximizing the total bank value, because debt D is continuously repriced.

We use V_i to denote the entrepreneur's expected payoff in intermediation:

$$V_i = \mathbb{E}\left[\int_0^\infty e^{-(\rho+\phi)t} \phi \left\{ \mathbb{1}_{\{\kappa \leq \kappa_i\}} p_H R_f + \mathbb{1}_{\{\kappa > \kappa_i\}} (p_L R_f + B) \right\} dt \right].$$

[14] A heuristic derivation of (19) goes as follows. The approximate probability that the project matures over a period of length dt is ϕdt, and the default probability is $1 - \hat{p}_i(\theta_t, D_t)$ within this period. Given there is zero recovery upon default, the promised payoff to creditors $1 + y_t dt$ needs to satisfy the break-even condition

$$1 = (1 - \phi dt)(1 + ydt) + \hat{p}_i \phi dt(1 + ydt) + (1 - \hat{p}_i)\phi dt \times 0.$$

The expression for y_t follows by ignoring higher-order terms.

[15] Here, $\Pi_i\left(\theta\right)$ is the bank's value function given its retention θ, which implicitly assumes debt issuance has been chosen at the optimal level. Therefore, this constraint involves a fixed point for the value function $\Pi_i\left(\theta\right)$. Hu et al. (2021) use the same technique to reduce the problem's dimensions. A similar problem is analyzed in Abel (2018).

The expression differs from (9) in that the threshold cost for monitoring is replaced by κ_i. We look for a Markov perfect equilibrium in state variable θ_t, henceforth, the intermediation equilibrium.

Definition 2. *An **intermediation equilibrium** is a Markov perfect equilibrium consisting of a price function $q\colon [0,1] \to \mathbb{R}_+$, a trading strategy $(\theta_t)_{t\geq 0}$, a debt-issuance policy $D^*: [0,1] \to \mathbf{R}_+$, and the interest-rate function $y\colon [0,1] \to \mathbf{R}_+$ that satisfy the following:*

1. *For all $\theta \in [0,1]$, the debt-issuance policy $D^*(\theta)$ solves (21).*
2. *Given $\theta_0 = 1$, $(\theta_t)_{t\geq 0}$ is Markov trading strategy that maximizes (22).*
3. *For all $\theta \in [0,1]$, the price $q(\theta)$ satisfies the break-even condition (2).*
4. *For all $\theta \in [0,1]$, the interest rate $y(\theta) := \hat{y}(\theta, D^*(\theta))$ satisfies (19).*

The analysis of the intermediation equilibrium has two steps: debt issuance and loan trading.

4.2.1 Short-term Debt Issuance

Dividing both sides of (21) by $\rho + \phi$, we can define

$$\mathcal{V}(D,\theta) := \Phi\pi_i(\theta) = \Phi\left[\hat{p}_i(\theta, D)\,\theta R_o - \int_0^{\kappa_i} \kappa dF(\kappa)\right] + (1-\Phi)\,D. \tag{23}$$

The term in the bracket is the net payoff to the bank and its creditors: with probability $\hat{p}_i(\theta, D)$, the project succeeds so that they receive θR_o; $\int_0^{\kappa_i} \kappa dF(\kappa)$ is the expected monitoring cost. The last term in (23) is the value from issuing debt. An increase in D reduces the bank's monitoring incentive and therefore reduces the first term. Meanwhile, an increase in D also reduces the bank's funding cost and therefore increases the last term. The optimal D balances the two effects. Under Assumption 2, $\mathcal{V}(D,\theta)$ is concave in D so that the solution is an interior one. Meanwhile, the bank's equity holders' limited liability constraint requires that for any θ, $D \leq \Pi_i(\theta)$. We have the following result.

Lemma 2. *Let $D^*(\theta)$ be the optimal choice of short-term debt. There exists a threshold θ_D where $D^*(\theta) = \Pi_i(\theta)$ if $\theta < \theta_D$.*

According to Lemma 2, the bank finances the loan using both short-term debt and bank capital when θ is relatively high. When θ is relatively low, the loan is only financed via short-term debt. This result implies that higher levels of bank capital are associated with more retention and monitoring.

4.2.2 Trading

Next, we turn to the maximization problem (22) and study how an intermediating bank trades its loans over time. Following similar steps in the certification equilibrium, the term $\dot{\theta}\left(\Pi_i'(\theta) - q_i(\theta)\right)$ must vanish in the smooth-trading region, so the bank's continuation value satisfies the HJB:

$$\rho\Pi_i\left(\theta\right) = \phi\Big[\pi_i\left(\theta\right) - \Pi_i\left(\theta\right)\Big], \tag{24}$$

and the equilibrium price is determined by the indifference condition $\Pi_i'\left(\theta\right) = q_i\left(\theta\right)$. The trading strategy follows from the investors' zero-profit condition that is similar to (13):

$$0 = \phi\Big[d_i\left(\theta\right) - q_i\left(\theta\right)\Big] + \dot{\theta}q_i'\left(\theta\right) \Longrightarrow \dot{\theta} = -\phi\frac{d_i\left(\theta\right) - q_i\left(\theta\right)}{q_i'\left(\theta\right)}. \tag{25}$$

Applying the Envelope Theorem in Milgrom and Segal (2002), we get

$$\pi_i'(\theta) = d_i(\theta) + \left[f\left(\kappa_i\right)\frac{\partial\kappa_i}{\partial\theta}\Delta\right]D^*(\theta) + z(\theta)\Pi_i'(\theta), \tag{26}$$

where $z(\theta)$ is the Lagrange multiplier of the debt-issuance constraint $D \leq \Pi_i(\theta)$.[16] The equilibrium trading rate becomes

$$\dot{\theta} = \phi\frac{(1-\Phi)(1-p\left(\theta\right))R_o + \Phi z(\theta)\left(\Pi_i'(\theta) - R_o\right)}{\Phi\pi_i''\left(\theta\right)}. \tag{27}$$

Lemma 3. *For any θ, $\dot{\theta} > 0$ holds under the optimal short-term debt issuance policy $D\left(\theta\right) = D^*\left(\theta\right)$.*

To complete the characterization of the equilibrium, we need to consider the case in which the bank trades an atom of loans. Similar to the certification equilibrium, we can show a unique $\theta_\dagger$ exists such that the bank sells off all the remaining loans at a price $q\left(0\right) = p_L R_o$ if $\theta < \theta_\dagger$. However, when $\theta = \theta_\dagger$, unlike in the certification equilibrium, the mixed strategy is no longer needed. Instead, the bank buys the loan smoothly so that θ will increase toward $\theta = 1$. Therefore, the price of the loan satisfies $q_i(\theta_\dagger) = \Phi\pi_i'(\theta_\dagger)$.[17] The following proposition describes the results.

Proposition 3 (Intermediation Equilibrium)**.** *A unique* ***intermediation equilibrium*** *exists in which the bank retains $\theta = 1$ until the project matures and issues short-term debt $D = D^*(1)$. For all*

[16] The expression for $z(\theta)$ is available in equation (43) of the appendix.

[17] Both in the case of certification and intermediation, the price function $q(\theta)$ is discontinuous. Whereas in certification, the bank trades toward the discontinuity point (i.e., $\dot{\theta}(\theta_*+) < 0$), in intermediation, the bank trades away from the discontinuity point (i.e., $\dot{\theta}(\theta_\dagger+) > 0$). The construction of the equilibrium (and the analysis of the bank's optimal control problem) is simpler in this latter case because the trajectory of θ_t does not "see" the discontinuity.

$\theta \in [0,1)$, there exists a unique optimal debt issuance policy $D^*(\theta)$, and $D^*(\theta) < \Pi_i(\theta)$ iff $\theta > \theta_D$. For all $\theta \in [0,1]$, the interest rate $y(\theta)$ satisfies $y(\theta) = \phi\left(1 - p_L - F\left(\Delta\left(R_o\theta - D^*(\theta)\right)\right)\Delta\right)$. For $\theta \in [\theta_\dagger, 1)$, the bank buys loans following (27) until θ_t reaches one. For $\theta \in (0, \theta_\dagger)$, the bank immediately sells all the remaining loans. The equilibrium loan price is

$$q_i(\theta) = \begin{cases} \Phi\pi_i'(\theta) & \theta \geq \theta_\dagger \\ p_L R_o & \theta < \theta_\dagger, \end{cases} \tag{28}$$

where $\pi_i'(\theta)$ is defined in (26).

Proposition 3 shows that an intermediation bank finds it optimal to retain the loan and issues short-term debt against it. In other words, the bank prefers issuing short-term debt against the loan over directly selling it. The reason is that investors would increase the interest rate on the short-term debt had the bank reduced its retention.

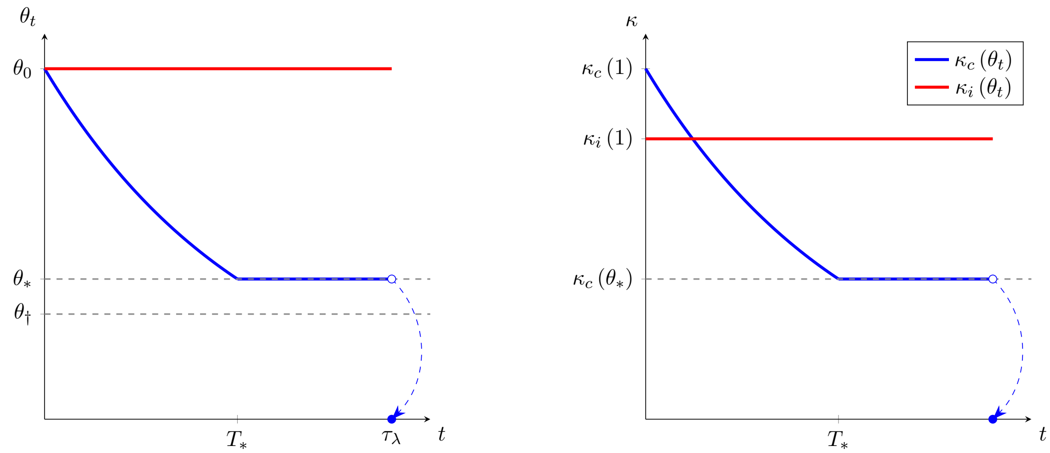
5 Equilibria Comparison, Mechanism, and Discussion

5.1 Comparing Retention and Monitoring

Figure 4 compares the bank's loan retention and the implied monitoring intensity after a loan is originated. In both panels, the blue lines describe the certification equilibrium, whereas the red ones describe the intermediation equilibrium. In the certification equilibrium, the bank with retention θ_0 first sells the loan smoothly. After θ_t reaches θ_*, the bank sells off all the remaining loans at time ν, following a stochastic delay. By contrast, an intermediation bank starts with $\theta_0 = 1$ and always retains the loan until the project matures.

The different patterns in retention also imply different probabilities of monitoring. Specifically, a comparison between the threshold monitoring costs κ_c in (3) and κ_i in (4) shows that under the same level of retention θ, an intermediating bank monitors less due to its outstanding debt D. Meanwhile, an intermediating bank could have higher retention θ, so the overall comparison of monitoring is ambiguous. The right panel of Figure 4 plots the monitoring threshold $\kappa_c(\theta_t)$ and $\kappa_i(\theta_t)$ when both banks retain the entire loan at $t = 0$, i.e., $\theta_0 = 1$.[18] A higher threshold is associated with a higher probability of monitoring if the project matures at time t. Early on, κ_c is higher, but eventually, κ_c falls below κ_i. Our model thus predicts that a certifying bank conducts more monitoring if the loan matures early, whereas an intermediating bank conducts more monitoring if the loan matures late.

[18] The blue line will shift downwards if the certification bank starts with $\theta_0 < 1$.



(a) Retention (b) Monitoring

Figure 4: Retention and Monitoring Dynamics

To understand the different retention patterns between the two structures, we compare the marginal value of retention to the bank's payoff in the smooth trading region. Given $\Pi_c(\theta) = \Phi\pi_c(\theta)$ and $\Pi_i(\theta) = \Phi\pi_i(\theta)$, we directly compare $\pi_c'(\theta)$ with $\pi_i'(\theta)$. For certification, (7) shows $\pi_c'(\theta) = d_c(\theta)$, so the marginal value equals expected cash flows per loan share. However, $\Phi\pi_c'(\theta) = \Phi d_c(\theta)$ falls below investors' $d_c(\theta)$ due to discount rate differences, making the bank an inefficient loan holder and creating gains from trade. While all creditors benefit from monitoring, the bank alone bears the cost. This monitoring externality implies when selling, the bank does not internalize the reduced the value of others' loan stakes, leading to suboptimal monitoring over time under certification.

In intermediation, (26) shows an increase in θ leads to two additional benefits. The first benefit is characterized by the term $\left[f\left(\kappa_i\right)\frac{\partial\kappa_i}{\partial\theta}\Delta\right]D^*$, which shows the marginal effect of retention θ on $f\left(\kappa_i\right)\frac{\partial\kappa_i}{\partial\theta}\Delta$, the incremental probability that debt will be repaid. As a result, more retention enables the bank to issue *cheaper* debt. The second benefit is captured by the last term $z(\theta)\Pi_i'(\theta)$, where $z(\theta) > 0$ if $D^*\left(\theta\right) = \Pi_i(\theta)$. Here, an increase in θ relaxes the constraint on debt issuance so that the bank can issue *more* debt. Note that Lemma 3 continues to hold even if this second benefit disappears, that is, even if $z(\theta) = 0$ so that the debt-issuance constraint is slack. The two additional benefits associated with an increase in θ lead to the opposite result in intermediation when compared to certification. The bank wants to increase its retention because more retention enables an intermediating bank to issue cheaper and more short-term debt. The role of short-term debt is to help the bank internalize the monitoring externality. Indeed, the interest rate of short-term debt reflects the probability of monitoring so that the bank and its creditors share both the benefits and costs of monitoring. In other words, short-term debt creates a market for the services offered by the bank, that is, monitoring, to be fairly priced and constantly repriced.

Bank's choice between certification and intermediation. It is easily shown that $\Pi_i(\theta) > \Pi_c(\theta)$ holds for any θ. Therefore, given any θ, if the bank can choose between the two structures, i.e., if the bank cannot commit to its financial structure at $t = 0$, it will prefer intermediation over certification. The reason is that in our model, debt is short-term (and there is no bankruptcy cost). Under short-term debt, there is no commitment issue in debt issuance. Because short-term debt resolves the commitment problem, it is better than committing to a fixed level of debt (which is zero in certification). If, instead, the bank issues long-term debt, the ability to issue debt is not necessarily better for the bank than committing to some fixed level of debt (such as zero in certification). This result has been shown by Admati et al. (2018) and DeMarzo and He (2021) in the context of leverage-ratchet effect.

5.2 Free-riding and Short-term Debt

Our analysis has thus far highlighted the crucial feature of short-term debt: that the interest rate can adjust as soon as the bank changes its retention to reflect the credit risks borne by creditors. By doing so, short-term debt resolves the free-riding problem in bank monitoring. This subsection further explores the mechanism by introducing long-term debt and/or loss-absorbing equity to an intermediating bank.

Intermediating Bank with Long-term Debt and Loss-Absorbing Equity

Drechsler et al. (2021) argue that for traditional banks, even though deposits are nominally short-term, they effectively constitute long-term financing. We solve the intermediation equilibrium with two modifications. First, if the project fails and generates nothing, the bank is able to pay the creditors up to $X \in (0, D)$. One can think about X as the level of the bank's risk-absorbing equity or the liquidity required to put aside in case of bank failure. Second, the bank can issue only long-term debt $D_0 = D$ at $t = 0$, that is, debt that only matures with the project at τ. After $t = 0$, the bank is no longer allowed to issue any further debt. Following these modifications, the bank's incentive compatibility constraint in monitoring becomes

$$p_H(\theta R_o - D) + (1 - p_H)(-X) - \kappa \geq p_L(\theta R_o - D) + (1 - p_L)(-X) \Rightarrow \kappa \leq \kappa_i := \Delta(R_o\theta + X - D).$$

The definition for $\hat{p}_i(\theta, D)$ stays unchanged from (18), whereas $\hat{\pi}_i(\theta, D)$ becomes

$$\hat{\pi}_i(\theta, D) = \hat{p}_i(\theta, D)(\theta R_o - D) + (1 - \hat{p}_i(\theta, D))(-X) - \int_0^{\kappa_i} \kappa dF(\kappa)$$

to include equity holders' losses when the project fails. Because $D_0 = D$ is only chosen at $t = 0$, we suppress these functions' dependence on D, and therefore refer to $\kappa_i(\theta)$, $p_i(\theta)$, and $\pi_i(\theta)$, respectively. After $t = 0$, the bank only chooses its trading strategy and solves

$$\max_{\{\theta_t\}_{t\geq 0}} \mathbb{E}\left[\int_0^\infty e^{-(\rho+\phi)t}\Big(\phi\pi_i(\theta_t)\,dt + dG(\theta_t)\Big)\right],$$

which is isomorphic to the certifying bank's problem in (8). Therefore, the solution method, as well as the equilibrium outcomes, follow the certification equilibrium in subsection 4.1. In particular, it continues to hold that after $t = 0$, the bank sells its loans and reduces its retention over time. We leave the rest of the analysis to the appendix.

Intermediating Bank with Short-term Debt and Loss-Absorbing Equity

Next, let us turn to the model where the intermediating bank issues short-term debt. If the bank is mandated to issue riskless short-term debt so that a constraint $D \leq X$ is imposed, then $\hat{y}(\theta, D) = 0$. In equilibrium, equity holders will always issue up to the limit so that $D^*(\theta) = X$. The rest of the equilibrium follows exactly the certification equilibrium given that there are no more decisions of debt issuance.

Without the constraint of $D \leq X$, it is never optimal for the bank to issue riskless debt. In other words, the endogenous choice of debt always satisfies $D > X$, with an interest rate $\hat{y}(\theta, D) = \phi\big(1-\hat{p}(\theta, D-X)\big)\left(1-\frac{X}{D}\right)$.[19] Let us define $\tilde{D} \equiv D-X$, then the problem becomes similar to the intermediation equilibrium, except that the bank chooses $\tilde{D}$ as opposed to D. Therefore, the equilibrium in this case is similar to the intermediation equilibrium described in Proposition 3.

Summary. The analysis in this subsection shows the following results. First, an intermediating bank that issues long-term debt and has loss-absorbing equity behaves as a certifying bank. Second, an intermediating bank that issues short-term debt that is not exposed to credit risks also acts as a certifying bank. Third, with loss-absorbing equity, an intermediating bank optimally chooses a level of short-term debt that is exposed to credit risks. Taken together, these exercises demonstrate that short-term debt can resolve the commitment problem only if it is exposed to credit risks.

[19] Following the same step as in the derivation of equation (23), we can write the bank's payoff function in terms of net debt $\tilde{D} \equiv D - X$:

$$\mathcal{V}(\tilde{D}, \theta) := X + \Phi\left[\hat{p}_i(\theta, \tilde{D})\theta R_o - \int_0^{\kappa_i} \kappa dF(\kappa)\right] + (1-\Phi)\,\tilde{D}.$$

5.3 A Linear-Quadratic Example

This subsection presents a linear-quadratic example that allows us to obtain closed-form solutions for the initial payoffs. With the closed-form solutions, we also study the entrepreneur's initial choice between certification and intermediation at time 0. In particular, we specialize the analysis to the case in which the probability of success is $p_L = 0$ if the entrepreneur shirks and the monitoring cost $\tilde{\kappa}$ follows the uniform distribution on $[0, \bar{\kappa}]$. In addition, we assume $\bar{\kappa} \geq \Phi\Delta^2 R_o$, which guarantees the bank never monitors with probability one. In this special case, the bank will never sell off all its loans atomistically (that is, $\theta_* = 0$ and $T_* \to \infty$ in the certification equilibrium), because the resulting price will be zero.

In certification, for any initial θ_0, the amount that the entrepreneur can borrow at $t = 0$ and her payoff from the project maturing are

$$
\begin{aligned}
L_c(\theta_0) &= \Pi_c(\theta_0) + q_c(\theta_0)(1-\theta_0) = \frac{\Phi}{2\bar{\kappa}}(\Delta R_o \theta_0)^2 + \frac{\Phi}{\bar{\kappa}}(\Delta R_o)^2 \theta_0 (1-\theta_0) \\
V_c(\theta_0) &= \Phi B - \frac{\Phi}{2-\Phi}\frac{\Delta R_o \theta_0}{\bar{\kappa}}(B-b).
\end{aligned}
$$

Define $\theta_0^* = 1 - \frac{1}{2-\Phi}\frac{B-b}{\Delta R_o}$, which maximizes the entrepreneur's overall payoff $V_c(\theta_0) + L_c(\theta_0)$. The entrepreneur chooses θ_0 to maximize $V_c(\theta_0) + L_c(\theta_0)$, subject to the feasibility constraint that $L_c(\theta_0) \geq I - A$. Given that $L_c(\theta_0)$ is increasing in θ_0, it is only feasible to finance the project if $L_c(1) \geq I - A$. Assuming this is the case, we can solve for the optimal θ_0 in closed form:

$$
\begin{cases}
\theta_0^* & \text{if } I - A \leq L_c(\theta_0^*) \\
\theta_0^{\min} & \text{if } I - A \in (L_c(\theta_0^*), L_c(1)],
\end{cases}
$$

where $\theta_0^{\min}$ is the minimum value of θ_0 satisfying the feasibility constraint $L_c(\theta_0) \geq I - A$.

In the case of intermediation, the uniform distribution leads to a result that the optimal amount of short-term debt without the issuance constraint $D \leq \Pi_i(\theta)$ is a constant $\frac{1-\Phi}{\Phi}\frac{\bar{\kappa}}{\Delta^2}$. We show in the appendix that the issuance constraint is slack if Φ is sufficiently high.[20] Given that $\theta_t \equiv \theta_0 = 1$ in the intermediation equilibrium, the amount that the entrepreneur is able to borrow at $t = 0$ and

[20]The optimal debt issuance is

$$
D^*(\theta) = \min\left\{\frac{1-\Phi}{\Phi}\frac{\bar{\kappa}}{\Delta^2}, \sqrt{\left(\frac{\bar{\kappa}}{\Delta^2}\right)^2 + (R_o\theta)^2} - \frac{\bar{\kappa}}{\Delta^2}\right\}.
$$

The constraint in debt issuance binds at $\theta = 1$ if and only if $\Phi < \sqrt{\frac{\left(\bar{\kappa}/\Delta^2\right)^2}{R_o^2 + \left(\bar{\kappa}/\Delta^2\right)^2}}$

her payoff from the project maturing are

$$L_i(1) = \Pi_i(1) = \Phi \frac{\Delta^2 \left[R_o^2 - (D^*(1))^2\right]}{2\bar{\kappa}} + (1-\Phi)D^*(1)$$
$$V_i = \Phi B - \Phi \frac{\Delta (R_o - D^*(1))}{\bar{\kappa}} (B - b).$$

Initial Choice

How does the entrepreneur choose between certification and intermediation? First, note that the entrepreneur is always able to borrow more in intermediation. Therefore, if $I - A \in (L_c(1), L_i(1))$, only intermediation is feasible to the entrepreneur. For the remainder of this subsection, we focus on the case of $I - A < L_c(\theta_0^*)$ so that the entrepreneur can borrow enough in certification under the level of θ_0 that maximizes her initial payoff. We have the following result.

Proposition 4. *Under uniform distribution and $p_L = 0$, define $\Delta W \equiv (V_c(\theta_0^*) + L_c(\theta_0^*)) - (V_i + L_i(1))$ as the difference in the entrepreneur's $t = 0$ payoff between certification and intermediation. If $L_c(\theta_0^*) > I - A$, then:*

1. *ΔW monotonically increases in the entrepreneur's private benefit B, the project maturity rate ϕ, and the difference in success probabilities Δ.*
2. *ΔW monotonically decreases in the upper bound on monitoring costs $\bar{\kappa}$.*
3. *There exist thresholds B, ϕ^*, $\bar{\kappa}^*$ and Δ^* such that:*
   - *For $B > B^*$, $\phi > \phi^*$, $\bar{\kappa} < \bar{\kappa}^*$, or $\Delta > \Delta^*$, certification is preferred ($\Delta W > 0$).*
   - *For $B < B^*$, $\phi < \phi^*$, $\bar{\kappa} > \bar{\kappa}^*$, or $\Delta < \Delta^*$, intermediation is preferred ($\Delta W < 0$).*

The result that ΔW increases with B is reminiscent of previous work in corporate governance on the potential of over-monitoring by large shareholders (Pagano and Röell, 1998).[21] In particular, while bank monitoring introduces a positive externality to investors, it also imposes a negative externality on the entrepreneur by restricting her from choosing the high private benefit B. From the perspective of the entrepreneur, she cares not only about the market value of the project but also her benefits as the manager. Therefore, a level of monitoring that maximizes the firm value may be excessive to the entrepreneur. Our result implies that entrepreneurs with larger private benefits (or control rents) tend to borrow from certifying banks.

The rest of the comparative static analysis is straightforward. Projects with a higher ϕ mature earlier on average, in which case a certifying bank still has a relatively large retention and is

[21] See section 9.2.2 in Tirole (2010) for a summary of this literature.

therefore preferred. Both Δ and $\bar{\kappa}$ affect the bank's monitoring decision sensitivity to its retention θ. A higher Δ implies that the bank's monitoring incentives are higher for a given θ (see equations (3) and (4)). Therefore, the decision to monitor is less sensitive to its retention. As a result, certification is relatively preferable. Finally, an increase in $\bar{\kappa}$ is interpreted as an increase in the expected monitoring cost, in which case more retention is needed. Consequently, intermediation becomes relatively preferable.

5.4 Related Literature

Short-Term Debt. In our paper, short-term debt resolves the commitment issue through a repricing mechanism whereby the bank's retention choice directly affects the cost (and also the amount) of short-term debt. The role of short-term debt in aligning a bank's incentives has been discussed by the previous literature in banking (Calomiris and Kahn, 1991; Diamond and Rajan, 2001), which emphasizes the demandable feature of debt and the externalities from depositor runs. Calomiris and Kahn (1991) is about stopping a crime in progress through a run, and the prospect of a run creates a reward for information acquisition. In our paper, there is no run and, hence, no need to reward information acquisition. In Diamond and Rajan (2001), there is no crime in progress to be stopped. Instead, uninformed depositors just have to run when being held up — they are solving a severe incentive/rent extraction problem. Our paper is similar in that it also solves the incentive problems by bankers, but different in that repricing with no runs also gives the bank enough incentives to monitor. Our results are also related to Flannery (1994), which emphasizes the timing between investment and debt issuance. By contrast, the mechanism in our model relies crucially on the mismatch of liabilities and assets. The repricing of short-term debt prevents reduced monitoring by the financial intermediary.[22]

Our model's role of short-term debt is closely related to the leasing solution to the durable goods monopoly problem. In particular, a monopolist can overcome the commitment problem by renting the good rather than selling it (Bulow, 1982). Intuitively, the short-term nature of the rental contract does not allow the monopolist to take advantage of early buyers. Instead, any change in rental prices simultaneously affects all buyers, eliminating the monopolist's temptation to discriminate against buyers over time. The associated commitment problem is, therefore, resolved. In our context, an intermediating bank that issues short-term debt against the project's cash flows can be thought of as a rentor of the claims. The short-term nature of the contract implies that any change in retention is immediately priced by all investors, which eliminates the bank's incentives to sell loans.

[22]We are particularly grateful to Raghuram Rajan for pointing out and interpreting the differences between the three papers discussed in this paragraph.

Loan Sales and Securitization in Banking. This paper is also related to a large literature on bank monitoring, loan sales, and securitization. Parlour and Plantin (2008) show that the informational advantage acquired via bank screening could lead to illiquidity in the secondary market. Our paper is dynamic, and the source of illiquidity comes from the lack of commitment rather than information asymmetry. The commitment problem is also mitigated when claims are collateralized. Rampini and Viswanathan (2019) emphasize the advantage of intermediaries in collateralizing claims. In their paper, certification and intermediation are equivalent.

Durable-goods monopoly. The lack of commitment problem was initially recognized by Coase (1972) in the context of durable goods monopoly. In our context, one can interpret the bank as the monopolist and the durable goods as claims to the cash flows of the entrepreneur's project. This paper belongs to the more recent literature that applies the related insights to corporate finance and banking. The certification equilibrium in our model is closely related to DeMarzo and Urošević (2006), who study a large shareholder's tradeoff between monitoring and diversification without commitment. Admati et al. (2018) and DeMarzo and He (2021) study the problem when a borrower cannot commit to its debt level, which leads to the leverage-ratchet effect. A main insight of this literature is that the Coasian force leads to no gains at all, which has also been shown in the context of bargaining by Fuchs and Skrzypacz (2010) and Daley and Green (2020). Similarly, our certification bank does not benefit from its ability to sell loans.

6 Extensions and Empirical Relevance

6.1 Extensions

Lender Dispersion

In the benchmark model, there are two differences between certification and intermediation. We have highlighted the importance of the intermediating bank's reliance on short-term debt. In addition, as depicted in Figure 1, another difference is that the intermediating bank stands as the sole lender, whereas the certifying bank is just one among several lenders. We now show that this latter difference does not impact our qualitative results.

First, let us consider the model in which the certifying bank is also the only lender. This corresponds to the model in subsection 4.1 with the extra constraint that $\theta_0 = 1$. The equilibrium dynamics in loan sales, retention, monitoring, and pricing stay qualitatively unchanged: the certifying bank sells the loan gradually until θ_t reaches θ_*, after which the bank sells the remaining loans at a Poisson rate λ.

Second, let us consider the model in which the intermediating bank can borrow from multiple lenders. This corresponds to the model in subsection 4.2, but $\theta_0 \in [0,1]$ is optimally chosen by the entrepreneur. In general, the optimal solution can be either $\theta_0 < 1$ or $\theta_0 = 1$. For $\theta_0 < 1$, as shown in Lemma 3, the intermediating bank will buy loans over time until θ_t reaches one, after which it retains the loan until the project matures. The result that banks buy back loans to increase retention could also be interpreted as banks issuing additional loans to the same borrower to increase its overall exposure. The main takeaway is that the bank has increased exposure to the borrower and more skin in the game over time.

The exercises here relate to Bolton and Scharfstein (1996), who highlight one benefit of having dispersed lenders. They show that borrowing from multiple creditors can discipline a firm's manager by reducing the rents to renegotiation after a strategic default. In our model, the bank has a commitment problem that reduces the ex-post incentives in monitoring, and lender dispersion is not the key reason that resolves the commitment problem.

Short-term debt pricing

In our model, short-term debt is fairly priced. In practice, however, short-term bank debt benefits from explicit or implicit guarantees (e.g., FDIC insurance) and is not fairly priced (Drechsler et al., 2021). We consider an extension where the bank only partially passes through interest rates to their main short-term debt holders due to these guarantees. One can think of these guarantees as either deposit insurance or the implicit guarantee from a government bailout.[23] We show that once the guarantee becomes sufficiently large, an intermediating bank no longer retains but instead sells loans over time, just as a certifying bank.

Specifically, we assume the bank only needs to pay a fraction ξ of the interest rate so that equation (19) becomes

$$y_t = \phi\xi\left(1 - \hat{p}_i\left(\theta, D\right)\right),$$

where $\xi \in (0,1)$. The analysis follows that in subsection 4.2, in which short-term debt issuance and trading are solved sequentially.

Proposition 5. *A $\xi_\dagger$ exists such that in the intermediation equilibrium, $\dot{\theta} < 0$ if $\xi < \xi_\dagger$ for θ sufficiently large.*

[23]In the U.S., deposit insurance takes the form of a maximum guaranteed amount that has been \$250,000 since 2010. There is a one-to-one mapping between the maximum insurance amount and the parameter ξ introduced later on. To see this, note one can think about the interest rate as $y_t = 0$ for deposits below $\$250,000$ but following (19) for deposits above the limit. Our parameter ξ captures the fraction of deposits that are above the limit.

If the short-term debt is mostly guaranteed by the government, the bank has reduced incentives to retain its loans. If θ_t is high enough, the debt-issuance constraint is slack $(D^* (\theta) < \Pi_i (\theta))$, Proposition 5 shows that the bank sells loans over time. In the extreme case of $\xi = 0$, the interest rate is independent of the bank's retention (and therefore monitoring), and the results go back to certification. If θ_t is low, the debt-issuance constraint binds $(D^* (\theta) = \Pi_i (\theta))$, then an increased retention still allows the bank to issue more short-term debt. In this case, the bank may still retain loans.

Continuous Monitoring Technology

In our model, the monitoring decision is only made when the project matures. A more realistic model would require the monitoring decision to be made continuously. We consider such an extension and show that our main result is robust in this setup.[24]

Suppose the project generates a payoff of R at maturity with a probability of 1 unless it fails before that maturity event arrives. The failure probability depends on the bank's monitoring effort: before maturity, the project fails with an arrival rate $\eta(1 - m_t)$, where m_t is the bank's monitoring effort chosen at time t. The cost of monitoring is given by a convex function $h(m_t) = \frac{1}{2} k m_t^2$. Note that in this model, monitoring directly affects the arrival rate of the cash flows.

The main results in our main model carry over in this extension. Specifically, we show that a certification bank will gradually sell its loans over time, whereas an intermediating bank never trades. The reason, again, is due to the role of short-term debt, which gets constantly repriced. Detailed analysis is supplemented in Appendix E.3.

6.2 Empirical Relevance

Next, we provide a discussion of the paper's empirical relevance. Given that the model focuses on borrowers with credit risks, we skip the large literature that focuses on mortgage lending by banks and non-bank financial institutions.

Our study is based on the empirical finding that there exists a strong correlation between bank monitoring activities and the bank's retention. Gustafson et al. (2021) use site visits, third-party evaluations, and demands for loan-specific information as proxies for bank monitoring. They show that in syndicated loans, monitoring increases with the lead bank's retention. Moreover, they observe an inverse relationship between bank monitoring intensity and loan interest rates. Focarelli et al. (2008) present evidence of a certification effect: syndicated loans tend to have lower interest rates when a greater portion of the loan facility is held by the arranger. In addition, Wang and

[24] We are grateful to an anonymous referee for this suggestion.

Xia (2014) banks tend to exert less effort in post-loan monitoring when they have the option to securitize loans.

Our model predicts that the price of the loan in the secondary market is negatively correlated with banks' skin in the game. Empirically, Irani et al. (2021) collect data on secondary market pricing of loan sales and show that syndicated loans with higher nonbank funding experienced greater downward pressure on secondary market prices during the recent financial crisis.

Our model predicts certifying banks, i.e., those who do not face constant capital market discipline such as those from short-term creditors, will sell their loans and reduce monitoring. In practice, a bank typically has multiple outstanding loan facilities to a single borrower (Term Loan A, Term Loan B, and revolver). Therefore, one should interpret the one loan in our model as the combination of the bank's credit exposure to a borrower. Our model predicts that a certifying bank without any commitment to retention will gradually reduce its overall exposure to the borrower. Meanwhile, intermediating banks, i.e., those with short-term debt constantly repriced, are subject to discipline. Therefore, the intermediation bank should not be interpreted as a traditional commercial bank that relies on uninformed and insured retail deposits. Rather, this bank could be interpreted as institutions such as shadow banks, asset-backed commercial paper conduits (ABCP Conduit), or structured investment vehicles (SIV), which rely largely on short-term funding from institutional investors. Relatedly, one should not interpret investors as traditional retail depositors protected by FDIC insurance. Instead, they should be considered as institutional investors, such as non-banks or short-term wholesale funding. Our paper, therefore, emphasizes the role of informed creditors' discipline in bank lending. Irani and Meisenzahl (2017) find that when banks are more exposed to disruptions in the short-term wholesale funding market in the crisis, they have a higher probability of selling loan shares, and this depressed loan prices in the secondary market. Specifically, they show that the result is driven by the bank's exposure to short-term funds, including repos and interbank borrowing.

Our paper provides an explanation for why intermediaries rely on short-term debt. Jiang et al. (2020) document that shadow banks originate long-term loans by issuing short-term debt to a few informed lenders. In our model, banks should be more broadly interpreted as any financial institution capable of lending and monitoring. In that case, our paper can be related to the recent empirical studies that document the role of non-bank lenders and their substitution for bank lending. Specifically, Gopal and Schnabl (2022) study the borrowing decisions by small businesses and document the importance of nonbank lenders such as finance companies. Chernenko et al. (2022) study borrowing decisions by middle-market firms and document the prevalence of nonbank financial intermediaries in this market. To the best of our knowledge, no empirical research has studied the ex-post performance of loans extended by non-bank lenders and how it relates to these lenders' monitoring intensities. Our paper predicts that the performance crucially depends on the liability structure of these non-bank lenders, a hypothesis to be tested in the future.

7 Final Remarks

This paper develops a theory of intermediary financing when banks cannot commit to the retentions on the balance sheet. Our main message is that the intermediary's liability structure impacts the dynamics of lending and monitoring. A (certifying) bank that finances using long-term claims, such as long-term debt and equity, has incentives to sell loans over time, leading to a gradual reduction in monitoring. By contrast, a (intermediating) bank that finances itself by issuing short-term debt does not have incentives to sell loans. As a result, certification is associated with a lower lending capacity compared to intermediation. Certification has more monitoring during the early periods after loan origination, whereas intermediation has more monitoring during the later periods.

Throughout the paper, we follow Holmstrom and Tirole (1997) by assuming that all projects financed by an intermediary are perfectly correlated and thus abstract from the bank's ability to pool assets and diversify the risk (Diamond, 1984; DeMarzo, 2005). In the case of many loans, we can interpret the monitoring decision as the bank's investment in its monitoring technology (which might include the adoption of advanced information technology, the hiring of qualified loan officers, more efficient internal governance, etc.). Such investments improve the bank's ability to control bank-specific risk in its portfolio, which cannot be eliminated by diversification.

In our model, informed investors monitor the bank's balance sheet. It is important, though, that investors cannot write binding contracts based on their assessments of the riskiness of the bank. As argued by Flannery (1994), banks specialize in financing non-marketable, informationally intensive assets, and the composition of these assets change rapidly with new business opportunities. As a result, these assets do not have contractible, easily described risk properties.

Our main focus has been on the bank's ex-post monitoring rather than ex-ante screening (Ramakrishnan and Thakor, 1984). Hu and Varas (2021) shows how zombie lending will emerge in this context when screening takes time, as a relationship bank can signal through either dynamic retention or debt issuance.[25] Given that our focus is on how the bank's liability structure enables commitment to retention, we chose to stay away from these complications introduced by screening.

By focusing on Markov equilibria, we ignore the intermediary's concern for reputation (see Winton and Yerramilli (2021) for some recent work on the role of reputation concerns). While reputation does not directly affect the return to monitoring in our model, it can have an important

[25] See Leland and Pyle (1977) and Ross (1977) for related issues in the static environment.

effect on the dynamics of loan sales. In particular, one can construct an equilibrium in which the commitment problem is mitigated if the intermediary has a long-run reputation (see Ausubel and Deneckere (1989) for a study of the impact of durable good monopolist's reputation concerns and Malenko and Tsoy (2020) in the context of a corporate borrower who cannot commit to its debt level).

References

Abel, A. B. (2018). Optimal debt and profitability in the trade-off theory. *The Journal of Finance 73*(1), 95–143.

Admati, A. R., P. M. DeMarzo, M. F. Hellwig, and P. Pfleiderer (2018). The leverage ratchet effect. *Journal of Finance 73*(1), 145–198.

Ausubel, L. M. and R. J. Deneckere (1989). Reputation in bargaining and durable goods monopoly. *Econometrica*, 511–531.

Bardi, M. and I. Capuzzo-Dolcetta (2008). *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations.* Springer Science.

Barles, G., A. Briani, E. Chasseigne, and C. Imbert (2018). Flux-limited and classical viscosity solutions for regional control problems. *ESAIM: Control, Optimisation and Calculus of Variations 24*(4), 1881–1906.

Bolton, P. and D. S. Scharfstein (1996). Optimal debt structure and the number of creditors. *Journal of political economy 104*(1), 1–25.

Bulow, J. I. (1982). Durable-goods monopolists. *Journal of political Economy 90*(2), 314–332.

Calomiris, C. W. and C. M. Kahn (1991). The role of demandable debt in structuring optimal banking arrangements. *The American Economic Review*, 497–513.

Chernenko, S., I. Erel, and R. Prilmeier (2022). Why do firms borrow directly from nonbanks? *The Review of Financial Studies 35*(11), 4902–4947.

Coase, R. H. (1972). Durability and monopoly. *The Journal of Law and Economics 15*(1), 143–149.

Daley, B. and B. Green (2020). Bargaining and news. *American Economic Review 110*(2), 428–74.

DeMarzo, P. M. (2005). The pooling and tranching of securities: A model of informed intermediation. *The Review of Financial Studies 18*(1), 1–35.

DeMarzo, P. M. and Z. He (2021). Leverage dynamics without commitment. *The Journal of Finance 76*(3), 1195–1250.

DeMarzo, P. M. and B. Urošević (2006). Ownership dynamics and asset pricing with a large shareholder. *Journal of Political Economy 114*(4), 774–815.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The review of economic studies 51*(3), 393–414.

Diamond, D. W. and R. G. Rajan (2001). Liquidity risk, liquidity creation, and financial fragility: A theory of banking. *Journal of Political Economy 109*(2), 287–327.

Drechsler, I., A. Savov, and P. Schnabl (2021). Banking on deposits: Maturity transformation without interest rate risk. *The Journal of Finance 76*(3), 1091–1143.

Drucker, S. and M. Puri (2009). On loan sales, loan contracting, and lending relationships. *The Review of Financial Studies 22*(7), 2835–2872.

Flannery, M. J. (1994). Debt maturity and the deadweight cost of leverage: Optimally financing banking firms. *The American economic review 84*(1), 320–331.

Focarelli, D., A. F. Pozzolo, and L. Casolaro (2008). The pricing effect of certification on syndicated loans. *Journal of Monetary Economics 55*(2), 335–349.

Fuchs, W. and A. Skrzypacz (2010). Bargaining with arrival of new traders. *American Economic Review 100*(3), 802–36.

Gopal, M. and P. Schnabl (2022). The rise of finance companies and fintech lenders in small business lending. *The Review of Financial Studies 35*(11), 4859–4901.

Gustafson, M. T., I. T. Ivanov, and R. R. Meisenzahl (2021). Bank monitoring: Evidence from syndicated loans. *Journal of Financial Economics 139*(2), 452–477.

Holmstrom, B. and J. Tirole (1997). Financial intermediation, loanable funds, and the real sector. *Quarterly Journal of Economics 112*(3), 663–691.

Hu, Y. and F. Varas (2021). A dynamic theory of learning and relationship lending. *Journal of Finance 76*(4), 1813–1867.

Hu, Y., F. Varas, and C. Ying (2021). Debt maturity management. Technical report.

Irani, R. M., R. Iyer, R. R. Meisenzahl, and J.-L. Peydro (2021). The rise of shadow banking: Evidence from capital regulation. *The Review of Financial Studies 34*(5), 2181–2235.

Irani, R. M. and R. R. Meisenzahl (2017). Loan sales and bank liquidity management: Evidence from a us credit register. *The Review of Financial Studies 30*(10), 3455–3501.

Jiang, E., G. Matvos, T. Piskorski, and A. Seru (2020). Banking without deposits: Evidence from shadow bank call reports. Technical report.

Leland, H. E. and D. H. Pyle (1977). Informational asymmetries, financial structure, and financial intermediation. *The journal of Finance 32*(2), 371–387.

Malenko, A. and A. Tsoy (2020). Optimal time-consistent debt policies. *Available at SSRN 3588163*.

Milgrom, P. and I. Segal (2002). Envelope theorems for arbitrary choice sets. *Econometrica 70*(2), 583–601.

Milgrom, P. and C. Shannon (1994). Monotone comparative statics. *Econometrica: Journal of the Econometric Society*, 157–180.

Pagano, M. and A. Röell (1998). The choice of stock ownership structure: Agency costs, monitoring, and the decision to go public. *The Quarterly Journal of Economics 113*(1), 187–225.

Parlour, C. A. and G. Plantin (2008). Loan sales and relationship banking. *Journal of Finance 63*(3), 1291–1314.

Ramakrishnan, R. T. and A. V. Thakor (1984). Information reliability and a theory of financial intermediation. *The Review of Economic Studies 51*(3), 415–432.

Rampini, A. A. and S. Viswanathan (2019). Financial intermediary capital. *The Review of Economic Studies 86*(1), 413–455.

Ross, S. A. (1977). The determination of financial structure: the incentive-signalling approach. *The bell journal of economics*, 23–40.

Tirole, J. (2010). *The theory of corporate finance*. Princeton University Press.

Wang, Y. and H. Xia (2014). Do lenders still monitor when they can securitize loans? *The Review of Financial Studies 27*(8), 2354–2391.

Winton, A. and V. Yerramilli (2021). Monitoring in originate-to-distribute lending: Reputation versus skin in the game. *The Review of Financial Studies*.

Appendix

A Certification

In this section, we prove Proposition 2.

Lemma 4. *The bank with retention θ never sells a fraction of the loans.*

Proof. Suppose the bank with retention θ sells $\theta - \theta^+$, where $\theta^+ > 0$, and that after this it continues to trade smoothly. Multiple jumps are ruled out without loss of generality. In this case, the overall trading gains are $dG(\theta) + \Pi_c(\theta^+) - \Pi_c(\theta) = (\theta - \theta^+) q_c(\theta^+) + \Pi_c(\theta^+) - \Pi_c(\theta)$, where $dG(\theta)$ is the instant trading gain and $\Pi(\theta^+) - \Pi(\theta)$ are the gains (negative loss) in its continuation value. Block trading is suboptimal as long as

$$\theta = \arg\max_{\theta^+} \left\{ \Pi_c(\theta^+) + (\theta - \theta^+) q_c(\theta^+) \right\}. \tag{29}$$

It is easy to verify that the first order condition is always satisfied at $\theta^+ = \theta$, thus it suffice to show that the second order condition for global optimality is satisfied. $\square$

Next, we show that $L_c(\theta_0)$ is monotonically increasing in θ_0. Specifically, we have

$$L_c'(\theta_0) = \Phi\pi_c'(\theta_0) - \Phi\pi_c'(\theta_0) + (1-\theta_0)\Phi\pi_c''(\theta_0) = (1-\theta_0)\Phi\pi_c''(\theta_0) = (1-\theta_0)\Phi d_c'(\theta_0)$$

where

$$\begin{aligned} d_c(\theta_0) &= R_o p_c(\theta_0) = R_o(p_L + F(\kappa_c)\Delta) = R_o(p_L + F(\Delta R_o \theta_0)\Delta) \\ d_c'(\theta_0) &= f(\Delta R_o \theta_0)(\Delta R_o)^2. \end{aligned}$$

Then, we know

$$L_c'(\theta_0) = (1-\theta_0)\Phi d_c'(\theta_0) = (1-\theta_0)\Phi f(\Delta R_o \theta_0)(\Delta R_o)^2 \geq 0.$$

Verification of Optimality Trading Strategy

In this section, we complete the characterization of the equilibrium by verifying that the equilibrium trading strategy maximizes the bank's payoff given the price function $q(\theta)$. Because the payoff in a mixed strategy equilibrium is given by the payoff of any pure strategy in its support, we can restrict attention to pure strategies in the verification of optimality. A trading strategy for the bank is given by the right continuous function with left limits. A trading strategy is admissible if it can be decomposed as

$$\theta_t = \int_0^t \dot{\theta}_s^c ds + \sum_{k \geq 0} (\theta_{t_k}^d - \theta_{t_k -}^d),$$

for some bounded function $\dot{\theta}^c_t$. We denote the set of admissible trading strategies by Θ. The bank's optimization problem is to choose $\theta \in \Theta$ to maximize its payoff

$$\Pi^*(\theta_0) = \sup_{\theta\in\Theta} \int_0^\infty e^{-(\rho+\phi)t}\big(\phi\pi(\theta_t) - \dot{\theta}^c_t q(\theta_t)\big)dt - \sum_{k\geq 0} e^{-(\rho+\phi)t_k} q(\theta_{t_k})(\theta^d_{t_k} - \theta^d_{t_k-}). \tag{30}$$

Due to the discontinuity in the price function $q(\theta)$, the Hamilton-Jacobi-Bellman (HJB) equation is discontinuous at θ_*, so we need to resort to the theory of viscosity solutions for the analysis of the bank's problem. Our problem is a particular case of the general class of optimal control problems in stratified domains studied by Barles et al. (2018). Our proof relies on their characterization of the value function using viscosity solution methods. The analysis in Barles et al. (2018) does not consider the case in which the trajectory of the state variable can be discontinuous (impulse control). However, as we show below, we can approximate a trading $\theta_t \in \Theta$ by an absolutely continuous trading strategy with derivative $|\dot{\theta}_t| \leq N$ for some N large enough (the approximation is in the sense that it yields a similar payoff). Thus, we can consider a sequence of optimization problems

$$\Pi^*_N(\theta_0) = \sup_{|\dot{\theta}_t|\leq N} \int_0^\infty e^{-(\rho+\phi)t}\big(\phi\pi(\theta_t) - \dot{\theta}_t q(\theta_t)\big)dt, \tag{31}$$

and verify that, for any $\theta \in [0,1]$, $\Pi^*_N(\theta) \to \Pi(\theta)$, where

$$\Pi(\theta) = \begin{cases} \Phi\pi(\theta) & \text{if } \theta \in [\theta_*, 1] \\ q(0)\theta & \text{if } \theta \in [0, \theta_*) \end{cases}.$$

The following Lemma establishes that we can indeed consider the limit of bounded absolutely continuous strategies.

Lemma 5. *For any $\theta_0 \in [0,1]$, $\lim_{N\to\infty} \Pi^*_N(\theta_0) = \Pi^*(\theta_0)$.*

Proof. Let $\theta^{\epsilon *}_t$ be and ϵ-optimal policy (at this point in the proof, we have not established the existence of an optimal policy). For any $k \geq 0$, let $\Delta_k \equiv \inf\{\Delta > 0 : \theta^{\epsilon *}_{t_k - \Delta} + \text{sgn}(\theta^{\epsilon *}_{t_k} - \theta^{\epsilon *}_{t_k -})N\Delta = \theta^{\epsilon *}_{t_k}\}$ (we can find Δ_k if N is large enough as $|\dot{\theta}^{\epsilon c *}_t| \leq M$ for some finite M). Consider the policy $\hat{\theta}^N_t = \theta^{\epsilon *}_t$ if $t \notin \cup_{k\geq 0}(t_k - \Delta_k, t_k)$, and $\hat{\theta}^N_t = \theta^{\epsilon *}_{t_k - \Delta_k} + \text{sgn}(\theta^{\epsilon *}_{t_k} - \theta^{\epsilon *}_{t_k -})N(t - t_k + \Delta_k)$ if $t \in \cup_{k\geq 0}(t_k - \Delta_k, t_k)$. The difference between the payoff of $\theta_t^{\epsilon *}$ and $\hat{\theta}_t^N$ is

$$\begin{aligned}
\Pi^{\epsilon *}(\theta_0) - \hat{\Pi}_N &= \sum_{k\geq 0}\Bigg\{\int_{t_k-\Delta_k}^{t_k} e^{-(\rho+\phi)t}\big(\phi\pi(\theta_t^{\epsilon *}) - \dot{\theta}_t^{\epsilon c *}q(\theta_t^{\epsilon *}) - \phi\pi(\hat{\theta}_t^N)\big)dt \\
&\quad + \int_{t_k-\Delta_k}^{t_k} e^{-(\rho+\phi)t}\mathrm{sgn}(\theta_{t_k}^{\epsilon *} - \theta_{t_k-}^{\epsilon *})Nq(\hat{\theta}_t^N)dt - e^{-(\rho+\phi)t_k}q(\theta_{t_k}^{\epsilon *})(\theta_{t_k}^{\epsilon d *} - \theta_{t_k-}^{\epsilon d *})\Bigg\} - \epsilon \\
&= \sum_{k\geq 0}\Bigg\{\int_{t_k-\Delta_k}^{t_k} e^{-(\rho+\phi)t}\big(\phi\pi(\theta_t^{*}) - \dot{\theta}_t^{\epsilon c *}q(\theta_t^{*}) - \phi\pi(\hat{\theta}_t^N)\big)dt \\
&\quad + \frac{\theta_{t_k}^{\epsilon *} - \theta_{t_k-\Delta_k}^{\epsilon *}}{\Delta_k}\int_{t_k-\Delta_k}^{t_k} e^{-(\rho+\phi)t}q(\hat{\theta}_t^N)dt - e^{-(\rho+\phi)t_k}q(\theta_{t_k}^{\epsilon *})(\theta_{t_k}^{\epsilon d *} - \theta_{t_k-}^{\epsilon d *})\Bigg\} - \epsilon.
\end{aligned}$$

For all $k \geq 0$, we have that $\Delta_k \downarrow 0$ as $N \to \infty$. It follows that

$$\lim_{\Delta_k \downarrow 0}\frac{1}{\Delta_k}\int_{t_k-\Delta_k}^{t_k} e^{-(\rho+\phi)t}q(\hat{\theta}_t^N)dt = \begin{cases} e^{-(\rho+\phi)t_k}q(\theta_{t_k}^{\epsilon *}-) & \text{if } \theta_{t_k}^{\epsilon *} > \theta_{t_k-}^{\epsilon *} \\ e^{-(\rho+\phi)t_k}q(\theta_{t_k}^{\epsilon *}+) & \text{if } \theta_{t_k}^{\epsilon *} < \theta_{t_k-}^{\epsilon *}. \end{cases}$$

The price function is right continuous so $q(\theta_{t_k}^{\epsilon *}+) = q(\theta_{t_k}^{\epsilon *})$. We can conclude that

$$\lim_{N\to\infty}(\Pi^{\epsilon *}(\theta_0) - \hat{\Pi}_N(\theta_0)) = \sum_{k\geq 0} e^{-(\rho+\phi)t_k}\big(q(0) - q(\theta_*)\big)(\theta_{t_k}^* - \theta_{t_k-}^*)^+\mathbf{1}_{\{\theta_{t_k}^* = \theta_*\}} - \epsilon \leq 0.$$

Because this holds for any $\epsilon > 0$, we can conclude that $\lim_{N\to\infty}(\Pi^*(\theta_0) - \hat{\Pi}_N(\theta_0)) \leq 0$, and given that $\Pi^*(\theta_0) \geq \hat{\Pi}_N(\theta_0)$, we get $\lim_{N\to\infty}\hat{\Pi}_N(\theta_0) = \Pi^*(\theta_0)$. For N large enough, the policy $\hat{\theta}_t^N$ satisfies $|\dot{\hat{\theta}}_t^N| \leq N$ (this can be guarantee because for any ϵ there is M such that $|\dot{\theta}_t^{\epsilon c *}| \leq M$), so its payoff, $\hat{\Pi}_N(\theta_0)$ provides a lower bound to $\Pi_N^*(\theta_0)$, which means that $\lim_{N\to\infty}\Pi_N^*(\theta_0) = \Pi^*(\theta_0)$. □

This shows that the value function converges (pointwise) to the one in the equilibrium under consideration. Hence, we can verify the optimality of the bank's strategy by analyzing the control problem (31). For future reference, recall that the price function in the control problem (31) is given by

$$q(\theta) = \begin{cases} \Phi\pi'(\theta) & \text{if } \theta \geq \theta_* \\ p_L R\,(1-\alpha) & \text{if } \theta < \theta_* \end{cases}, \tag{32}$$

where the threshold θ_* is given by $\Phi\pi(\theta_*) = q(0)\theta_*$. Notice that we are not computing the equilibrium in a model in which the bank is restricted to use absolutely continuous trading strategies with bounded derivative $\dot{\theta}_t$, but rather considering the equilibrium price function in the general case, and then considering a sequence of auxiliary optimization problems to construct the value function. Because the expected payoff of the candidate equilibrium strategy is equal to the value function, it is necessarily optimal.

The Hamilton-Jacobi-Bellman equation (HJB) for the optimization problem (31) is

$$(\rho+\phi)\Pi_N(\theta) - H(\theta, \Pi_N'(\theta)) = 0, \tag{33}$$

where H

$$H(\theta,\Pi_N') \equiv \phi\pi(\theta) + \max_{|\dot{\theta}|\le N}\left\{\dot{\theta}\big(\Pi_N' - q(\theta)\big)\right\}. \tag{34}$$

We guess and verify that, for N large enough, the solution (in the viscosity sense) of the previous equation is

$$\Pi_N(\theta) = \begin{cases} \Phi\pi(\theta) & \text{if } \theta\in[\theta_*,1] \\ e^{-\frac{\rho+\phi}{N}(\theta_*-\theta)}\Phi\pi(\theta_*) + \frac{(\rho+\phi)}{N}\int_\theta^{\theta_*} e^{-\frac{\rho+\phi}{N}(y-\theta)}\left(\Phi\pi(y) - \frac{N}{\rho+\phi}q(0)\right)dy & \text{if } \theta\in[\tilde{\theta}_N,\theta_*) \\ \frac{N}{\rho+\phi}\left(1-e^{-\frac{(\rho+\phi)}{N}\theta}\right)q(0) + \frac{(\rho+\phi)}{N}\int_0^\theta e^{-\frac{(\rho+\phi)}{N}(\theta-y)}\Phi\pi(y)dy & \text{if } \theta\in[0,\tilde{\theta}_N), \end{cases} \tag{35}$$

where $\tilde{\theta}_N$ is the unique solution on $[0,1]$ to the equation

$$\begin{aligned} &\frac{N}{\rho+\phi}\left(1-e^{-\frac{(\rho+\phi)}{N}\tilde{\theta}_N}\right)q(0) + \frac{(\rho+\phi)}{N}\int_0^{\tilde{\theta}_N} e^{-\frac{(\rho+\phi)}{N}(\tilde{\theta}_N-y)}\Phi\pi(y)dy = \\ &\qquad e^{-\frac{\rho+\phi}{N}(\theta_*-\tilde{\theta}_N)}\Phi\pi(\theta_*) + \frac{(\rho+\phi)}{N}\int_{\tilde{\theta}_N}^{\theta_*} e^{-\frac{\rho+\phi}{N}(y-\tilde{\theta}_N)}\left(\Phi\pi(y) - \frac{N}{\rho+\phi}q(0)\right)dy \end{aligned} \tag{36}$$

A.1 Auxiliary Lemmas

Before proceeding with the verification theorem, we provide several Lemmas providing properties of our candidate value function $\Pi_N(\theta)$ that will be later used in the verification

Lemma 6. *If $\Phi\pi_c(1) > p_L R_o > 0$, then there exists a unique $\theta_*\in(0,1)$ solving the equation*

$$\theta_* q_c(0) = \Phi\pi_c(\theta_*) \tag{37}$$

If $p_L = 0$, then $\theta_* = 0$ is the unique solution to (37) on $[0,1]$.

Proof. As $\Phi\pi_c(0)=0$, equation (37) is trivially satisfied at $\theta_*=0$, we want to show that if $\Phi\pi_c(1) > p_L R_o = q_c(0)$, then there is a non trivial solution $\theta_*>0$ that also satisfies equation (37). First, if $\Phi\pi_c(1) > p_L R_o$, then the right hand side of equation (37) is strictly larger than its left hand side evaluated at $\theta_*=1$. Second, as $\Phi\pi_c'(0) < q_c(0)$ it follows that for ε small enough $\varepsilon q_c(0) > \Phi\pi_c(\varepsilon)$. Thus, it follows from continuity that a nontrivial solution exists on $(0,1)$. Uniqueness follows because

$$\begin{aligned} q_c(0) - \Phi\pi_c'(\theta_*) &= \frac{\Phi\pi_c(\theta_*)}{\theta_*} - \Phi p_c(\theta_*)R_o \\ &= \Phi\left[p_c(\theta_*)R_o - \frac{1}{\theta_*}\int_0^{\kappa_c(\theta_*)}\kappa dF(\kappa)\right] - \Phi p_c(\theta_*)R_o < 0, \end{aligned}$$

so the function $\theta q_c(0) - \Phi\pi_c(\theta)$ single crosses 0 from above, which implies $\theta q_c(0) > \Phi\pi_c(\theta)$ on $\theta\in(0,\theta_*)$ and $\theta q_c(0) < \Phi\pi_c(\theta)$ on $\theta\in(\theta_*,1]$. Finally, if $p_L=0$, then $\Phi\pi_c'(0) = q_c(0) = 0$. It follows then from the convexity of $\pi_c(\theta)$ that $\theta_*=0$ is a global maximum of $\theta q_c(0) - \Phi\pi_c(\theta)$, which means that $\theta q_c(0) < \Phi\pi_c(\theta)$ for all $\theta>0$. ☐

Lemma 7. *There is a unique solution $\tilde{\theta}_N \in (0, \theta_*)$ to equation (36).*

Proof. First, we show existence. Given the definition of θ_* and the convexity of $\pi(\theta)$ we have that $\Phi\pi(\theta) < \theta q(0)$ for all $\theta < \theta_*$. Hence,

$$\frac{N}{\rho+\phi}\left(1 - e^{-\frac{(\rho+\phi)}{N}\theta_*}\right) q(0) + \frac{(\rho+\phi)}{N}\int_0^{\theta_*} e^{-\frac{(\rho+\phi)}{N}(\theta-y)}\Phi\pi(y)dy < \Phi\pi(\theta_*).$$

We also have that

$$\begin{aligned} e^{-\frac{\rho+\phi}{N}\theta_*}\Phi\pi(\theta_*) + \frac{(\rho+\phi)}{N}\int_0^{\theta_*} e^{-\frac{\rho+\phi}{N}y}\left(\Phi\pi(y) - \frac{N}{\rho+\phi}q(0)\right)dy &\leq \Phi\pi(\theta_*) - \frac{N}{(\rho+\phi)}\left(1 - e^{-\frac{\rho+\phi}{N}\theta_*}\right)q(0) \\ &\leq \Phi\pi(\theta_*) - \theta_* q(0) = 0. \end{aligned}$$

The existence of a solution follows from the intermediate value theorem. To show uniqueness, we consider the derivative of the difference between the left and the right-hand sides of equation (36) evaluated at θ, which we denote by $G'(\theta)$.

$$\begin{aligned} G'(\theta) = e^{-\frac{(\rho+\phi)}{N}\theta}q(0) + \frac{(\rho+\phi)}{N}\Phi\pi(\theta) - \frac{(\rho+\phi)^2}{N^2}\int_0^{\theta} e^{-\frac{(\rho+\phi)}{N}(\theta-y)}\Phi\pi(y)dy \\ - \frac{(\rho+\phi)}{N}e^{-\frac{\rho+\phi}{N}(\theta_*-\theta)}\Phi\pi(\theta_*) + \frac{(\rho+\phi)}{N}\left(\Phi\pi(\theta) - \frac{N}{\rho+\phi}q(0)\right) \\ - \frac{(\rho+\phi)^2}{N^2}\int_{\theta}^{\theta_*} e^{-\frac{\rho+\phi}{N}(y-\theta)}\left(\Phi\pi(y) - \frac{N}{\rho+\phi}q(0)\right)dy \end{aligned}$$

From here we get that

$$\begin{aligned} G'(\theta_*) &= -\left(1 - e^{-\frac{(\rho+\phi)}{N}\theta_*}\right)q(0) - \frac{(\rho+\phi)^2}{N^2}\int_0^{\theta_*} e^{-\frac{(\rho+\phi)}{N}(\theta_*-y)}\Phi\pi(y)dy + \frac{(\rho+\phi)}{N}\Phi\pi(\theta_*) \\ &\leq \frac{(\rho+\phi)}{N}\left(\Phi\pi(\theta_*) - \theta_* q(0)\right) = 0 \\ G'(0) &= -\frac{(\rho+\phi)}{N}e^{-\frac{\rho+\phi}{N}\theta_*}\Phi\pi(\theta_*) - \frac{(\rho+\phi)^2}{N^2}\int_0^{\theta_*} e^{-\frac{\rho+\phi}{N}y}\left(\Phi\pi(y) - \frac{N}{\rho+\phi}q(0)\right)dy \\ &\geq -\frac{(\rho+\phi)}{N}\Phi\pi(\theta_*) + \left(1 - e^{-\frac{\rho+\phi}{N}\theta_*}\right)q(0) \\ &\geq \frac{(\rho+\phi)}{N}\left(\theta_* q(0) - \Phi\pi(\theta_*)\right) = 0 \end{aligned}$$

Moreover, we get that, for any $\theta \in (0, \theta_*)$, $G(\theta) = 0$ implies

$$\begin{aligned} G'(\theta) &= \frac{2(\rho+\phi)}{N}\left[\Phi\pi(\theta) - \left(\frac{N}{\rho+\phi}\left(1 - e^{-\frac{(\rho+\phi)}{N}\theta}\right)q(0) + \frac{(\rho+\phi)}{N}\int_0^{\theta} e^{-\frac{(\rho+\phi)}{N}(\theta-y)}\Phi\pi(y)dy\right)\right] \\ &\leq \frac{2(\rho+\phi)}{N}\left[\Phi\pi(\theta) - \theta q(0)\right] < 0. \end{aligned}$$

It follows that $G(\theta)$ single crosses 0, so there is a unique solution to the equation $G(\theta) = 0$. $\square$

Lemma 8. *There is $\tilde{N}$ such that, for all $N > \tilde{N}$, $\Pi'_N(\theta) < q(0)$ on $(0, \tilde{\theta}_N)$ and $\Pi'_N(\theta) > q(0)$ on $(\tilde{\theta}_N, \theta_*)$.*

Proof. First, we very that $\Pi'_N(\theta) < q(0)$ on $(0, \tilde{\theta}_N)$. The derivative of $\Pi_N(\theta) - \theta q(0)$ on $(0, \tilde{\theta}_N)$ is given by

$$\begin{aligned}\Pi'_N(\theta) - q(0) &= \frac{\rho+\phi}{N}\Phi\pi(\theta) - \left(1 - e^{-\frac{(\rho+\phi)}{N}\theta}\right) q(0) - \frac{(\rho+\phi)^2}{N^2}\int_0^\theta e^{-\frac{(\rho+\phi)}{N}(\theta-y)}\Phi\pi(y)dy \\ &\leq \frac{\rho+\phi}{N}\Phi\pi(\theta) - \left(1 - e^{-\frac{(\rho+\phi)}{N}\theta}\right) q(0) \leq \frac{\rho+\phi}{N}\left(\Phi\pi(\theta) - \theta q(0)\right) < 0.\end{aligned}$$

The derivative of $\Pi_N(\theta) - \theta q(0)$ on $(\tilde{\theta}_N, \theta_*)$ is given by

$$\Pi'_N(\theta) - q(0) = \frac{\rho+\phi}{N}\left[e^{-\frac{\rho+\phi}{N}(\theta_*-\theta)}\Phi\pi(\theta_*) - \Phi\pi(\theta) + \frac{(\rho+\phi)}{N}\int_\theta^{\theta_*} e^{-\frac{\rho+\phi}{N}(y-\theta)}\left(\Phi\pi(y) - \frac{N}{\rho+\phi}q(0)\right)dy\right]$$

Differentiating the HJB equation, we get that

$$\begin{aligned}\Pi''_N(\theta) &= \frac{(\rho+\phi)}{N}\left(\Pi'_N(\theta) - \Phi\pi'(\theta)\right) \\ \Pi'''_N(\theta) &= \frac{(\rho+\phi)}{N}\left(\Pi''_N(\theta) - \Phi\pi''(\theta)\right).\end{aligned}$$

From here we get that $\Pi''_N(\theta) = 0 \implies \Pi'''_N(\theta) < 0$, so we the function $\Pi'_N(\theta)$ is quasi-concave on $(\tilde{\theta}_N, \theta_*)$. Moreover, $\Pi'_N(\theta_*-) = q(0)$, and

$$\Pi'_N(\tilde{\theta}_N+) = q(0) + \frac{(\rho+\phi)}{N}\left(\Pi_N(\tilde{\theta}_N+) - \Phi\pi(\tilde{\theta}_N)\right) > q(0),$$

so we can conclude that $\Pi'_N(\theta) > q(0)$ on $(\tilde{\theta}_N, \theta_*)$ as long as $\Pi_N(\tilde{\theta}_N+) > \Phi\pi(\tilde{\theta}_N)$, which follows from

$$\begin{aligned}\Pi_N(\tilde{\theta}_N+) - \Phi\pi(\tilde{\theta}_N) &= \Pi_N(\tilde{\theta}_N-) - \Phi\pi(\tilde{\theta}_N) \\ &= \frac{N}{\rho+\phi}\left(1 - e^{-\frac{(\rho+\phi)}{N}\tilde{\theta}_N}\right) q(0) - \Phi\pi(\tilde{\theta}_N) + \frac{(\rho+\phi)}{N}\int_0^{\tilde{\theta}_N} e^{-\frac{(\rho+\phi)}{N}(\tilde{\theta}_N-y)}\Phi\pi(y)dy \\ &\geq \frac{N}{\rho+\phi}\left(1 - e^{-\frac{(\rho+\phi)}{N}\tilde{\theta}_N}\right) q(0) - \Phi\pi(\tilde{\theta}_N) \\ &= \tilde{\theta}_N q_0 - \frac{(\rho+\phi)\tilde{\theta}_N^2}{N} - \Phi\pi(\tilde{\theta}_N) + O\big(1/N^2\big).\end{aligned}$$

$\tilde{\theta}_N q_0 > \Phi\pi(\tilde{\theta}_N)$ because θq_0 single crosses $\Phi\pi(\theta)$ at $\theta_* \geq \tilde{\theta}_N$. Hence, there is $\tilde{N}$ such that, for all $N \geq \tilde{N}$, we have $\Pi_N(\tilde{\theta}_N+) > \Phi\pi(\tilde{\theta}_N)$. $\square$

Lemma 9. *Let*

$$\Pi(\theta) = \begin{cases} \Phi\pi(\theta) & \textit{if } \theta \in [\theta_*, 1] \\ q(0)\theta & \textit{if } \theta \in [0, \theta_*) \end{cases}.$$

Then, for any $\theta \in [0,1]$

$$\lim_{N\to 0} \Pi_N(\theta) = \Pi(\theta).$$

Proof. For all $\theta \geq \theta_*$, $\Pi_N(\theta) = \Pi(\theta)$, and, for any $\theta < \theta_*$, $\lim_{N\to\infty} \Pi_N(\theta) = \theta q(0) = \Pi(\theta)$ by L'Hopital's rule. $\square$

A.2 Verification of Optimality

We start providing the necessary definitions from the theory of viscosity solutions, together with the relevant results from the theory of optimal control in stratified domains in Barles et al. (2018). We make some changes in notation to make it consistent with our setting, and to translate their minimization problem into a maximization one. While Barles et al. (2018) considers the state space to be the complete real line, the state space in our case is $[0,1]$. However, we can extend the state space by letting the payoff on the complement of $[0,1]$ be sufficiently low. This can be achieved by adding a penalization term and setting the flow payoff equal to $\phi\pi(1) - \dot{\theta}q(1) - k|\theta - 1|$ for $\theta > 1$, and $\phi\pi(0) - \dot{\theta}q(0) - k|\theta|$ for $\theta < 0$. By choosing k large enough, we can ensure that the optimal solution never exits the interval $[0,1]$. Due to the discontinuity in the Hamiltonian at θ_*, a viscosity solution might fail to be unique. In order to fully characterize the value function we need to specify its behavior at θ_*. This is done in Barles et al. (2018) by considering the concept of *Flux-limited* sub- and supersolutions. Letting $\Omega_0 = (-\infty, \theta_*)$ and $\Omega_1 = (\theta_*, \infty)$, we consider the equation

$$\begin{cases} (\rho+\phi)\Pi - H_0(\theta,\Pi) = 0 & \text{in } \Omega_0 \\ (\rho+\phi)\Pi - H_1(\theta,\Pi) = 0 & \text{in } \Omega_1 \\ (\rho+\phi)\Pi - \phi\pi(\theta_*) = 0 & \text{in } \{\theta_*\}, \end{cases} \tag{38}$$

where

$$H_0(\theta,\Pi') = \phi\pi(\theta) + k\min\{0,\theta\} + \max_{|\dot{\theta}|\leq N} \Big\{\dot{\theta}\big(\Pi' - q(0)\big)\Big\}$$
$$H_1(\theta,\Pi') = \phi\pi(\theta) - k\max\{0,\theta-1\} + \max_{|\dot{\theta}|\leq N} \Big\{\dot{\theta}\big(\Pi' - q(\theta)\big)\Big\}$$

In $\Omega_0 \cup \Omega_1$, the definitions are just classical viscosity sub- and supersolution, which we provide next for completeness.

Definition 3 (Bardi and Capuzzo-Dolcetta (2008), Definition 1.1)**.** *A function $u \in C(\mathbb{R})$ is a viscosity subsolution of (33) if, for any $\varphi \in C^1(\mathbb{R})$,*

$$(\rho+\phi)u(\theta_0) - H(\theta_0, \varphi'(\theta_0)) \leq 0, \tag{39}$$

at any local maximum point $\theta_0 \in \mathbb{R}$ of $u - \varphi$. Similarly, $u \in C(\mathbb{R})$ is a viscosity supersolution of (33) if, for any $\varphi \in C^1(\mathbb{R})$,

$$(\rho+\phi)u(\theta_1) - H(\theta_1, \varphi'(\theta_1)) \geq 0, \tag{40}$$

at any local minimum point $\theta_1 \in \mathbb{R}$ of $u - \varphi$. Finally, u is a viscosity solution of (33) if it is simultaneously a viscosity sub- and supersolution.

Before providing the definition of sub- and supersolution on $\{\theta_*\}$, we introduce the following space $\Im$ of real valued test functions: $\varphi \in \Im$ if $\varphi \in C(\mathbb{R})$ and there exist $\varphi_0 \in C^1(\overline{\Omega}_0)$ and $\varphi_1 \in C^1(\overline{\Omega}_1)$ such that $\varphi = \varphi_0$ in $\overline{\Omega}_0$, and $\varphi = \varphi_1$ in $\overline{\Omega}_1$. Next, we introduce two Hamiltonians that are needed to define a flux-limited sub- and supersolution at $\{\theta_*\}$.

$$
\begin{aligned}
H_1^+(\theta_*, \Pi') &\equiv \phi\pi(\theta) + \sup_{0<\dot{\theta}\leq N} \Big\{ \dot{\theta}\big(\Pi' - q(\theta_*)\big) \Big\} \\
H_0^-(\theta_*, \Pi') &\equiv \phi\pi(\theta) + \sup_{0>\dot{\theta}\geq -N} \Big\{ \dot{\theta}\big(\Pi' - q(0)\big) \Big\}.
\end{aligned}
$$

Definition 4 (Barles et al. (2018), Definition 2.1)**.** *An upper semi-continuous, bounded function $u : \mathbb{R} \to \mathbb{R}$ is a flux-limited subsolution on $\{\theta_*\}$ if for any test function $\varphi \in \Im$ such that $u - \varphi$ has a local maximum at θ_*, we have*

$$
(\rho + \phi)u(\theta_*) - \max\{\phi\pi(\theta_*), H_0^-(\theta_*, \varphi_0'(\theta_*)), H_1^+(\theta_*, \varphi_1'(\theta_*))\} \leq 0. \tag{41}
$$

A lower semi-continuous, bounded function $v : \mathbb{R} \to \mathbb{R}$ is a flux-limited supersolution on $\{\theta_*\}$ if for any test function $\varphi \in \Im$ such that $u - \varphi$ has a local minimum at θ_*, we have

$$
(\rho + \phi)v(\theta_*) - \max\{\phi\pi(\theta_*), H_0^-(\theta_*, \varphi_0'(\theta_*)), H_1^+(\theta_*, \varphi_1'(\theta_*))\} \geq 0. \tag{42}
$$

The Hamiltonians H_0^- and H_1^+ are needed to express the optimality conditions at the discontinuity θ_*. H_1^+ consider controls that starting at θ_* take θ_t towards the interior of $[\theta_*, 1]$, and H_0^- considers controls that starting at θ_* take θ_t towards the interior of $[0, \theta_*]$. The use of the Hamiltonians H_0^- and H_1^+ at $\{\theta_*\}$, instead of H_0 and H_1, distinguishes *flux-limited* viscosity solutions from the traditional (discontinuous) viscosity solutions.

We consider the following control problem, equivalent to the one defined in (31),

$$
\begin{aligned}
\tilde{\Pi}_N^*(\theta_0) &= \sup_{|\theta_t|\leq N} \int_0^\infty e^{-(\rho+\phi)t} \Big(\phi\tilde{\pi}(\theta_t) - \theta_t \tilde{q}(\theta_t)\mathbf{1}_{\{\theta_t \neq \theta_*\}} - k\big(\max\{0, \theta_t - 1\} - \min\{0, \theta_t\}\big)\Big) dt \\
\tilde{\pi}(\theta) &= \pi(0)\mathbf{1}_{\{\theta<0\}} + \pi(\theta)\mathbf{1}_{\{\theta\in[0,1]\}} + \pi(1)\mathbf{1}_{\{\theta>1\}} \\
\tilde{q}(\theta) &= q(0)\mathbf{1}_{\{\theta<0\}} + q(\theta)\mathbf{1}_{\{\theta\in[0,1]\}} + q(1)\mathbf{1}_{\{\theta>1\}}.
\end{aligned}
$$

The following Theorem characterizes the value function $\tilde{\Pi}_N^*$ in terms of flux-limited viscosity solutions.

Theorem 1 (Barles et al. (2018), Theorem 2.9)**.** *The value function $\tilde{\Pi}_N^*$ is the unique flux-limited viscosity solution to equation (38).*

We can now proceed to apply Theorem 1 to verify that Π_N defined in (35) is the value function of the control problem (31).

Verification Lemmas 7 and 8 imply that Π_N is a classical solution on $\Omega \setminus \{\tilde{\theta}_N, \theta_*\}$ so we only need to verify the conditions for a viscosity solution on $\{\tilde{\theta}_N, \theta_*\}$. Π_N defined in (35) is a classical solution on $(\theta_*, 1)$. At $\theta = \theta_*$, Π_N has a convex kink, so we only need to verify the supersolution property. That is, that for any $\varphi'(\theta_*)$ in the subdifferential of $\Pi_N(\theta)$ at θ_*, which is $[\Pi'_N(\theta_*-), \Pi'_N(\theta_*+)]$, inequality (42) is satisfied. $H_1^+\big(\theta_*, \varphi'(\theta_*)\big)$ is nondecreasing in $\varphi'(\theta_*)$ and $H_0^-\big(\theta_*, \varphi'(\theta_*)\big)$ is nonincreasing in $\varphi'(\theta_*)$; thus, the supersolution property follows from

$$(\rho+\phi)\Pi_N(\theta_*) - H_1^+\big(\theta_*, \Pi'_N(\theta_*+)\big) = (\rho+\phi)\Pi_N(\theta_*) - \phi\pi(\theta_*) = 0$$
$$(\rho+\phi)\Pi_N(\theta_*) - H_0^-\big(\theta_*, \Pi'_N(\theta_*-)\big) = (\rho+\phi)\Pi_N(\theta_*) - \phi\pi(\theta_*) = 0.$$

As $\Pi'_N(\tilde{\theta}_N-) < q(0) < \Pi'_N(\tilde{\theta}_N+)$, $\Pi_N(\theta)$ has a convex kink at $\tilde{\theta}_N$, we only need to verify the property for a supersolution. Thus, we need to verify that for any $\varphi'(\tilde{\theta}_N) \in [\Pi'_N(\tilde{\theta}_N-), \Pi'_N(\tilde{\theta}_N+)]$, inequality (40) is satisfied. This amounts to verifying that

$$(\rho+\phi)\Pi_N(\tilde{\theta}_N) - \phi\pi(\tilde{\theta}_N) \geq N \max\left\{\left|\Pi'_N(\tilde{\theta}_N-) - q(0)\right|, \left|\Pi'_N(\tilde{\theta}_N+) - q(0)\right|\right\}$$

By definition of $\tilde{\theta}_N$, we have

$$(\rho+\phi)\Pi_N(\tilde{\theta}_N) - \phi\pi(\tilde{\theta}_N) = N\big(\Pi'_N(\tilde{\theta}_N+) - q(0)\big) = N\big(q(0) - \Pi'_N(\tilde{\theta}_N-)\big),$$

so it follows that

$$(\rho+\phi)\Pi_N(\tilde{\theta}_N) - \phi\pi(\tilde{\theta}_N) = N\big|\Pi'_N(\tilde{\theta}_N-) - q(0)\big| = N\big|\Pi'_N(\tilde{\theta}_N+) - q(0)\big|.$$

Finally, at $\theta = 1$, by choosing k large enough, we have that the solution of the HJB equation on $\{\theta > 1\}$ entails $\dot{\theta}(\theta) = -N$. Moreover, $\Pi'(1-) = q(1)$ implies that the value function is differentiable at $\theta = 1$ (in the extended problem) and that $\dot{\theta}(1) \leq 0$ is optimal. Thus, the state constraint is satisfied at $\theta = 1$. A similar argument applies at $\theta = 0$. Thus, we can conclude that $\Pi_N(\theta)$ is a flux-limited viscosity solution, so, by Theorem 1, it is the value function of the optimal control problem.

The uniqueness proof follows from the fact that the HJB has a unique solution in the smooth trading region, under the given boundary conditions. In the region $[0, \theta_*)$, smooth trading and no trading are both ruled out. This implies that trading must be atomistic. Under atomistic trading, the price must be at least $p_L R_o$, but the bank will sell everything, even below this price. The solution θ_* is easily verified as being unique.

B Intermediation

Proof of Lemma 1

Proof. Using integration by parts and a Transversality condition $\lim_{t\to\infty} e^{-(\rho+\phi)t}D_t = 0$, we have

$$\int_0^\infty e^{-(\rho+\phi)s} dD_s = \int_0^\infty e^{-(\rho+\phi)s}(\rho+\phi)D_s ds - D_0.$$

(20) can be rewritten as

$$\max_{(D_t,\theta_t)_{t\geq 0}} \mathbb{E}\left[\int_0^\infty e^{-(\rho+\phi)s}\left[(\phi\hat{\pi}(\theta_s, D_s) - y\left(\theta_s, D_s\right) D_s + (\rho+\phi)D_s\right) ds + dG\left(\theta_s\right)\right]\right] - D_0$$

The optimization in the lemma follows directly from the definition of $\phi\pi(\theta)$ □

Proof of Lemma 2 and 3

Proof. Let $\mathcal{L} = \mathcal{V}\left(D,\theta\right) + \Phi z\left(\theta\right)\left[\Pi_i(\theta) - D\right]$, where $\Phi z\left(\theta\right)$ is the Lagrange multiplier of the debt issuance constraint. From the first order condition, $\mathcal{L}_D = 0$, we get that the optimal solution $D^*(\theta)$ solves

$$\phi\left[f(\kappa_i)\Delta^2\right] D^*(\theta) = \rho - \phi z(\theta) \Rightarrow z(\theta) = \left(\frac{\rho}{\phi} - \left[f(\kappa_i)\Delta^2\right]\Pi_i(\theta)\right)^+. \tag{43}$$

To finish the derivation of optimal debt choice $D^*\left(\theta\right)$, we need one more assumption to guarantee that the second-order condition of the constrained maximization problem is satisfied. Assumption 2 imposes lower and upper bounds on $f'\left(\kappa\right)$, and will be satisfied in the uniform-distribution case. The upper bound guarantees that, for any given θ, the second-order partial derivative of (23) will satisfy $\mathcal{L}_{DD} < 0$. The lower bound guarantees that, in the region where the equity holders' limited liability binds, the value function $\Pi_i\left(\theta\right)$ will be convex in θ.

Lemma 2 follows from the fact that $z'\left(\theta\right) < 0$, which is necessarily the case given Assumption 2 and equation (43). Moreover, the constraint clearly binds at $\theta = 0$ with $D^*\left(0\right) = \Pi_i\left(0\right) = 0$.

There are two cases to be considered: when the borrowing constraint is binding and when it is not. When the borrowing constraint is slack, then $z\left(\theta\right) = 0$, so the conclusion follows directly. Thus, we only need to verify the case in which the borrowing constraint, $D \leq \Pi_i\left(\theta\right)$, is binding.

Note that from the solution to $D^*\left(\theta\right)$, it is immediately clear that $z\left(\theta\right) \leq \frac{\rho}{\phi}$. In the constrained region,

$\Pi_i(\theta) = \mathcal{V}(\Pi_i(\theta), \theta)$. Using the Envelope Theorem to solve for $\Pi_i'(\theta)$, we get

$$\begin{aligned}\Pi_i'(\theta) &= \frac{d\mathcal{V}(\Pi_i(\theta), \theta)}{d\theta} \\ &= \left.\frac{\partial \mathcal{L}}{\partial \theta}\right|_{D=\Pi_i(\theta)} \\ &= \frac{\Phi}{1-\Phi z(\theta)}\left[p(\theta) R_o + f(\kappa_i)\Delta^2 R_o \Pi_i(\theta)\right].\end{aligned}$$

Since the *F.O.C.* implies

$$f(\kappa_i)\Delta^2\Pi_i(\theta) = \frac{\rho - \phi z(\theta)}{\phi},$$

we can show

$$\Pi_i'(\theta) = \frac{\Phi}{1-\Phi z(\theta)}\left[p(\theta) + \frac{\rho - \phi z(\theta)}{\phi}\right] R_o.$$

Substituting this expression in equation (27), we get

$$\begin{aligned}\dot{\theta} &= \phi\frac{(1-\Phi)(1-p(\theta))R_o + \Phi z(\theta)(\Pi_i'(\theta) - R_o)}{\Phi \pi_i''(\theta)} \\ &= \phi\frac{(1-\Phi)(1-p(\theta)) + \Phi z(\theta)\left(\phi\frac{p(\theta)-1}{(\rho+\phi)-\phi z(\theta)}\right)}{\Phi\pi_i''(\theta)}R_o \\ &= \phi\frac{(1-\Phi) - \frac{\Phi z(\theta)\phi}{(\rho+\phi)-\phi z(\theta)}}{\Phi\pi_i''(\theta)}(1-p(\theta))R_o.\end{aligned}$$

Clearly, $\dot{\theta} > 0$ if and only if

$$(1-\Phi) - \frac{\Phi z(\theta)\phi}{(\rho+\phi)-\phi z(\theta)} > 0 \Longrightarrow \frac{\rho}{\phi} > \frac{\phi}{(\rho+\phi)-\phi z(\theta)} z(\theta),$$

which always holds because of the upper bound on the Lagrange multiplier, $z(\theta) \leq \frac{\rho}{\phi}$.

□

Proof of Proposition 3

Proof. The convexity of $\Pi(\theta)$ in the unconstrained region $D^*(\theta) < \Pi_i(\theta)$ follows under Assumption 2. In the constrained region where $D^*(\theta) = \Pi_i(\theta)$, we differentiate $\Pi_i'(\theta)$ and get

$$\Pi_i''(\theta) = \frac{\Phi}{1-\Phi z(\theta)}\left[d_i'(\theta) + z'(\theta)(\Pi_i'(\theta) - R_o)\right].$$

Substituting the first order condition for D in $\Pi_i'(\theta)$, we get that

$$\Pi_i'(\theta) - R_o = -\frac{\Phi}{1-\Phi z(\theta)}(1-p(\theta)) < 0.$$

Hence, it suffices to show that $z'(\theta) \leq 0$. Note that (43) implies

$$z'(\theta) = -\left[f' \Delta^2 \Pi_i(\theta) \frac{\partial \kappa_i}{\partial \theta} + f \Delta^2 \Pi_i'(\theta) \right].$$

Note that $\Pi_i'(\theta) = \Phi \pi_i'(\theta)$, so that

$$\Pi_i'(\theta) = \frac{\Phi}{1 - \Phi z(\theta)} \left[p(\theta) R_o + f \Delta^2 R_o \Pi_i(\theta) \right].$$

Therefore, for $z'(\theta) < 0$, it suffices to have $\Delta f' \Pi_i(\theta) + \frac{\Phi}{1-\Phi z(\theta)} \left[p(\theta) + f \Delta^2 \Pi_i(\theta) \right] > 0$, which always holds under the lower bound imposes by Assumption 2.

Next, at the boundary θ_D where the equity holder's limited liability constraint binds, $\Pi_i'(\theta)$ is continuous, which follows from (26), and the continuity of $d_i(\theta)$, $D^*(\theta)$, as well as the fact that $\lim_{\theta \uparrow \theta_D} z(\theta) = z(\theta_D) = 0$. Therefore, we conclude that $\Pi_i(\theta)$, the value function in the smooth-trading case is globally convex.

Finally, note that at $\theta = 0$, $\Pi_i'(\theta) < p_L R_o$. At $\theta = 1$, $\Pi_i(1) > \Pi_c(1)$, where $\Pi_c(1) > p_L R_o$ follows from Assumption 1. The intersection between $\Pi_i(\theta)$ and $p_L R_o \theta$ then admits the unique solution $\theta_\dagger$. Moreover, $\Pi_i(\theta) < p_L R_o \theta$ if and only if $\theta < \theta_\dagger$.

□

C Analysis of Uniform Distribution

In this appendix, we provide the detailed calculations for the case of a uniform distribution with $p_L = 0$.

Certification: Let

$$v(\theta) = F(\kappa_c) p_H (R - R_o) + (1 - F(\kappa_c)) B \tag{44}$$

be the borrower's expected payoff if the asset matures. The borrower's expected payoff at $t = 0$ is

$$V_c(\theta_0) = \int_0^\infty e^{-(\rho+\phi)t} \phi v(\theta_t)\, dt = \Phi B + \frac{\phi(b - B)}{\bar{\kappa}} \frac{\Delta R_o \theta_0}{2\rho + \phi}, \tag{45}$$

where we have substituted $R_o = R - \frac{b}{\Delta}$ and therefore $p_H(R - R_o) = b$. So, the borrower's problem at time 0 is

$$W_c = \max_{\theta_0} \quad V_c(\theta_0) + L_c(\theta_0) \tag{46}$$

$$s.t. \quad L_c(\theta_0) \geq I - A. \tag{47}$$

It can be easily verified that, $L_c(\theta_0)$ is maximized at $\theta_0 = 1$. However, $V_c(\theta_0) + L_c(\theta_0)$ is maximized at

$$\theta_0 = 1 - \frac{\rho + \phi}{2\rho + \phi} \frac{B - b}{\Delta R_o}. \tag{48}$$

Intermedation: If the borrowing constraint $D \leq \Pi$ is binding, $\Pi_i(\theta)$ solves

$$\frac{\Delta^2}{2\bar{\kappa}} \left[(R_o \theta)^2 - D^2 \right] = D.$$

The previous equation has two roots, and the positive root is

$$\Pi_i(\theta) = -\frac{\bar{\kappa}}{\Delta^2} + \sqrt{\left(\frac{\bar{\kappa}}{\Delta^2}\right)^2 + (R_o \theta)^2}.$$

1. If $\Phi \geq \underline{\Phi} := \sqrt{\frac{(\bar{\kappa}/\Delta^2)^2}{(\bar{\kappa}/\Delta^2)^2 + (R_o)^2}}$, the bank's debt choice satisfies $D^*(1) = \tilde{D}(1) = \frac{\rho\bar{\kappa}}{\phi\Delta^2}$.
2. Otherwise, the bank's debt choice satisfies $D^*(1) = \Pi_i(1) = -\frac{\bar{\kappa}}{\Delta^2} + \sqrt{\left(\frac{\bar{\kappa}}{\Delta^2}\right)^2 + (R_o)^2}$.

The borrowing capacity is $L_i(1) = \Pi_i(1) = \Phi \pi_i(1)$ and the optimal debt when $\theta = 1$ is

$$D^*(1) = \min\left\{ \frac{\rho\bar{\kappa}}{\phi\Delta^2}, -\frac{\bar{\kappa}}{\Delta^2} + \sqrt{\left(\frac{\bar{\kappa}}{\Delta^2}\right)^2 + (R_o)^2} \right\}.$$

The debt-issuance constraint is slack if

$$\tilde{D}(1) = \frac{\rho\bar{\kappa}^2}{\phi\Delta^2} \leq \Pi_i(1) = -\frac{\bar{\kappa}}{\Delta^2} + \sqrt{\left(\frac{\bar{\kappa}}{\Delta^2}\right)^2 + R_o^2}.$$

Simple derivation shows this is satisfied if and only if

$$\Phi > \sqrt{\frac{(\bar{\kappa}/\Delta^2)^2}{R_o^2 + (\bar{\kappa}/\Delta^2)^2}},$$

which holds if and only if ρ is sufficiently low.

Whenever the borrowing constraint is slack (that is $\tilde{D}(1) < \Pi_i(1)$), we can plug in $D^*(1)$ to get $\pi_i(1) = \frac{\Delta^2}{2\bar{\kappa}} \left(R_o - \frac{\rho\bar{\kappa}}{\phi\Delta^2} \right)^2 + \frac{\rho + \phi \frac{\Delta^2}{\bar{\kappa}} \left(R_o - \frac{\rho\bar{\kappa}}{\phi\Delta^2} \right)}{\phi} \frac{\rho\bar{\kappa}}{\phi\Delta^2}$.

According to (4), $\kappa_i = \Delta \left(R_o - \frac{\rho\bar{\kappa}}{\phi\Delta^2} \right)$ and $p_i(1) = \frac{\kappa_i}{\bar{\kappa}} \Delta = \frac{\Delta^2}{\bar{\kappa}} \left(R_o - \frac{\rho\bar{\kappa}}{\phi\Delta^2} \right)$. Consequently, $\hat{\pi}_i(1, D^*)$ and $\pi_i(1)$ defined in (17) and (21) become

$$\hat{\pi}_i(1, D^*) = \frac{\Delta^2}{2\bar{\kappa}} \left(R_o - \frac{\rho\bar{\kappa}}{\phi\Delta^2} \right)^2 \tag{49}$$

$$\pi_i(1) = \frac{\Delta^2}{2\bar{\kappa}} \left(R_o - \frac{\rho\bar{\kappa}}{\phi\Delta^2} \right)^2 + \frac{\rho + \phi \frac{\Delta^2}{\bar{\kappa}} \left(R_o - \frac{\rho\bar{\kappa}}{\phi\Delta^2} \right)}{\phi} \frac{\rho\bar{\kappa}}{\phi\Delta^2}. \tag{50}$$

As in the certification case, we can define the entrepreneur and bank's payoff as

$$v(1) = F(\kappa_i)\, p_H (R - R_o) + (1 - F(\kappa_i)) B = B + \frac{\Delta\left(R_o - \frac{\rho\bar{\kappa}}{\phi\Delta^2}\right)}{\bar{\kappa}} (b - B) \tag{51}$$

$$V_i(1) = \int_0^\infty e^{-(\rho+\phi)t} \phi v(\theta_t)\, dt = \Phi v(1) \tag{52}$$

$$L_i(1) = \Pi_i(1) = \Phi \pi_i(1). \tag{53}$$

Thus, the borrower's payoff at time 0 is

$$W_i = V_i(1) + L_i(1). \tag{54}$$

Proof of Proposition 4

Proof. Note that $L_c(\theta_0) = \Pi_c(\theta_0) + (1 - \theta_0)\, q_c(\theta_0)$. In the proposition, $L_c(\theta_0^*) > I - A$ guarantees that the borrowing constraints are slack in both certification and intermediation. Under certification

$$W_c = \Phi B + \frac{\phi(b - B)}{\bar{\kappa}} \frac{\Delta R_o}{2\rho + \phi} \theta_0 + \frac{\Phi}{\bar{\kappa}} (\Delta R_o)^2 \theta_0 \left(1 - \frac{\theta_0}{2}\right),$$

where θ_0 is evaluated at

$$\theta_0^* = 1 - \frac{\rho + \phi}{2\rho + \phi} \frac{B - b}{\Delta R_o} = 1 - \frac{1}{2 - \Phi} \frac{B - b}{\Delta R_o},$$

Under intermediation,

$$W_i = \Phi \left[B + \frac{\Delta\left(R_o - \frac{\rho\bar{\kappa}}{\phi\Delta^2}\right)}{\bar{\kappa}} (b - B)\right] + \frac{\Phi}{2\bar{\kappa}} \Delta^2 \left(R_o - \frac{\rho\bar{\kappa}}{\phi\Delta^2}\right)^2 + \frac{\rho + \phi \frac{\Delta^2}{\bar{\kappa}} \left(R_o - \frac{\rho\bar{\kappa}}{\phi\Delta^2}\right)}{\rho + \phi} \frac{\rho\bar{\kappa}}{\phi\Delta^2}.$$

Certification dominates intermediation if $W_c > W_i$, letting $\Delta W \equiv W_c - W_i$,

$$\begin{aligned}
\Delta W =& \Phi \left[\frac{1}{2 - \Phi} \frac{(b - B)}{\bar{\kappa}} \left(\Delta R_o - \frac{1}{2 - \Phi}(B - b)\right) + \frac{1}{2\bar{\kappa}} \left(\Delta R_o - \frac{1}{2 - \Phi}(B - b)\right) \left(\Delta R_o + \frac{1}{2 - \Phi}(B - b)\right)\right. \\
& \left. - \frac{(b - B)}{\bar{\kappa}} \left(\Delta R_o - \frac{\rho\bar{\kappa}}{\phi\Delta}\right) - \frac{1}{2\bar{\kappa}} \left(\Delta R_o - \frac{\rho\bar{\kappa}}{\phi\Delta}\right)^2 - \left(\frac{1 - \Phi}{\Phi} + \frac{\Delta}{\bar{\kappa}} \left(\Delta R_o - \frac{\rho\bar{\kappa}}{\phi\Delta}\right)\right) \frac{\rho\bar{\kappa}}{\phi\Delta^2}\right] \\
=& \frac{\Phi}{2\bar{\kappa}} \left[\left(\frac{B - b}{2 - \Phi}\right)^2 + 2\left(\frac{1 - \Phi}{2 - \Phi} \Delta R_o - \frac{\rho\bar{\kappa}}{\phi\Delta}\right)(B - b) - \left(\frac{\rho\bar{\kappa}}{\phi\Delta}\right)^2\right]
\end{aligned}$$

Comparative statics *w.r.t.* B**.** The previous equation is negative for B on $(\underline{B}, \bar{B})$, where

$$\underline{B} = b - (2-\Phi)^2 \left[\sqrt{\left(\frac{1-\Phi}{2-\Phi}\Delta R_o + \frac{\rho\kappa}{\phi\Delta}\right)^2 + \left(\frac{\rho\bar{\kappa}}{\phi\Delta(2-\Phi)}\right)^2} - \left(\frac{1-\Phi}{2-\Phi}\Delta R_o - \frac{\rho\kappa}{\phi\Delta}\right)\right]$$

$$\bar{B} = b + (2-\Phi)^2 \left[\sqrt{\left(\frac{1-\Phi}{2-\Phi}\Delta R_o - \frac{\rho\kappa}{\phi\Delta}\right)^2 + \left(\frac{\rho\bar{\kappa}}{\phi\Delta(2-\Phi)}\right)^2} - \left(\frac{1-\Phi}{2-\Phi}\Delta R_o - \frac{\rho\kappa}{\phi\Delta}\right)\right]$$

Because $\underline{B} < b$, we only need to consider the upper bound, so, after substituting R_o, we get that the expression is negative as long as B satisfies

$$b < B < b + (2-\Phi)^2 \left[\sqrt{\left(\frac{1-\Phi}{2-\Phi}(\Delta R - b) - \frac{\rho\kappa}{\phi\Delta}\right)^2 + \left(\frac{\rho\bar{\kappa}}{\phi\Delta(2-\Phi)}\right)^2} - \left(\frac{1-\Phi}{2-\Phi}(\Delta R - b) - \frac{\rho\kappa}{\phi\Delta}\right)\right].$$

Thus, we can conclude that certification dominates if

$$B > B^* \equiv b + (2-\Phi)^2 \left[\sqrt{\left(\frac{1-\Phi}{2-\Phi}(\Delta R - b) - \frac{\rho\kappa}{\phi\Delta}\right)^2 + \left(\frac{\rho\bar{\kappa}}{\phi\Delta(2-\Phi)}\right)^2} - \left(\frac{1-\Phi}{2-\Phi}(\Delta R - b) - \frac{\rho\kappa}{\phi\Delta}\right)\right].$$

Comparative statics *w.r.t.* ϕ**.** We have

$$\frac{\partial \Delta W}{\partial \phi} = \frac{\rho}{2\bar{\kappa}(\phi+\rho)^2}\left[\frac{2+\Phi}{2-\Phi}\left(\frac{B-b}{2-\Phi}\right)^2 + \frac{2\phi+\rho}{\rho}\left(\frac{\rho\bar{\kappa}}{\phi\Delta}\right)^2 + \frac{2(B-b)}{\Delta}\left[\Delta^2 R_o\left(1 - \frac{2}{(2-\Phi)^2}\right) + \bar{\kappa}\right]\right].$$

Note that when the debt issuance constraint is slack, we have $D^* = \frac{\rho\bar{\kappa}}{\phi\Delta^2}$, which implies

$$p_i(\theta_0) = \hat{p}_i(\theta_0, D^*) = F(\kappa_i)\Delta = \frac{\Delta^2 (R_o\theta_0 - D^*)}{\bar{\kappa}} = \frac{\Delta^2}{\bar{\kappa}}\left(R_o\theta_0 - \frac{\rho\bar{\kappa}}{\phi\Delta^2}\right).$$

The constraint $p_i(\theta_0) \le 1$ for all $\theta_0 \in [0,1]$ implies that a constraint $\bar{\kappa} \ge \Phi\Delta^2 R_o$ is required. Under this constraint, we get

$$\Delta^2 R_o\left[1 - \frac{2}{(2-\Phi)^2}\right] + \bar{\kappa} \ge \Delta^2 R_o\left[1 - \frac{2}{(2-\Phi)^2}\right] + \Phi\Delta^2 R_o = \frac{\Delta^2 R_o}{(2-\Phi)^2}(\Phi-1)(\Phi-1-\sqrt{3})(\Phi-1+\sqrt{3}) \ge 0,$$

implying that $\frac{\partial \Delta W}{\partial \phi} \ge 0$. Note that ΔW is continuous in ϕ. Moreover, if $\phi \uparrow \infty$, we have

$$\Delta W = \frac{(B-b)^2}{2\bar{\kappa}} > 0,$$

and if $\phi \downarrow 0$, we have

$$\Delta W < 0.$$

Letting $\Delta W = 0$, there exists a unique positive root ϕ^* of ϕ. If $\phi > \phi^*$, we have $W_c > W_i$; if $\phi < \phi^*$, we have $W_c < W_i$.

Comparative statics *w.r.t.* $\bar{\kappa}$**.** Letting $\Delta W = 0$, we get a quadratic equation of $\bar{\kappa}$

$$\left(\frac{B-b}{2-\Phi}\right)^2 + 2(B-b)\left(\frac{1-\Phi}{2-\Phi}\Delta R_o - \frac{\rho\bar{\kappa}}{\phi\Delta}\right) - \left(\frac{\rho\bar{\kappa}}{\phi\Delta}\right)^2 = 0.$$

The positive root is

$$\bar{\kappa}^* \equiv \frac{\phi\Delta}{\rho}\left[\sqrt{\frac{(2-\Phi)^2+1}{(2-\Phi)^2}(B-b)^2 + \frac{2\Delta R_o(1-\Phi)(B-b)}{2-\Phi}} - (B-b)\right].$$

Since

$$\frac{\partial \Delta W}{\partial \bar{\kappa}} = -\frac{\Phi}{2\bar{\kappa}^2}\left[\left(\frac{B-b}{2-\Phi}\right)^2 + 2(B-b)\frac{1-\Phi}{2-\Phi}\Delta R_o + \left(\frac{\rho\bar{\kappa}}{\phi\Delta}\right)^2\right] \leq 0,$$

$\bar{\kappa}^*$ is the unique positive root. Then $W_c < W_i$ if $\bar{\kappa} > \bar{\kappa}^*$, and $W_c > W_i$ if $0 < \bar{\kappa} < \bar{\kappa}^*$.

Comparative statics *w.r.t.* Δ Since

$$\frac{\partial \Delta W}{\partial \Delta} = \frac{\Phi}{2\bar{\kappa}}\left[2(B-b)\left(\frac{1-\Phi}{2-\Phi}R + \frac{\rho\bar{\kappa}}{\phi\Delta^2}\right) + \frac{2}{\Delta}\left(\frac{\rho\bar{\kappa}}{\phi\Delta}\right)^2\right] \geq 0,$$

ΔW is increasing in Δ. Letting $\Delta W = 0$, we get

$$\left(\frac{B-b}{2-\Phi}\right)^2 + 2(B-b)\left(\frac{1-\Phi}{2-\Phi}(\Delta R - b) - \frac{\rho\bar{\kappa}}{\phi\Delta}\right) - \left(\frac{\rho\bar{\kappa}}{\phi\Delta}\right)^2 = 0,$$

which determines a unique positive root Δ^*. Thus $W_c < W_i$ if $\Delta < \Delta^*$, and $W_c > W_i$ if $\Delta > \Delta^*$.

□

Proof of Proposition 5

Proof. We can similarly define $\mathcal{V}$, the bank's objective function without trading gains as

$$\mathcal{V}(D,\theta) := \Phi\left[\hat{p}_i(\theta,D)\,\theta R_o - \int_0^{\kappa_i}\kappa dF(\kappa)\right] + (1-\Phi)\,D + \Phi\,(1-\hat{p}_i(\theta,D))\,(1-\xi)\,D,$$

where the new term $\Phi\,(1-\hat{p}_i(\theta,D))\,(1-\xi)\,D$ stands for the benefit of the government subsidy.

If θ is sufficiently large such that the debt issuance constraint is slack, simple derivation shows that the optimal debt issuance satisfies

$$\tilde{D}(\theta) = \frac{(1-\Phi) - \Phi\,(1-p(\theta))\,(1-\xi)}{\Phi\,[1-(1-p(\theta))\,(1-\xi)]\,f(\kappa_i)\,\Delta^2}.$$

Recall that in Lemma 3, the constraint is slack whenever θ is sufficiently high.

In the region where the debt-issuance constraint is slack, the HJB equation implies

$$\dot{\theta} = \phi \frac{R_o \xi \frac{(1-\Phi)-\Phi(1-p(\theta))(1-\xi)}{1-(1-p(\theta))(1-\xi)} - (1-\Phi)\, p(\theta) R_o}{\Phi \pi_i''(\theta)}.$$

Clearly, when $\xi = 1$, we get the results in subsection 4.2 that $\dot{\theta} = \phi \frac{R_o(1-\Phi)(1-p(\theta))}{\Phi \pi_i''(\theta)} > 0$. Moreover, when $\xi = 0$ so that the entire interest rate is subsidized by the government, $\dot{\theta} = \phi \frac{-(1-\Phi)p(\theta)R_o}{\Phi \pi_i''(\theta)} < 0$, implying the bank sells loans over time. In general, there exists a $\xi_\dagger$ and $\dot{\theta} < 0$ if $\xi < \xi_\dagger$. $\square$

D Discrete-time Model

D.1 Model Setup and Main Results

We present the full model behind the example in Section 2. With slight abuse of notation, let Φ be the one-period discount rate of the entrepreneur and the bank. Investors do not discount the future. The model has a total of four dates: $t = 0, 1, 2, 3$. At $t = 0$, the entrepreneur signs a contract, which specifies the bank's retention θ_0 and the investors' retention $1 - \theta_0$.

The project matures at $t = 3$, and its outcome depends on the entrepreneur's effort. Specifically, the project produces R with probability p_H if the entrepreneur works but with probability $p_L = p_H - \Delta$ if the entrepreneur shirks. The two shirking options bring private benefits B and b. Let $R_f = b/\Delta$ be the entrepreneur's retention and $R_o = R - R_f$ be the loans. The financial market opens at $t = 1$ and $t = 2$, in which the bank can trade loans as well as issue debt. Figure 5 describes the timing.

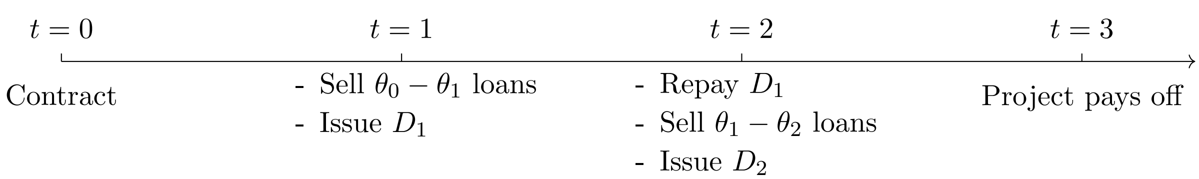


Figure 5: Timing with three periods

Without the Financial Market. Let us first analyze the model without the financial market; that is, neither loan trading nor short-term debt issuance is allowed at $t = 1$ or $t = 2$. As a result, the bank and investors always retain their shares θ_0 and $1 - \theta_0$ until $t = 3$, and the amount of short-term debt outstanding satisfies $D_1 = D_2 = 0$. Given any θ_0, the amount that the entrepreneur is able to raise at $t = 0$ is

$$L(\theta_0) = \underbrace{\Phi^3 \left(q(\theta_0)\theta_0 - \int_0^{\Delta R_o \theta_0} \kappa dF(\kappa) \right)}_{\text{bank's lending}} + \underbrace{q(\theta_0)(1 - \theta_0)}_{\text{investors' lending}},$$

where $q(\theta_0) = (p_L + F(\Delta R_o \theta_0)\Delta) R_o$ is the price of the loan under bank's retention θ_0. The entrepreneur's payoff is

$$\max_{\theta_0 \in [0,1]} \underbrace{L(\theta_0) - (I - A)}_{\text{extra borrowing}} + \underbrace{\Phi^3 (p_L + F(\Delta R_o \theta_0)\Delta) R_f}_{\text{payoff from project maturing}}$$
$$\text{s.t.} \quad L(\theta_0) \geq I - A$$

Let the optimal solution to the problem be θ_0^* and we focus below on the case $L(\theta_0^*) > I - A$.

Financial market opens once. Now we analyze the model in which the financial market opens once at $t = 2$. Therefore, the bank's retention at the end of $t = 1$ (equivalently the beginning of $t = 2$) equals $\theta_1 = \theta_0^*$. In the financial market at $t = 2$, the bank can either sell the loan $\theta_1 - \theta_2$ or issue debt D_2 against it. The bank's payoff at $t = 2$ is

$$\hat{V}_2(\theta_2, D_2, \theta_1) = \underbrace{\Phi \left(p(\theta_2, D_2)(\theta_2 R_o - D_2) - \int_0^{\Delta(\theta_2 R_o - D_2)} \kappa dF(\kappa) \right)}_{\text{bank equity value}} + \underbrace{p(\theta_2, D_2) D_2}_{\text{debt issuance}} + \underbrace{q(\theta_2, D_2)(\theta_1 - \theta_2)}_{\text{trading gains}}.$$

The bank's problem is

$$V_2(\theta_1) = \max_{\theta_2 \in [0,1], D_2 \geq 0} \hat{V}_2(\theta_2, D_2, \theta_1)$$
$$\text{s.t.} \quad D_2 \leq \theta_2 R_o.$$

It is easily derived that $\frac{\partial \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial \theta_2} = -R_o \frac{\partial \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial D_2}$. Under the issuance constraint $D_2 \leq \theta_2 R_o$, the choice between θ_2 and D_2 is undetermined.

Financial market opens twice. Now we turn to the model in which the financial market opens at both $t = 1$ and $t = 2$. Given any θ_0, the bank's payoff from loan trading and issuing one-period debt at $t = 1$ is

$$\hat{V}_1(\theta_1, D_1, \theta_0) = D_1 + q_1(\theta_1, D_1)(\theta_0 - \theta_1) + \Phi (V_2(\theta_1) - D_1).$$

Note that the one-period debt D_1 will be paid with certainty at $t = 2$ and is therefore riskless. Moreover, the price of the loan satisfies $q_1(\theta_1, D_1) = q_2(\theta_2(\theta_1), D_2(\theta_1))$, where $\theta_2(\theta_1)$ and $D_2(\theta_1)$ are the bank's optimal decisions given θ_1. The bank's problem at $t = 1$ is

$$
\begin{aligned}
V_1(\theta_0) = & \max_{\theta_1 \in [0,1], D_1 \geq 0} \hat{V}_1(\theta_1, D_1, \theta_0) \\
& \text{s.t.} \quad D_1 \leq V_2(\theta_1).
\end{aligned}
$$

Once again, the issuance constraint $D_1 \leq V_2(\theta_1)$ arises because otherwise the bank will immediately default. By formulating Lagrangian, we can show that the first-order conditions imply

$$
\begin{aligned}
\theta_1 &= \theta_0 \\
D_1 &= V_2(\theta_1).
\end{aligned}
$$

Financial market without short-term debt issuance. Finally, let us study the model in which the financial market opens on both dates, but debt issuance is not allowed. This model will map into the certification model analyzed later in section 4.1. Now that $D_2 = 0$, for any given θ_1, there is a unique θ_2 that maximizes the bank's payoff. Moreover, we show both $\theta_1 \leq \theta_0$ and $\theta_2 \leq \theta_1$ hold, so that the bank always sells the loan at both dates.

D.2 Detailed Analysis and Proof

Problem at $t = 2$ given any θ_1

From

$$
\begin{aligned}
p_2(\theta_2, D_2) &= p_L + \Delta F\left(\Delta(\theta_2 R_o - D_2)\right) \\
q_2(\theta_2, D_2) &= p_2(\theta_2, D_2) R_o = \left[p_L + \Delta F\left(\Delta(\theta_2 R_o - D_2)\right)\right] R_o,
\end{aligned}
$$

it is easily derived that $\frac{\partial p_2(\theta_2, D_2)}{\partial \theta_2} = -R_o \frac{\partial p_2(\theta_2, D_2)}{\partial D_2}$ holds. Therefore, we get

$$
\begin{aligned}
\frac{\partial \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial \theta_2} &= (\Phi - 1) p_2(\theta_2, D_2) R_o + \frac{\partial p_2(\theta_2, D_2)}{\partial \theta_2} \left[R_o(\theta_1 - \theta_2) + D_2\right] \\
\frac{\partial \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial D_2} &= (1 - \Phi) p_2(\theta_2, D_2) + \frac{\partial p_2(\theta_2, D_2)}{\partial D_2} \left[R_o(\theta_1 - \theta_2) + D_2\right] \\
\Rightarrow \frac{\partial \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial \theta_2} &= -R_o \frac{\partial \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial D_2}.
\end{aligned}
$$

By formulating the Lagrangian of the bank's problem, we get

$$
\begin{aligned}
\frac{\partial \mathcal{L}_2}{\partial \theta_2} &= \frac{\partial \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial \theta_2} + \eta_2 R_o = 0 \\
\frac{\partial \mathcal{L}_2}{\partial \theta_2} &= \frac{\partial \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial \theta_2} - \eta_2 = 0,
\end{aligned}
$$

where η_2 is the Lagrange multiplier on the constraint $D_2 \leq \theta_2 R_o$. It means for any (θ_2, D_2) that solves the first F.O.C, it also solves the second one. Therefore, the choice of (θ_2, D_2) is undetermined. If we restrict $D_2 = 0$, then the F.O.C. implies

$$(1 - \Phi)p_2(\theta_2, D_2)R_o = \frac{\partial p_2(\theta_2, D_2)}{\partial \theta_2} R_o(\theta_1 - \theta_2).$$

Given $\frac{\partial p_2(\theta_2, D_2)}{\partial \theta_2} > 0$, $\theta_2 < \theta_1$. Note that there is no closed-form solution for θ_2 as a function of θ_1.

Problem at $t = 1$ given any θ_0

We know

$$\hat{V}_1(\theta_1, D_1, \theta_0) = D_1 + q_1(\theta_1, D_1)(\theta_0 - \theta_1) + \Phi\left(-D_1 + V_2(\theta_1)\right).$$

The bank's problem at $t = 1$ is

$$\max_{\theta_1, D_1} \hat{V}_1(\theta_1, D_1, \theta_0)$$
$$s.t.\ D_1 \leq V_2(\theta_1).$$

Let us formulate Lagrangian. We get

$$\begin{aligned}
\frac{\partial \mathcal{L}_1}{\partial D_1} &= 1 - \Phi - \eta_1 = 0 \Rightarrow \eta_1 = 1 - \Phi. \\
\frac{\partial \mathcal{L}_1}{\partial \theta_1} &= -q_1(\theta_1, D_1) + \frac{\partial q_1(\theta_1, D_1)}{\partial \theta_1}(\theta_0 - \theta_1) + (\Phi + \eta)\frac{\partial V_2(\theta_1)}{\partial \theta_1} \\
&= -q_1(\theta_1, D_1) + \frac{\partial q_1(\theta_1, D_1)}{\partial \theta_1}(\theta_0 - \theta_1) + (\Phi + \eta)q_2(\theta_2, D_2) \\
&= -q_1(\theta_1, D_1) + \frac{\partial q_1(\theta_1, D_1)}{\partial \theta_1}(\theta_0 - \theta_1) + (\Phi + \eta)q_1(\theta_1, D_1) \\
&= \frac{\partial q_1(\theta_1, D_1)}{\partial \theta_1}(\theta_0 - \theta_1),
\end{aligned}$$

where the second line follows the envelope condition. From $\frac{\partial \mathcal{L}_1}{\partial \theta_1} = 0$, we know $\theta_1 = \theta_0$. Moreover, the result $\eta_1 > 0$ implies the constraint $D_1 \leq V_2(\theta_1)$ always binds. Therefore, given any θ_0, it is always the case that $\theta_1 = \theta_0$ and $D_1 = V_1(\theta_1)$. Note that in contrast with the result at $t = 2$, loan sales and debt issuance are not equivalent. The choice of θ_1 and D_1 is uniquely determined.

We can also restrict both $D_1 = 0$ and $D_2 = 0$, similar to certification. In this case, $\eta_1 = 0$ and

$$\frac{\partial \mathcal{L}}{\partial \theta_1} = -q_1(\theta_1, 0) + \frac{\partial q_1(\theta_1, 0)}{\partial \theta_1}(\theta_0 - \theta_1) + \Phi q_1(\theta_1, 0).$$

From $\frac{\partial \mathcal{L}}{\partial \theta_1} = 0$, we get

$$\frac{\partial q_1(\theta_1, 0)}{\partial \theta_1}(\theta_0 - \theta_1) = (1 - \Phi)q_1(\theta_1, 0) > 0,$$

where
$$\frac{\partial q_1(\theta_1, 0)}{\partial \theta_1} = \frac{\partial q_2}{\partial \theta_2}\frac{\partial \theta_2}{\partial \theta_1} = (R_o\Delta)^2 f\left(\Delta(\theta_2 R_o)\right)\frac{\partial \theta_2}{\partial \theta_1}.$$

It remains to derive the sign of $\frac{\partial \theta_2}{\partial \theta_1}$. Noticing that

$$\frac{\partial \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial \theta_2} = (\Phi - 1)p_2(\theta_2, D_2)R_o + \frac{\partial p_2(\theta_2, D_2)}{\partial \theta_2}\left[R_o(\theta_1 - \theta_2) + D_2\right],$$

so
$$\frac{\partial^2 \hat{V}_2(\theta_2, D_2, \theta_1)}{\partial \theta_2 \partial \theta_1} = \frac{\partial p_2(\theta_2, D_2)}{\partial \theta_2}R_o > 0,$$

which means that $\hat{V}_2(\theta_2, D_2, \theta_1)$ satisfies the single crossing property in (θ_1, θ_2). It follows from the monotone comparative results in Milgrom and Shannon (1994) that $\frac{\partial \theta_2}{\partial \theta_1} \geq 0$.

Again, note that the closed-form solution for θ_1 given any θ_0 is not available.

E Detailed Analysis

E.1 Intermediating Bank with Long-term Debt and Loss-Absorbing Equity

Following the solution, we have the bank's value function under $\theta_t = \theta$ is

$$\Pi_i(\theta) = \Phi\pi_i(\theta) = \Phi\left[\hat{p}_i(\theta, D)(\theta R_o - D) + (1 - \hat{p}_i(\theta, D))(-X) - \int_0^{\kappa_i} \kappa dF(\kappa)\right].$$

At $t = 0$, the bank chooses $\{\theta_0, D\}$ to maximize

$$\Pi_i(\theta_0, D) + q_D(\theta_0, D)D,$$

where
$$q_D(\theta_0, D) = \mathbb{E}\left[p_i(\theta) + (1 - p_i(\theta))\frac{X}{D}\right]$$

is the price of long-term debt at $t = 0$, and the expectation operator takes into account the trajectory of θ_t after $\theta_0 = \theta$ and the shock for project maturing.

E.2 Lender Dispersion

We study the optimal choice of θ_0 in the Linear-Quadratic example and show that it can be either interior or 1, depending on parameters. Specifically, when the short-term debt issuance constraint is slack, we have

$$\begin{aligned}
W_i(\theta_0) =& L_i(\theta_0) + V_i(\theta_0) \\
=& -\Phi\frac{(\Delta R_o)^2}{2\bar{\kappa}}\theta_0^2 + \frac{\Delta R_o}{\bar{\kappa}}\left(\Phi\Delta R_o + \phi\frac{(b-B)}{2\rho+\phi}\right)\theta_0 + \left(\frac{\bar{\kappa}-\Phi R_o\Delta^2\theta_0}{\bar{\kappa}-\Phi R_o\Delta^2}\right)^{-\frac{\rho+\phi}{\rho}}\frac{\rho(b-B)}{2\rho+\phi}\left(\frac{\Phi\Delta R_o}{\bar{\kappa}} - \frac{1}{\Delta}\right) \\
&+ \frac{\rho}{\Delta}\frac{(b-B)}{2\rho+\phi} + \Phi\left(B - \frac{\rho(b-B)}{\phi\Delta} + \frac{\rho}{2\phi}\frac{\rho\bar{\kappa}}{\phi\Delta^2}\right),
\end{aligned}$$

and

$$\begin{aligned}
W_i'(\theta_0) &= \frac{\Phi(\Delta R_o)^2}{\bar{\kappa}}(1-\theta_0) - \frac{\phi\Delta R_o}{\bar{\kappa}}\frac{B-b}{2\rho+\phi}\left[1 - \left(\frac{\bar{\kappa}-\Phi R_o\Delta^2\theta_0}{\bar{\kappa}-\Phi R_o\Delta^2}\right)^{-\frac{2\rho+\phi}{\rho}}\right], \\
W_i''(\theta_0) &= \frac{\phi\Delta R_o}{\bar{\kappa}}\frac{\Phi R_o\Delta^2}{\bar{\kappa}-\Phi R_o\Delta^2}\frac{B-b}{\rho}\left(\frac{\bar{\kappa}-\Phi R_o\Delta^2\theta_0}{\bar{\kappa}-\Phi R_o\Delta^2}\right)^{-\frac{2\rho+\phi}{\rho}-1} - \frac{\Phi(\Delta R_o)^2}{\bar{\kappa}}.
\end{aligned}$$

Clearly,

$$\begin{aligned}
W_i'(1) &= 0, \\
W_i''(1) &= \frac{\Phi(\Delta R_o)^2}{\bar{\kappa}}\left[\frac{\phi\Delta}{\bar{\kappa}-\Phi R_o\Delta^2}\frac{B-b}{\rho} - 1\right].
\end{aligned}$$

The sign of $W_i''(1)$ depends on the choices of parameters. If $B-b$ is small, $W_i''(1) < 0$ maybe negative and as a result $\theta_0 = 1$ can be optimal. If $B-b$ is large, $W_i''(1) > 0$ and the optimal $\theta_0 \in (0,1)$.

E.3 Continuous Monitoring Technology

Suppose the project generates a payoff of R at maturity with a probability of 1 unless it fails before that maturity event arrives. The failure probability depends on the bank's monitoring effort: before maturity, the project fails with an arrival rate $\eta(1-m_t)$, where m_t is the bank's monitoring effort chosen at time t. The cost of monitoring is given by a convex function $h(m_t)$. Let $m_t = m$.

Certification The HJB equation for the intermediary is

$$(\rho+\eta)\Pi_c(\theta) = \max_{\dot{\theta},m}\ \phi\Big[R_o\theta - \Pi_c(\theta)\Big] + \eta m\Pi_c(\theta) - h(m) + \dot{\theta}\Big[\Pi_c'(\theta) - q_c(\theta)\Big]$$

The first order conditions for $\dot{\theta}$ and m are

$$\begin{aligned}
\Pi_c'(\theta) &= q_c(\theta) \\
h'(m) &= \eta\Pi_c(\theta).
\end{aligned}$$

The price of the asset is

$$\eta(1-m(\theta))q_c(\theta)=\phi\Big[R_o-q_c\left(\theta\right)\Big]+\dot{\theta}q_c'\left(\theta\right),$$

where

$$m(\theta)=(h')^{-1}\left[\eta\Pi_c\left(\theta\right)\right]$$

By the envelope theorem, we have

$$\left(\rho+\eta(1-m(\theta))\right)\Pi_c'\left(\theta\right)=\phi\Big[R_o-\Pi_c'\left(\theta\right)\Big]+\dot{\theta}\Big[\Pi_c''\left(\theta\right)-q_c'\left(\theta\right)\Big]$$

Substituting $\Pi_c'\left(\theta\right)=q_c\left(\theta\right)$ and $\Pi_c''\left(\theta\right)=q_c'\left(\theta\right)$ we get

$$\left(\rho+\eta(1-m(\theta))\right)q_c\left(\theta\right)=\phi\Big[R_o-q_c\left(\theta\right)\Big]$$

It follows that

$$\rho q_c(\theta)=-\dot{\theta}q_c'(\theta)\Longrightarrow\dot{\theta}=-\rho\frac{q_c(\theta)}{q_c'(\theta)}=-\rho\frac{\Pi_c'\left(\theta\right)}{\Pi_c''\left(\theta\right)}$$

In the particular case that $h(m)=km^2/2$ we have $(h')^{-1}(x)=x/k$, so

$$m(\theta)=\frac{\eta}{k}\Pi_c\left(\theta\right).$$

Substituting in the HJB equation, and defining $\upsilon\equiv\eta^2/k$ we get

$$\frac{\upsilon}{2}\Pi_c\left(\theta\right)^2-\left(\rho+\phi+\eta\right)\Pi_c\left(\theta\right)+\phi R_o\theta=0.$$

The solution must be increasing in θ. Hence the relevant root is

$$\Pi_c\left(\theta\right)=\frac{\left(\rho+\phi+\eta\right)-\sqrt{\left(\rho+\phi+\eta\right)^2-2\upsilon\phi R_o\theta}}{\upsilon}.$$

The derivatives for the value function are

$$\begin{aligned}
\Pi_c'(\theta)&=\phi R_o\left[\left(\rho+\phi+\eta\right)^2-2\upsilon\phi R_o\theta\right]^{-1/2}>0\\
\Pi_c''(\theta)&=\upsilon(\phi R_o)^2\left[\left(\rho+\phi+\eta\right)^2-2\upsilon\phi R_o\theta\right]^{-3/2}>0
\end{aligned}$$

The optimal monitoring is

$$m(\theta)=\frac{\left(\rho+\phi+\eta\right)-\sqrt{\left(\rho+\phi+\eta\right)^2-2\upsilon\phi R_o\theta}}{\eta}.$$

In equilibrium, trading satisfies

$$\dot{\theta} = -\rho \frac{(\rho + \phi + \eta)^2 - 2\upsilon\phi R_o \theta}{\upsilon \phi R_o} < 0.$$

Finally, note that

$$\Pi_c'(0) = \frac{\phi R_o}{\rho + \phi + \eta} < q_c(0) = \frac{\phi R_o}{\phi + \eta},$$

which implies that the bank has incentives to sell an atom for θ sufficiently low. As in the benchmark model, when the $\theta < \theta^*$, where θ_* satisfies $\Pi_c'(0) = q_c(0)$, the bank liquidates the remaining loans immediately. At θ_*, the bank follows a mixed strategy like the one in the benchmark model.

Intermediation In the case of intermediation, we have

$$\phi \pi_i(\theta) := \max_{D \le \Pi_i(\theta)} \{\phi \theta R_o + \rho D\}$$

The constraint is $D \le \Pi(\theta)$ is always binding so we get $\phi \pi_i(\theta) = \phi \theta R_o + \rho \Pi_i(\theta)$. Following similar steps in the certification equilibrium, the term $\dot{\theta}\left(\Pi_i'(\theta) - q_i(\theta)\right)$ must vanish in the smooth-trading region, so the bank's continuation value satisfies the HJB:

$$(\rho + \eta)\Pi_i(\theta) = \max_m \ \phi \Big[\pi_i(\theta) - \Pi_i(\theta)\Big] + \eta m \Pi_i(\theta) - h(m).$$

Substituting $\phi \pi_i(\theta) = \phi \theta R_o + \rho \Pi_i(\theta)$, we get

$$\eta \Pi_i(\theta) = \max_m \ \phi \Big[\theta R_o - \Pi_i(\theta)\Big] + \eta m \Pi_i(\theta) - h(m).$$

The first order condition is

$$h'(\theta) = \eta \Pi_i(\theta)$$

In the case that $h(m) = km^2/2$, the HJB equation reduces to

$$\frac{\upsilon}{2}\Pi_i(\theta)^2 - (\eta + \phi)\Pi_i(\theta) + \phi \theta R_o = 0,$$

so we get

$$\Pi_i(\theta) = \frac{(\phi + \eta) - \sqrt{(\phi + \eta)^2 - 2\upsilon \phi R_o \theta}}{\upsilon}$$

and

$$m(\theta) = \frac{(\phi + \eta) - \sqrt{(\phi + \eta)^2 - 2\upsilon \phi R_o \theta}}{\eta}$$

Using the envelope theorem, we get that

$$\eta(1 - m(\theta))\Pi_i'(\theta) = \phi \Big[R_o - \Pi_i'(\theta)\Big].$$

From the loan pricing equation,

$$\eta(1-m)q_i(\theta) = \phi\left(R_o - q_i(\theta)\right) + \dot{\theta}q_i(\theta),$$

we plug in $q_i(\theta) = \Pi_i'(\theta)$ and $q_i'(\theta) = \Pi_i''(\theta)$ to get that

$$\dot{\theta} = 0.$$

Notice that

$$(\phi+\eta) - \sqrt{(\phi+\eta)^2 - 2\upsilon\phi R_o\theta} > (\rho+\phi+\eta) - \sqrt{(\rho+\phi+\eta)^2 - 2\upsilon\phi R_o\theta}$$

so $\Pi_i(\theta) > \Pi_c(\theta)$ and $m_i(\theta) > m_c(\theta)$.