# Intermediary Financing without Commitment

Yunzhi Hu and Felipe Varas*

September 2, 2022

**Abstract**

Intermediaries reduce agency frictions through monitoring. To be a credible monitor, an intermediary needs to retain a fraction of its loans; we study the credit market dynamics when it cannot commit to doing so. We compare certification (investors directly lend to borrowers) with intermediation (investors lend through intermediaries). With commitment to retentions, they are equivalent. Without commitment, they are very different. A certifying bank sells its loans and reduces monitoring over time. By contrast, an intermediating bank issues short-term debt to internalize the monitoring externalities and retains its loans. The problem relates to durable-goods monopoly and the leasing solution.

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

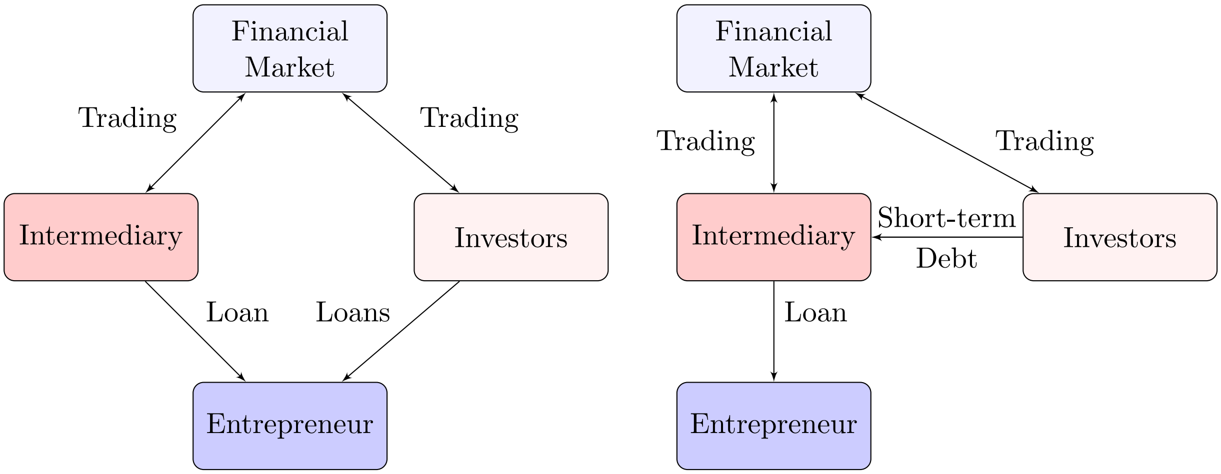
*Hu: Kenan-Flagler Business School, UNC Chapel Hill; Yunzhi_Hu@kenan-flagler.unc.edu. Varas: Fuqua School of Business, Duke University; felipe.varas@duke.edu. The authors would like to thank Douglas Diamond and Raghuram Rajan for very detailed feedback, comments and suggestions. The authors are also 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: 60% of the loans are sold within one month after origination, and nearly 90% are sold within one year (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. Being a credible monitor requires the bank to retain a sufficient fraction of loans on its balance sheet. However, in the presence of the financial market, the bank has incentives to sell its loans, due to its high cost of capital. The more it sells, the less likely it will monitor, and the price of the loans will drop more as well. Motivated by Holmstrom and Tirole (1997), we study two types of intermediary structures: certification and intermediation. As shown in Figure 1, in certification, both the bank and investors directly invest in the borrower's venture, and the role of the bank is to certify that it will monitor. In intermediation, investors put their money in the bank, and the bank then invests a collection of its own funds and those from investors into the borrower's venture. Although the two structures are equivalent in the static framework with the bank's commitment to its retention, without commitment they lead to very different dynamics in loan sales and monitoring. In certification, the lack of commitment induces the bank to sell its loans gradually, and the bank's monitoring intensity declines over time. In intermediation, the bank issues short-term debt to investors, and the repricing of short-term times helps the bank commit to the retention and the decision to monitor. As a result, an intermediating bank never sells its loans, and the entrepreneur is always able to borrow more. Even though the intermediation structure has a higher borrowing capacity, the entrepreneur may still end up choosing certification. In other words, the structure that maximizes borrowing may not be the one that maximizes the borrower's expected payoff.

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




(a) Certification (b) Intermediation

**Figure 1: Certification vs. Intermediation**

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. 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 too aggressively.

Although certification and intermediation are equivalent in a two-period model under bank's commitment to its retention, they lead to different dynamics in loan sales and monitoring if there is no such a commitment. By formulating the problem in continuous time, we characterize the equilibrium and the related trading dynamics under both certification and intermediation. In certification, the lack of commitment hurts the bank as in the standard durable-goods monopoly problem (Coase, 1972). 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. Indeed, a certifying bank has incentives to sell the loans, because its marginal valuation is below that of investors. After its initial loan sell, the bank has incentives to keep selling to exploit the remaining gains of trade. The price of the loans reflects the expectation that the bank will gradually sell the loan, reducing the likelihood of monitoring, which in turn reduces the bank's proceeds from selling. Hence, 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. Similar to the result in the Coase conjecture, the certifying bank does not benefit from its ability to trade loans at all.

By contrast, an intermediating bank does not have incentives to sell, due to its ability to issue short-term debt that matures instantaneously. Indeed, when the bank has access to short-term debt as a source of cheap financing, selling loans not only depresses the valuation of the loan, but also increases the interest rate of short-term debt. The elevated interest rate acts as a mechanism that deters the bank from selling. As a result, the bank finds it optimal to retain the entire loan on its balance sheet even though it has not committed to doing so.

This distinction between certification and intermediation is related to an important externality in bank monitoring. As in Diamond (1984), monitoring suffers from the free-rider problem in that investors enjoy the benefits but do not share the cost with the bank. Therefore, the equilibrium monitoring effort is too low. In certification, this externality leads to the bank reducing its probability of monitoring over time. In intermediation, however, short-term debts help the bank internalize the externalities. In particular, short-term debt creates a market for the monitoring services to be fairly priced: the interest rate would have immediately increased, had the bank sold its retention. As a result, the bank and its investors share both the benefits and costs of monitoring.

Next, 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, even though she borrows less. The reason is that intermediation can lead to excessive monitoring (Pagano and Röell, 1998). Intuitively, the entrepreneur cares about not only the value of the project, but also her future private benefits. The level of monitoring if the bank retains the entire loan may be excessive. Although bank monitoring introduces a positive externality to investors, it also imposes a negative externality to the entrepreneur by reducing her private benefits. This effect dominates if private benefits are sufficiently large, which is why the entrepreneur prefers the certification structure. Our result thus implies that entrepreneurs with higher private benefits – such as those value corporate perks and enjoy control rents – tend to borrow from certifying banks, whereas entrepreneurs that are more financially-constrained – such as small-business owners – tend to borrow from intermediation banks.

Finally, we explore the impact of common policies designed to address the bank's commitment problem. In particular, we look at the impact of a lock-up period and minimum retention levels. Although the qualitative nature of the equilibrium outcome remains unchanged, these policies mitigate the commitment problem and allow the entrepreneur to borrow more upfront. Moreover, if an intermediating bank's liabilities have long maturity or are subsidized by the government, the bank's commitment to retention is also impaired.

### Related Literature

The lack of commitment problem was initially recognized by Coase (1972) in the context of durable-goods monopoly. 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).

The role of short-term debt as a commitment device 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. Put differently, all that the demandable deposits ensure is severe punishment to the banker if defaulted on. 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. Hence, the result in our paper applies to well-capitalized banks or banks with short term but not demandable deposits. Whereas both Calomiris and Kahn (1991) and Diamond and Rajan (2001) emphasize how run externalities create commitment, our paper shows how the market pricing of short-term debt (independent of run externalities) resolves the commitment issue. Our results are also related to Flannery (1994), who presents a model where liabilities and asset mature at the same time, and a one-time investment opportunity arrives before the final maturing date. Flannery (1994) emphasizes the timing between investment and debt issuance, and studies the impact on investment distortions. In Flannery (1994), the notion of maturity is not explicitly specified, Instead. the main driving force is the relative timing between debt issuance and investment opportunity. Debt repricing prevents bad investments by the entrepreneur. 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.[1]

This insight is also 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 affect all buyers, eliminating the monopolist's temptation to discriminate buyers over time. The associated commitment problem is therefore resolved. In our context, an intermediating bank who issues short-term debt against the project's cash flows can be thought 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.

This paper is also related to a large literature on bank monitoring, loan sales and securitization. One example is Parlour and Plantin (2008), who 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 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 still equivalent.

Finally, this paper also makes a technical contribution by applying the method from the optimal control problems theory on stratified domains. Another paper applies similar techniques is Aguiar et al. (2015), who study debt rollovers in monetary unions.

# 2 The Model

Our model builds on the fixed-size investment setup in Holmstrom and Tirole (1997). A key innovation is to introduce a competitive financial market in which the bank can trade its loans. We assume the bank cannot commit to its loan retention, and thus the decisions to monitor. The more it sells, the less likely it will monitor, and the loans will be valued lower as well.

## 2.1 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 liabilities. The entrepreneur starts out with cash level $A$, whereas banks and investors have deep pockets. We assume investors do not discount future cash flows,

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

whereas the entrepreneur and intermediaries discount the future at a rate $\rho > 0$. Investors in our model should be interpreted as informed 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 $\tau_\phi$, 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 $\tau_\phi$, 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 $\tau_\phi$, 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 associated with the low option. We assume the project's expected payoff is always higher if the entrepreneur works; that is, $p_H R > p_L R + B$.

## 2.2 Monitoring, Financial Structures, and Contracts

A competitive set of banks are present at $t = 0$, 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 and the free-rider problem (Diamond, 1984). At $\tau_\phi$, the project matures, and the bank can monitor to eliminate the high shirking option. To do so, it needs to pay a 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 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.[2] Figure 2 describes the 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 *directly* invest in the venture as well. One example of this type of bank is the lead investment bank in loan syndication. In certification, the entrepreneur directly signs the initial contract with the bank and investors. 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.

[2] 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 (which is increasing in $\theta$).

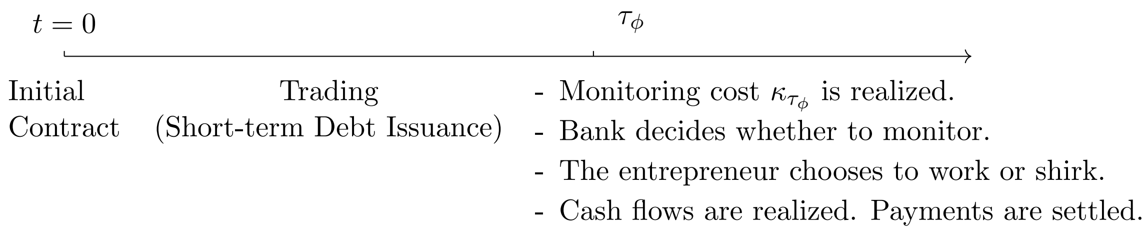



**Figure 2: Timing**

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 invest in the bank, which in turn lends to the entrepreneur a collection of its own funds and the money from investors.[3] One example of this type of bank is a shadow bank which borrows short-term debt from informed lenders. In intermediation, the entrepreneur signs an initial contract with a bank and promises to repay the bank $R_o$ if the project succeeds. The bank, in turn, 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. Below we sometimes refer to investors in the intermediation bank as creditors. We model the short-term debt as one with instant maturity. Debt with instant maturity is the continuous-time analogous to one-period debt in discrete time. Whenever the bank fails to repay its creditors, the remaining loans are sold for repayments. For simplicity, we assume no bankruptcy cost (or loss of charter value) is incurred if the bank fails, and a positive bankruptcy cost will only quantitatively change the results.

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, $R_o = R - b/\Delta$.[4]

[3] 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 bank in the form of short-term debt. In practice, one can interpret the 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.

[4] The entrepreneur can also borrow directly from investors, but the borrowing capacity is lower.

### 2.3 Trading and Pricing in the Financial Market

A competitive financial market is open after time 0 in which loans can be traded.[5] We normalize the total share of loans outstanding to one and use $\theta_t$ to denote the bank's retention at time $t$. In our model, $\theta_t$ will be the payoff-relevant state variable. Sometimes $\theta_t$ is also referred to as the bank's skin in the game, which is publicly observable. Before trading starts, the bank's initial retention is $\theta_0 \in [0, 1]$ in certification and $\theta_0 = 1$ 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. Conditional on the project maturing, let $p(\theta)$ be the equilibrium probability of success; investors of the loan receive

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

per share. Let $q(\theta)$ 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 of the asset:

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

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

The probability of success $p(\theta)$ will differ in certification and intermediation. Let $\kappa$ be the realization of the stochastic monitoring cost $\tilde{\kappa}$ at $\tau_\phi$. In certification, the bank with retention $\theta$ 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 $\theta$ 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.

For the rest of this paper, we restrict the (expected) monitoring cost to be sufficiently low.

[5] 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.

**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., $\theta_t \equiv 1,\ \forall t \leq \tau_\phi$), 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.

# 3 Equilibrium

Subsections 3.1 and 3.2 define and derive the certification and intermediation equilibrium, respectively. In subsection 3.3, we compare the two equilibria and present the constrained-efficient solution. Subsection 3.4 studies the choice between certification and intermediation under the case in which the monitoring cost $\tilde{\kappa}$ follows the uniform distribution and $p_L = 0$, where we obtain closed-form solutions in primitives.

## 3.1 Certification Equilibrium

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

$$\max_{\{\theta_t\}_{t\geq 0}} \mathbb{E}\left[\int_0^\infty e^{-(\rho+\phi)t}\Big(\phi\pi_c(\theta_t)\,dt + dG(\theta_t)\Big)\right], \tag{7}$$

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

$$V_c = \mathbb{E}\left[\int_0^\infty \phi e^{-(\rho+\phi)t}\left\{\mathbb{1}_{\{\kappa\leq\kappa_c\}}p_H R_f + \mathbb{1}_{\{\kappa>\kappa_c\}}(p_L R_f + B)\right\} dt\right]. \tag{8}$$

If the realized monitoring cost is lower than the threshold $\kappa_c$ 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 $\theta$, henceforth, the certification equilibrium.[6]

**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 (7).*
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 $\theta$.[7] In the smooth-trading region, $\Pi_c(\theta)$ satisfies the following Hamilton-Jacobi-Bellman (HJB) equation:

$$\rho\Pi_c(\theta) = \max_{\dot\theta}\ \phi\Big[\pi_c(\theta) - \Pi_c(\theta)\Big] + \dot\theta\Big[\Pi_c'(\theta) - q_c(\theta)\Big]. \tag{9}$$

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

$$\Pi_c'(\theta) = q_c(\theta), \tag{10}$$

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 (10) into

[6] 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 $p_L = 0$.

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

(9), we get that in the region of smooth trading,

$$\Pi_c(\theta) = \Phi \pi_c(\theta). \tag{11}$$

Note that the bank value is equal to the payoff if it never sells any loan until the project matures. In other words, 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 already 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 the one that it always retains the loan, it does not imply the bank will never trade loans on the equilibrium path. In fact, the price of loans following a no-trade strategy will be too high, which gives the bank strict incentives to sell. Indeed, the bank sells the loans for two reasons. First, due to the higher cost of capital, the bank's marginal valuation is below that of investors. Second, after the initial sell and the reduction of retention, the bank is willing to sell again because the price impact only accrues to a smaller number of shares. 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 following asset-pricing equation whenever the bank trades smoothly:

$$0 = \phi \Big[ d_c(\theta) - q_c(\theta) \Big] + \dot{\theta} q_c'(\theta), \tag{12}$$

where $\phi \Big[ d_c(\theta) - q_c(\theta) \Big]$ resembles the dividend income and $\dot{\theta} q_c'(\theta)$ the capital gain.[8] Combining (10), (11), and (12), and using the relation $d_c(\theta) = \pi_c'(\theta)$, one can derive the following equilibrium trading strategies:

$$\dot{\theta} = -\phi \frac{(1-\Phi)\, \pi_c'(\theta)}{\Phi \pi_c''(\theta)} < 0. \tag{13}$$

In the smooth-trading region, the bank sells loans over time and its retentions 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 focused on the case of smooth trading. Meanwhile, the bank also has the

[8] 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. Atomic 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 $\theta_*$ 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 $\theta < \theta_*$.

The final step in the equilibrium construction is to 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; shortly afterwards, 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 $\theta_t$ reaches $\theta_*$, because if so, the price of the loan will experience a deterministic downward jump, inconsistent with the asset-pricing equation (2). The only (stationary) trading strategy at $\theta_*$ consistent with (2) is for the bank to adopt a mixed strategy:[9] the bank sells off all its remaining loans at $\tau_\lambda$, which arrives upon a Poisson event at intensity $\lambda$ that satisfies

$$q_c(\theta_*) = \mathbb{E}\left[d_c(\theta_{\tau_\phi})|\theta_t = \theta_*\right] = \frac{\lambda}{\phi+\lambda} d_c(0) + \frac{\phi}{\phi+\lambda} d_c(\theta_*). \tag{14}$$

Proposition 1 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 in the appendix using results from the theory of optimal control in stratified domains.[10]

**Proposition 1** (Certification Equilibrium)**.** *A unique **certification equilibrium** exists. Given the bank's initial retention $\theta_0$, the bank sells its loans smoothly at a rate given by equation (13) until $T_* := \inf\{t > 0 : \theta_t = \theta_*\}$, after which it sells off its remaining loans at some Poisson rate $\lambda$ that satisfies (14). The equilibrium loan price is*

$$q_c(\theta_t) = \begin{cases} \Phi\Big(p_L + F\left(\Delta R_o \theta_t\right)\Delta\Big) R_o & t < T_* \\ \Big(p_L + \frac{\phi}{\lambda+\phi} F\left(\Delta R_o \theta_*\right)\Delta\Big) R_o & T_* \le t < \tau_\lambda \\ p_L R_o & t \ge \tau_\lambda. \end{cases} \tag{15}$$

[9] 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 $\theta_*$. The price $q_t$ would not be stationary, and it would depend on the trading history before time $t$.

[10] Due to the discontinuity in the price function $q_c(\theta)$, the HJB equation (9) is discontinuous at $\theta_*$. This technical problem can be sidestepped using (discontinuous) viscosity solution methods.

*The contract that maximizes the initial borrowing amount has $\theta_0 = 1$, in which case the borrowing amount is*

$$L_c = \Phi \pi_c (1) . \tag{16}$$

A contract that has the entrepreneur borrow exclusively from the bank at $t = 0$ enables the most upfront borrowing. Intuitively, under this contract, the bank needs to spend the longest period to fully offload its loans. This result is in contrast to that in Holmstrom and Tirole (1997), where the entrepreneur prefers to use as little bank capital as possible, because bank capital is expensive. In our setup, however, higher bank retention slows down the bank's selling process. The longer the bank takes to sell off the entire loan, the more likely the bank will monitor, and the value of the loan is also higher. This channel dominates the one from expensive bank capital.

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. In that case, our model predicts that a bank without any commitment to retention will gradually reduce its exposure to the borrower.

### 3.2 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:

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

where the project succeeds with probability

$$\hat{p}_i (\theta, D) \coloneqq p_L + F (\kappa_i) \Delta, \tag{18}$$

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

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

compensates the default risk borne by creditors.[11] In intermediation, the bank trades loans and issues short-term debt to maximize the expected payoff upon the project maturing, together with the net income from short-term debt issuance and trading gains $dG(\theta_t)$; that is,

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

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 $\theta_t$ as the state variable. From (21), we know the choice of $D$ only involves static tradeoff. A higher $D$ reduces the probability of success $\hat{p}_i\left(\theta, D\right)$ (see equation (4) and (18)), thereby reducing the value of loans. On the other hand, the term $\rho D$ emerges because debt is a cheaper source of funding. Debt issuance is bounded by the endogenous constraint $D \leq \Pi_i\left(\theta\right)$, which arises from the bank's limited liabilities.[12] 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 $E\left(\theta, D\right)$ is equivalent to maximizing the total bank value $\Pi_i\left(\theta\right)$, because debt $D$ is fairly priced.

[11] A heuristic derivation of (19) goes as follows. The approximate probability that the project matures over a period of length $dt$ is $\phi dt$, 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.

[12] Here, $\Pi_i\left(\theta\right)$ is the bank's value function given its retention $\theta$, 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).

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

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

The expressions differs from (8) in that the threshold cost for monitoring is replaced by $\kappa_i$. We look for a Markov perfect equilibrium in state variable $\theta_t$, 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 $\theta_0 \in [0,1]$, $(\theta_t)_{t\geq 0}$ 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.

### 3.2.1 Short-term Debt Issuance

Plugging (17) and (19) into (21) and dividing both sides by $\rho + \phi$, we get:

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

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 $\theta R_o$; $\int_0^{\kappa_i}\kappa dF(\kappa)$ 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. We impose conditions (Assumption 2 in the appendix) such that $\mathcal{V}(D,\theta)$ 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 $\theta$, $D \leq \Pi_i(\theta)$. We have the following result.

**Lemma 2.** *Let $D^*(\theta)$ be the optimal choice of short-term debt. There exists a threshold in $\theta$ where $D^*(\theta) = \Pi_i(\theta)$ if and only if $\theta$ falls below this threshold.*

According to Lemma 2, the bank finances the loan using both short-term debt and bank capital when $\theta$ is high. When $\theta$ is 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.

### 3.2.2 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:

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

and the equilibrium price is determined by the indifference condition $\Pi_i'(\theta) = q_i(\theta)$. The trading strategy follows from the asset-pricing equation

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

Applying the envelope theorem in Milgrom and Segal (2002), we get

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

where $z(\theta)$ is the Lagrange multiplier of the debt-issuance constraint $D \leq \Pi_i(\theta)$. The equilibrium trading rate (derived using condition (43) in the appendix) is

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

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

Recall that in certification, the bank has incentives to sell the loan because investors have a higher valuation. Lemma 3 shows in intermediation, these incentives vanish. Instead, the bank wants to increase its retentions. This distinction arises because more retention enables an intermediating bank to issue cheaper and more short-term debt. To see this, compare $\pi_c'(\theta) = d_c(\theta)$ with (26), which clearly shows an increase in $\theta$ leads to two additional benefits in intermediation. The first benefit is characterized by the term $\left[ f(\kappa_i) \frac{\partial \kappa_i}{\partial \theta} \Delta \right] D^*$, which shows the marginal effect of retention $\theta$ on $f(\kappa_i) \frac{\partial \kappa_i}{\partial \theta} \Delta$, 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

$z(\theta)\Pi_i'(\theta)$, which is only positive if $D^*(\theta) = \Pi_i(\theta)$. Here, an increase in $\theta$ relaxes the constraint on debt issuance so that the bank can issue *more* debt. Note this second benefit disappears whenever the debt-issuance constraint is slack, that is, $z(\theta) = 0$.

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 price $q(0) = p_L R_o$ 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 $\theta$ will increase toward $\theta = 1$. Therefore, the price of the loan satisfies $q_i(\theta_\dagger) = \Phi\pi'(\theta_\dagger)$.[13] The following proposition describes the results.

**Proposition 2** (Intermediation Equilibrium)**.**

1. *There exists an unique $\theta_\dagger$. If $\theta \geq \theta_\dagger$, the bank buys loans smoothly at the rate given by equation* (27)*, and the price of the loan satisfies $q_i(\theta) = \Phi\pi_i'(\theta)$, where $\pi_i'(\theta)$ satisfies* (26)*. If $\theta < \theta_\dagger$, the bank sells off the remaining loan immediately at the price $q_i(\theta) = p_L R_o$.*

2. *In the unique* ***intermediation equilibrium*** *where $\theta_0 = 1$, the bank holds $\theta_t = 1$ until the project matures, and maintains a constant debt level $D_t = D^*(1)$. The equilibrium loan price is $q_i(1) = \Phi\pi_i'(1)$.*

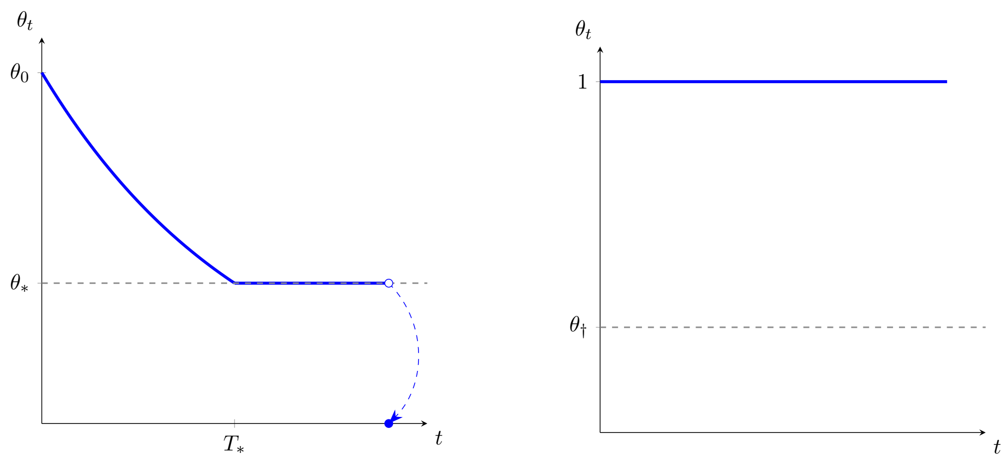
Proposition 2 shows that an intermediation bank finds it optimal to retain the loan and issues short-term debt against it. If, for some reason, the bank's retention fell below one, the intermediating bank would buy the loan back from outside investors over time. The result that banks buy back loans to increase retention could also be interpreted as banks issuing additional loans to the same borrower. As a result, the bank has more exposures and skin in the game. Regardless of the interpretation, the broader message is that the bank prefers issuing short-term debt against the loan over directly selling it, because the former offers commitment on monitoring. A crucial element behind the commitment is that investors are informed and they would increase the interest rate on the short term debt had the bank reduced its retention. Therefore, the intermediation bank should not be interpreted as the traditional commercial banks which rely 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. As shown by Jiang et al. (2020), shadow banks originate long-term loans by issuing short-term debt to a few informed lenders.

[13] 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 $\theta_t$ does not "see" the discontinuity.

## 3.3 Equilibria Comparison and Constrained Efficiency

### 3.3.1 Comparison and Discussion

Figure 3 compares the bank's loan retention dynamics in certification (left) and intermediation (right). In the certification equilibrium, the bank with retention $\theta_0$ first sells the loan smoothly. After $\theta_t$ reaches $\theta_*$, the bank sells off all the remaining loans, following a stochastic delay. By contrast, an intermediation bank starts with $\theta_0 = 1$ and always retains the loan until the project matures.




(a) Certification (b) Intermediation

**Figure 3: Retention Patterns**

The different patterns in retention also imply different probability of monitoring. Specifically, a comparison between the threshold monitoring costs $\kappa_c$ in (3) and $\kappa_i$ in (4) shows that under the same level of retention $\theta$, an intermediating bank monitors less due to its outstanding debt $D$. Meanwhile, an intermediating bank also has a higher retention $\theta$, so that the overall comparison of monitoring is ambiguous. 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., $\theta_0 = 1$. A higher threshold is associated with a higher probability of monitoring if the project matures at time $t$. Early on, $\kappa_c$ is higher, but eventually, $\kappa_c$ falls below $\kappa_i$. Our model thus predicts that a certifying bank conducts more monitoring during the early stage of the loan, whereas an intermediating bank conducts more monitoring during the late stage.

What are the fundamental mechanisms behind the differences in retention and monitoring dynamics? Monitoring has the property of the public goods in that investors can free ride on the bank. All creditors share the benefits from monitoring, whereas the bank exclusively bears the

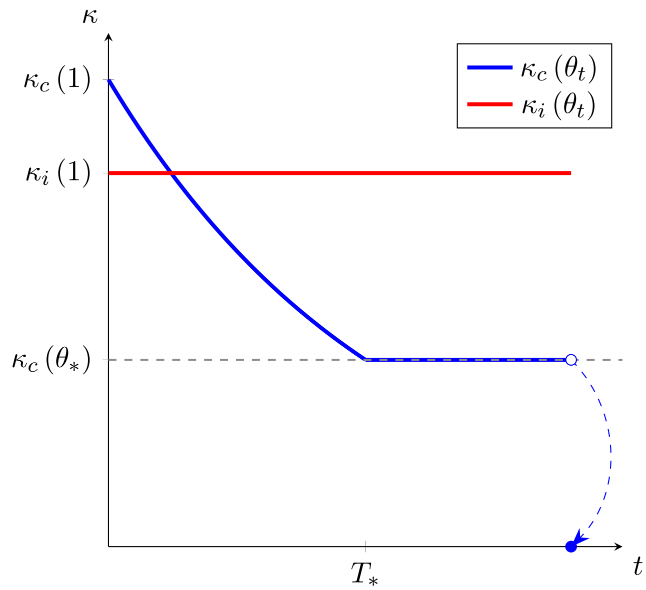



**Figure 4: Threshold Monitoring Cost**

cost. Therefore, the equilibrium monitoring effort is inefficiently low, and in certification, the bank reduces its probability of monitoring over time.

In intermediation, the role of short-term debt is to help the bank internalize the externality from monitoring. 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. Essentially, short-term debt creates a market for the services offered by the bank, that is, monitoring, to be fairly priced. This argument highlights the importance of short-term debt in aligning the bank's monitoring incentives. Calomiris and Kahn (1991) and Diamond and Rajan (2001) also emphasize the role of short-term debt in commitment, the mechanism of these papers relies on the demandable feature of deposits and run externalities. By contrast, the channel in our paper is fundamentally different: it does not rely on run externalities but depends on the pricing mechanism whereby the bank's retention choice directly affects the cost (and also the amount) of short-term debt.

### 3.3.2 The Constrained Social Planner's Problem

The social planner chooses the bank's retention $\{\theta_t\}_{t\geq 0}$ to maximize the aggregate social welfare, subject to the constraint that the bank decides whether to monitor under its retention. Given there is commitment to $\{\theta_t\}_{t\geq 0}$, the choices of debt $D_t$ is redundant. The planner's problem is as follows:

$$W = \max_{(\theta_t)_{t\geq 0}} \int_0^\infty \phi e^{-\phi t}\Bigg\{ \left(1-e^{-\rho t}\right)(1-\theta_t)d\left(\theta_t\right) \\ + e^{-\rho t}\left[p(\theta_t)R + \left(1-F(\Delta R_o \theta_t)\right)B - \int_0^{\Delta R_o\theta_t} \kappa dF\left(\kappa\right)\right]\Bigg\}dt, \quad (28)$$

subject to constraint (3). We show in the appendix that the optimal retention always satisfies $\theta_t < 1$. Due to differences in time discounting, the flow payoff in (28) is a weighted sum of the payoff to investors $(1-\theta_t)d\left(\theta_t\right)$ and the payoff to the bank and the entrepreneur in the bracket. The project succeeds with probability $p(\theta_t)R$, and the entrepreneur shirks to receive the high private value $B$ with probability $(1-F(\Delta R_o\theta_t))$. The optimal retention $\theta_t$ balances the benefits and costs of bank monitoring. Moreover, the difference in time discounting makes this tradeoff time-varying, which implies the optimal retention is in general time-varying as well.

### 3.4 The Initial Choice between Certification and Intermediation

This subsection studies the initial choice between certification and intermediation at time 0. To simplify the analysis and obtain closed-form solutions for the initial payoffs, we specialize the analysis to the case in which the monitoring cost $\tilde{\kappa}$ follows the uniform distribution on $[0, \bar{\kappa}]$ and the probability of success is $p_L = 0$ if the entrepreneur shirks.[14] As a result, 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. The constrained optimal solution satisfies

$$\theta_t = \frac{1}{2-e^{-\rho t}}\left[\left(1-e^{-\rho t}\right) + e^{-\rho t}\frac{\Delta R - B}{\Delta R - b}\right],$$

with $\theta_0 = \frac{\Delta R - B}{\Delta R - b}$ and $\lim_{t\to\infty}\theta_t = \frac{1}{2}$. Depending on the value of $B$ relative to $b$, $\theta_t$ can be either decreasing or increasing.

Next, we describe the certification and intermediation equilibrium and study the entrepreneur's choice between the two. In certification, for any initial $\theta_0$, the amount that the entrepreneur can borrow at $t = 0$ and her payoff from the project maturing are

$$L_c\left(\theta_0\right) = \Phi_c(\theta_0) + p_c(\theta_0)(1-\theta_0) = \frac{\Phi}{2\bar{\kappa}}\left(\Delta R_o\theta_0\right)^2 + \frac{\Phi}{\bar{\kappa}}\left(\Delta R_o\right)^2\theta_0\left(1-\theta_0\right)$$
$$V_c\left(\theta_0\right) = \Phi B - \frac{\Phi}{2-\Phi}\frac{\Delta R_o\theta_0}{\bar{\kappa}}\left(B-b\right).$$

[14] In the durable-goods monopoly literature, this is referred to as the "no-gap" case.

Define $\theta_0^* = 1 - \frac{1}{2-\Phi}\frac{B-b}{\Delta R_o}$, which maximizes the entrepreneur's overall payoff $V_c(\theta_0) + L_c(\theta_0)$. One interesting property of $\theta_0^*$ is that it decreases with $B - b$. Intuitively, this result holds because bank monitoring imposes negative externalities on the entrepreneur by reducing her private benefits. The entrepreneur chooses $\theta_0$ to maximize $V_c(\theta_0) + L_c(\theta_0)$, subject to the feasibility constraint that $L_c(\theta_0) \geq I - A$. Note that $L_c(\theta_0)$ is increasing in $\theta_0$, so 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 $\theta_0$ in closed form:

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

where $\theta_0^{\min}$ is the minimum value of $\theta_0$ 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 $\Phi$ is sufficiently high. Intuitively, a high $\Phi$ implies the bank has a relatively low cost of capital; that is, $\rho$ is sufficiently low.[15] 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 her payoff from the project maturing are

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

**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 $I - A \in \left(L_c(1), L_i(1)\right)$, only intermediation is feasible to the entrepreneur. For the remainder of this subsection, we focus on the case that both financing structures are feasible; that is, $I - A < L_c(\theta_0^*)$. Our result shows that the entrepreneur prefers certification to intermediation if and only if the private benefit without monitoring is sufficiently high.

[15] 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 $\Phi < \sqrt{\frac{\left(\bar{\kappa}/\Delta^2\right)^2}{R_o^2 + \left(\bar{\kappa}/\Delta^2\right)^2}}$

**Proposition 3.** *Under uniform distribution and $p_L = 0$, a threshold $B^*$ exists. If $L_c(\theta_0^*) > I - A$, then $V_c(\theta_0^*) + L_c(\theta_0^*) \geq V_i + L_i(1)$ if and only if $B > B^*$; that is, the entrepreneur prefers certification to intermediation if and only if $B > B^*$.*

Let us offer some explanations to this result. Because the entrepreneur is always able to borrow more under intermediation, the only reason why she may still prefer certification is that she receives a higher payoff from the project maturing; that is, $V_c(\theta_0^*) > V_i$. Simple calculation shows that

$$V_c(\theta_0^*) - V_i = \Phi \frac{\Delta}{\bar{\kappa}} (B - b) \left[ R_o \left( 1 - \frac{1}{2 - \Phi} \theta_0^* \right) - D^*(1) \right].$$

The first term in the bracket $R_o \left(1 - \frac{1}{2-\Phi}\theta_0^*\right)$ captures the additional benefit that the entrepreneur is able to receive due to the gradual reduction in monitoring in the case of certification. The second term, $D^*(1)$, captures the reduction in the intermediating bank monitoring caused by debt. Conditional on $V_c > V_i$, a higher $B$ leads to a larger difference in the entrepreneur's payoff. If the difference becomes sufficiently large, it can offset the difference in the amount of borrowing, $L_i(1) - L_c(\theta_0)$. In this case, the entrepreneur ends up choosing certification.

The previous result is reminiscent of previous work in corporate governance on the potential of over-monitoring by large shareholders (Pagano and Röell, 1998).[16] 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 future benefits as the manager. Therefore, a level of monitoring that maximizes the firm value may be excessive to the entrepreneur. Our result implies more financially-constrained entrepreneurs borrows from intermediating banks, whereas entrepreneurs with more net worth and who potentially enjoy higher private benefits (or control rents) tend to borrow from certifying banks.

One may wonder whether the entrepreneur may benefit by simultaneously borrowing from both an intermediating bank and investors; that is, $\theta_0 \in (0, 1)$ in the intermediation setup. The answer is yes. In some situations, the constrained-optimal allocation features the bank's retention $\theta_t$ increasing over time. One example is given by the solution to the planner's problem when $\theta_0 = \frac{\Delta R - B}{\Delta R - b} < \frac{1}{2}$. This solution can be better approximated by combining certification and intermediation, in which the bank buys loans over time.

**Numerical Example** We illustrate the comparison using the following numerical example. Let $\phi = 1$, $b = 0$, $R = 2$, $p_H = 0.8$, $p_L = 0$, $\rho = 0.01$, and $\kappa$ follow the uniform distribution on $[0, 2]$. Under given parameters, we can get that $D^*(1) = 0.0312$, $\Pi_i(1) = 0.3122$, and $\kappa_i = 1.5750$.

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

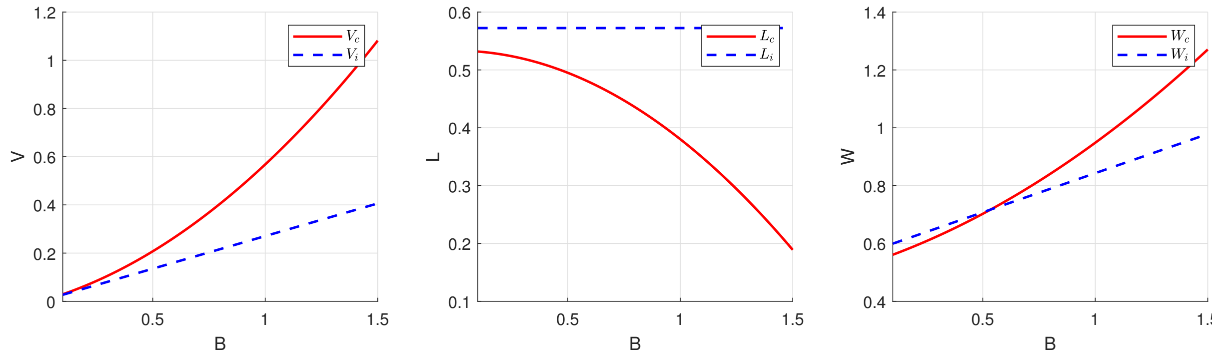



**Figure 5: Valuation under Certification, Intermediation, and Direct Lending**

Figure 5 illustrates the payoffs when the private benefit $B$ varies. The red (solid) and blue (dashed) lines respectively stand for certification and intermediation. The left panel describes the entrepreneur's payoff, which increases with $B$ in both cases. The entrepreneur always obtains a higher payoff in certification, due to the negative externalities of bank monitoring on the entrepreneur. The middle panel shows the maximum amount of borrowing, which is independent of $B$ in intermediation (because $\theta_t \equiv 1$ ) but decreases in certification (because $\theta_t$ declines over time). Finally, the right panel compares the overall payoff. Clearly, certification has a higher overall payoff than intermediation once $B$ becomes sufficiently high.

**Bank's choice between certification and intermediation.** We have studied the entrepreneur's choice between the two structures. What are the banks' incentives to choose between the two structures? For any $\theta_0$ chosen by the entrepreneur, we can show that a certifying bank's payoff $\Pi_c(\theta_0)$ always falls below an intermediating bank's payoff $\Pi_i(\theta_0)$. This is because the problem of the certifying bank is the one of the intermediation bank with the additional constraint that $D_t \equiv 0$, $\forall t$. Therefore, a bank chooses to issue short-term debt if it has the option.[17]

# 4 Extensions

## 4.1 Certification: Lockup Period and Minimum Retention

### Lockup Period

We introduce a lockup arrangement that allows the bank to commit to $\theta_0$ for a period $[0, t_\ell]$. Due to the stationary environment, the subgame starting from $t_\ell$ and the associated equilibrium

[17] This result holds in the general model when the monitoring cost $\tilde{\kappa}$ no longer follows the uniform distribution and $p_L > 0$.

are unchanged from those in 3.1. Between $[0, t_\ell]$, no trading occurs, and the bank's flow payoff is $\phi\pi_c(\theta_0)$. Let $L_\ell$ be the total lending at $t = 0$,

$$L_\ell = \Pi_c(\theta_0) + (1-\theta_0)\left[d_c(\theta_0)\left(1 - e^{-\phi t_\ell}\right) + e^{-\phi t_\ell} q_c(\theta_0)\right].$$

Again, this expression confirms the earlier result that due to the lack of commitment, the certifying bank does not benefit from its ability to trade loans at all: it is able to lend exactly $\Pi_c(\theta_0)$ regardless of $t_\ell$. Meanwhile, investors are willing to lend more as the lockup period $t_\ell$ becomes longer, because $q_c(\theta_0) = \Phi d_c(\theta_0)$. Therefore, $\frac{\partial L_\ell}{\partial t_\ell} > 0$ so that the lockup period increases the total amount of lending: the incremental lending comes from investors' willingness to lend due to the bank's commitment during $[0, t_\ell]$. Moreover, we use $V_\ell$ to denote the entrepreneur's overall payoff. Under uniform distribution and $p_L = 0$,

$$V_\ell = \Phi\left[\frac{\Delta R_o \theta_0}{\bar{\kappa}} b + \left(1 - \frac{\Delta R_o \theta_0}{\bar{\kappa}}\right) B\right]\left(1 - e^{-(\rho+\phi)t_\ell}\right) + e^{-(\rho+\phi)t_\ell} V_c(\theta_0).$$

Intuitively, during the lockup period $[0, t_\ell]$, the entrepreneur is monitored and receives $b = \Delta \cdot R_f$ with probability $\Delta R_o \theta_0 / \bar{\kappa}$; otherwise, she receives $B$. It is easily derived that $\frac{\partial V_\ell}{\partial t_\ell} < 0$, so that a longer lockup period leads to a lower payoff to the entrepreneur. Intuitively, the entrepreneur receives $b$ with monitoring but $B$ without monitoring. Although a longer lockup period increases monitoring, it also reduces the entrepreneur's equilibrium payoff.

Let $W_\ell = V_\ell + L_\ell$ be the aggregate payoff, which is also the objective function that the entrepreneur tries to maximize at $t = 0$. Under uniform distribution and $p_L = 0$, a longer lock-up period always increases the aggregate social welfare.

**Corollary 1** (Optimal Lockup Period)**.** *Given the optimal initial loan retention, $\theta_0 = \theta_0^*$, the aggregate social welfare attains the maximum as $t_\ell \to \infty$.*

**Minimum Retention**

According to section 941 of the Dodd-Frank Act, securitizers are required to retain no less than 5% of the credit risks associated with any securitization to perform intermediation services. This rule is commonly known as risk retention. In February 2018, the circuit court exempted CLO funds, We evaluate such a policy by imposing a minimum retention requirement on the bank.[18]

Suppose the bank must hold at least a fraction $\underline{\theta}$ of the loans on its balance sheet. The equilibrium is qualitatively similar to the one in Proposition 1: some threshold $\tilde{\theta}$ exists such that smooth trading occurs for $\theta \in (\tilde{\theta}, 1]$, and an atom $\tilde{\theta}$ exists where the holdings jump to $\underline{\theta}$. Similar to

[18]Securitization is identical to loan sales in our setup, given the binary outcome of the final cash flows.

the certification equilibrium, the jump from $\tilde{\theta}$ to $\underline{\theta}$ must also happen with a random delay. Details are available in the appendix.

### 4.2 Intermediation: Alternative Bank Liability Structures

In this subsection, we study the intermediation equilibrium under alternative liability structures. This exercise highlights 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. Specifically, we are going to show that the bank's commitment to retention will no longer hold under long-term debt, nor under debt that is sufficiently subsidized (such as FDIC-insured deposits).

#### Long-term Debt/Outside Equity

We solve the intermediation equilibrium in which the bank can issue only long-term debt at $t = 0$, that is, debt that only matures with the project at $\tau_\phi$. Given that the final cash flow has a binary outcome of either $R$ or 0, long-term debt is identical to outside equity. Let $D_0 = D$ be the amount of long-term debt that the bank issues at $t = 0$. Note that after $t = 0$, the bank can no longer raise any further debt. The definitions for $\kappa_i$, $\hat{p}_i(\theta, D)$, and $\hat{\pi}_i(\theta, D)$ follow from those earlier. Because $D_0 = D$ is only chosen at $t = 0$, we suppress these functions' dependence on $D$, and therefore refer to $\kappa_i(\theta)$, $p(\theta)$, and $\pi_i(\theta)$, respectively. The bank solves

$$\max_{\{\theta_t,\, D_0\}} \, (1 - y_0) D_0 + \mathbb{E}\left[\int_0^\infty e^{-(\rho+\phi)t}\Big(\phi\pi_i(\theta_t)\, dt + dG(\theta_t)\Big)\right],$$

where $y_0$ compensates the creditors for the default risk. Note that after $t = 0$, the bank only chooses its trading strategy, and the solution method as well as the equilibrium outcomes naturally follow the certification equilibrium in subsection 3.1. In fact, these two are identical because creditors are effectively investors who directly lend to the entrepreneur.[19] This extension highlights the importance of short-term debt in commitment. Given this result, a certifying bank can be equivalently understood as an intermediating bank but issues long-term debt.

#### Debt Subsidies

Next, we introduce an extension where the interest rate of short-term debt is partially subsidized by the government. One can think about this subsidy as either deposit insurance or the implicit

[19] A subtle difference is that long-term creditors cannot trade their debt, whereas investors in certification can sell the loans. However, given that in the certification equilibrium, the bank sells the loan (and, equivalently, investors buy the loan) over time, this difference does not affect the result. In other words, in the equilibrium with long-term debt, creditors buy loans in the secondary market after $t = 0$.

guarantee from a government bailout.[20] We show that once the subsidy becomes sufficiently high, an intermediating bank can no longer commit to its retention but instead sells loans over time, just as a certifying bank. This exercise highlights the importance of the interest rate in aligning the bank's incentives to commit to its retentions.

Specifically, we assume the bank only needs to pay a fraction $\xi$ 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 3.2, in which short-term debt issuance and trading are solved sequentially.

**Proposition 4.** *A $\xi_\dagger$ exists such that in the intermediation equilibrium, $\dot{\theta} < 0$ if $\xi < \xi_\dagger$ for $\theta$ sufficiently large.*

Intuitively, the bank no longer has the incentives to retain its loans if the short-term debt is mostly subsidized by the government and when $\theta_t$ is high such that the debt-issuance constraint is slack. If $\theta_t$ is low, this result is no longer true, because the debt-issuance constraint binds; that is, $D^*(\theta) = \Pi_i(\theta)$. In this case, the bank will have incentives to retain loans. The overall effect in this case combines the two.

### Bail-in vs. Bailout

Let us modify the model by assuming if the project fails and generates nothing, the bank is able to pay the creditors up to $X > 0$. 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.[21] The bank's incentive compatibility constraint in monitoring becomes $\kappa_i = \Delta(R_o\theta + X - D)$. Moreover, it is never optimal for the bank to issue risk-less debt. In other words, the endogenous choice of debt always satisfies $D > X$, with an interest rate $\hat{y}(\theta, D) = \phi\big(1 - \hat{p}(\theta, D - X)\big)\left(1 - \frac{X}{D}\right)$.[22] The equilibrium in this case

[20] 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 $\xi$ 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 $\xi$ captures the fraction of deposits that are above the limit.

[21] One example is the liquidity-coverage ratio (LCR).

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

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

is qualitatively unchanged from the intermediation equilibrium described in Proposition 2, instead of the one in Proposition 4.

The difference between $X$ and $\xi$ can be interpreted as the bail-ins vs. bailouts. Once again, this difference highlights the importance of monitoring externalities. When short-term debt is subsidized, only a fraction of the externalities is internalized, so the bank's incentives to monitor are also reduced. In the extreme case where the interest rate is independent of the bank's retention (and therefore monitoring), the results go back to the case of certification. By contrast, if a bail-in occurs and the bank compensates the creditors' part of the losses ($X/D$) in the case of a bank failure, the monitoring externalities are still internalized.

# 5 Final Remarks and Discussion

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 liability structure of the intermediary has an impact on 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 issuing short-term debt does not have incentives to sell loans. As a result, certification is associated with a lower amount of lending capacity compared to intermediation. However, the structure that maximizes lending capacity may not be the one that also maximizes the entrepreneur's expected payoff. We show an entrepreneur with high net-worth chooses to borrow under certification if the private benefits are sufficiently high. Certification has more monitoring during the early periods after loan origination, whereas intermediation has more monitoring during the later periods.

Throughout the paper, we have made several simplifying assumption. First, the project does not generate any interim cash flows, and that the entrepreneur's effort and the bank's monitoring are only required when the project matures. Due to the exponential arrival of final cash flows, this model is equivalent to a model in which independent cash flows are generated over time, and the levels of interim cash flows require continuous monitoring and entrepreneur effort. Second, following Holmstrom and Tirole (1997), we assume that all projects financed by an intermediary are perfectly correlated and thus abstract from bank's ability to pooling 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 mighty include the adoption of advanced information technology, the hiring of qualified loan officers, a 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 about 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 changes rapidly with new business opportunities. As a result, these assets do not have contractible, easily described risk properties.

Our main focus has been on 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.[23] 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 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 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).

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

# 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.

Aguiar, M., M. Amador, E. Farhi, and G. Gopinath (2015). Coordination and crisis in monetary unions. *The Quarterly Journal of Economics 130*(4), 1727–1779.

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.

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.

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.

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.

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

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.

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.

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.

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 1.

**Lemma 4.** *The bank with retention $\theta$ never sells a fraction of the loans.*

*Proof.* Suppose the bank with retention $\theta$ sells $\theta^+ - \theta$, where $\theta^+ > 0$, and that after this it continuous trading 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

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

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. $\square$

### 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 right continuous function with left limits. A trading strategy is admissible if it can be decomposed as

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

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

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

Due to the discontinuity in the price price function $q(\theta)$, the Hamilton-Jacobi-Bellman (HJB) equation is discontinuous at $\theta_*$, 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 $\theta_t^{\epsilon*}$ be and $\epsilon$-optimal policy (at this point in the proof we have not established 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 $\Delta_k$ if $N$ is large enough as $|\dot{\theta}_t^{\epsilon c*}| \leq M$ for some finite $M$). Consider the policy $\hat{\theta}_t^N = \theta_t^{\epsilon*}$ if $t \notin \cup_{k\geq 0}(t_k - \Delta_k, t_k)$, and $\hat{\theta}_t^N = \theta^{\epsilon*}_{t_k-\Delta_k} + \text{sgn}(\theta^{\epsilon*}_{t_k} - \theta^{\epsilon*}_{t_k-})N(t - t_k + \Delta_k)$ 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

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

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

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

The price function is right continuous so $q(\theta^{\epsilon*}_{t_k}+) = q(\theta^{\epsilon*}_{t_k})$. 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 $\epsilon$ 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 $\lim_{N\to\infty} \Pi_N^*(\theta_0) = \Pi^*(\theta_0)$. $\square$

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 $\theta_*$ 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 $H$

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

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 $\varepsilon$ small enough $\varepsilon q_c(0) > \Phi\pi_c(\varepsilon)$. Thus, it follows from continuity that a non trivial 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 $\theta_*$ 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 $\theta$, which we denote by $G'(\theta)$.

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

From here we get that

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

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

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

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

**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

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

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

$$\Pi_N''(\theta) = \frac{(\rho+\phi)}{N}\left(\Pi_N'(\theta) - \Phi\pi'(\theta)\right)$$
$$\Pi_N'''(\theta) = \frac{(\rho+\phi)}{N}\left(\Pi_N''(\theta) - \Phi\pi''(\theta)\right).$$

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

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

$\tilde{\theta}_N q_0 > \Phi\pi(\tilde{\theta}_N)$ because $\theta q_0$ 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 $\Pi_N(\tilde{\theta}_N+) > \Phi\pi(\tilde{\theta}_N)$. □

**Lemma 9.** *Let*

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

*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. □

## 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 $[0, 1]$ to 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 $\theta_*$, a viscosity solution might fail to be unique. In order to fully characterize the value function we need to specify its behavior at $\theta_*$. 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

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

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 $\{\theta_*\}$.

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

**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*

*$\theta_*$, 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 $\theta_*$, 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 $\theta_*$. $H_1^+$ consider controls that starting at $\theta_*$ take $\theta_t$ towards the interior of $[\theta_*, 1]$, and $H_0^-$ considers controls that starting at $\theta_*$ take $\theta_t$ 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 $\Pi_N$ defined in (35) is the value function of the control problem (31).

**Verification** Lemmas 7 and 8 imply that $\Pi_N$ is a classical solution on $\Omega \setminus \{\tilde{\theta}_N, \theta_*\}$ so we only need to verify the conditions for a viscosity solution on $\{\tilde{\theta}_N, \theta_*\}$. $\Pi_N$ defined in (35) is a classical solution on $(\theta_*, 1)$. At $\theta = \theta_*$, $\Pi_N$ 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 $\theta_*$, 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

$$\begin{aligned}
(\rho+\phi)\Pi_N(\theta_*) - H_1^+\big(\theta_*, \Pi_N'(\theta_*+)\big) &= (\rho+\phi)\Pi_N(\theta_*) - \phi\pi(\theta_*) = 0 \\
(\rho+\phi)\Pi_N(\theta_*) - H_0^-\big(\theta_*, \Pi_N'(\theta_*-)\big) &= (\rho+\phi)\Pi_N(\theta_*) - \phi\pi(\theta_*) = 0.
\end{aligned}$$

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 verify that

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

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 $\theta_*$ is easily verified as being unique.

# B Intermediation

We start by supplementing an assumption on the distribution of the monitoring cost $\kappa$.

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

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

$$\frac{f'}{f^2} < \frac{\Phi}{1-\Phi}\Delta.$$

Assumption 2 is satisfied by most commonly-used distribution functions such as the uniform distribution.

## 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

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

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

## Proof of Lemma 2 and 3

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

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

To finish the derivation of optimal debt choice $D^*(\theta)$, 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 $f'(\kappa)$, and will be satisfied in the uniform-distribution case. The upper bound guarantees that, for any given $\theta$, 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 $\Pi_i(\theta)$ will be convex in $\theta$.

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

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

Note that from the solution to $D^*(\theta)$, it is immediately clear that $z(\theta) \leq \frac{\rho}{\phi}$. 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 2

*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 $\theta_D$ 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 $\theta < \theta_\dagger$.

□

## The Constrained social planner's problem

*Proof.* Given any $\{\theta_t\}_{t\geq 0}$, the entrepreneur's payoff is $\int_0^\infty e^{-(\rho+\phi)t}\phi v(\theta_t)dt$, whereas the bank receives $\int_0^\infty e^{-(\rho+\phi)t}\phi\pi(\theta_t)dt$ and investors receive $\int_0^\infty e^{-\phi t}\phi(1-\theta_t)d(\theta_t)dt$.

$$\begin{aligned} v(\theta) &= p(\theta)(R - R_o) + (1 - F(\kappa_c))B \\ \pi(\theta) &= p(\theta) R_o \theta - \int_0^{\kappa_c} \kappa dF(\kappa) \\ d(\theta) &= p(\theta) R_o. \end{aligned}$$

From here, we get that the planner's problem is

$$\begin{aligned} W &= \max_{(\theta_t)_{t\geq 0}} \int_0^\infty \phi e^{-\phi t}\Big[(1-\theta_t)d(\theta_t) + e^{-\rho t}\left(v_c(\theta) + \pi_c(\theta)\right)\Big]dt \\ &= \max_{(\theta_t)_{t\geq 0}} \int_0^\infty \phi e^{-\phi t}\left\{\left(1-e^{-\rho t}\right)\left[(1-\theta_t)p(\theta_t)R_o\right] + e^{-\rho t}\left[p(\theta_t)R + (1-F(\kappa_c))B - \int_0^{\kappa_c}\kappa dF(\kappa)\right]\right\}dt. \end{aligned}$$

We can maximize the previous expression pointwise to get the first order condition

$$(1-\theta_t)f(\kappa_c)\Delta R_o\left(1-e^{-\rho t}\right)\frac{\partial \kappa_c}{\partial \theta} - p(\theta_t)\left(1-e^{-\rho t}\right)R_o + e^{-\rho t}f(\kappa_c)(\Delta R - B - \kappa_c)\frac{\partial \kappa_c}{\partial \theta} = 0$$

Substituting $\kappa_c = \Delta R_o \theta$ we get

$$(1-\theta_t)f(\kappa_c)\left(1-e^{-\rho t}\right)(\Delta R_o)^2 - (p_L + F(\kappa_c)\Delta)\left(1-e^{-\rho t}\right)R_o + e^{-\rho t}f(\kappa_c)(\Delta R - B + \kappa_c)\Delta R_o = 0$$

Under uniform distribution and $p_L = 0$, this simplifies to

$$\theta_t = \frac{\Delta R - b - e^{-\rho t}(B - b)}{(\Delta R - b)\left(2 - e^{-\rho t}\right)}.$$

Differentiating we get

$$\dot{\theta}_t = \frac{\rho e^{\rho t}(2B - b - \Delta R)}{(\Delta R - b)\left(2 - e^{\rho t}\right)^2}.$$

Therefore,

$$\dot{\theta}_t < 0 \Leftrightarrow \frac{\Delta R - B}{B - b} > 1.$$

At time zero we have

$$\theta_0 = \frac{\Delta R - B}{\Delta R - b},$$

and as $t \to \infty$

$$\lim_{t\to\infty} \theta_t = \frac{\Delta R - b}{2(\Delta R - b)} = \frac{1}{2}.$$

The expression for the welfare function is straightforward. The optimal retention satisfies the first order condition

$$\left(1 - e^{-\rho t}\right)\left[(1-\theta_t) f(\kappa_c)(\Delta R_o)^2 - p(\theta_t) R_o\right] + e^{-\rho t} f(\kappa_c)(\Delta R - B - \kappa_c)\Delta R_o = 0.$$

Plugging in $\theta_t = 1$, we can show that the FOC satisfies

$$\left(1 - e^{-\rho t}\right)\left[-p(1) R_o\right] + e^{-\rho t} f(\kappa_c)(b - B)\Delta R_o < 0.$$

The solution, as well as the verification that the second-order condition is satisfied. □

# 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

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

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

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 $D \le \Pi$ is slack, $\Pi_i(\theta)$ 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 $\Phi \ge \underline{\Phi} := \sqrt{\frac{(\bar{\kappa}/\Delta^2)^2}{(\bar{\kappa}/\Delta^2)^2 + (R_o)^2}}$, 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

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

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 $\rho$ 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 3

*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 $\theta_0$ 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 $\Delta W \equiv W_c - W_i$,

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

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

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

We need to compare now $W_c$ and $W_d$.

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

□

## Proof of Corollary 1

*Proof.* For any $\theta_0$, it is easily verified that

$$\frac{\partial W_\ell}{\partial t_\ell} = 0 \Longrightarrow \frac{\partial^2 W_\ell}{\partial t_\ell^2} > 0,$$

which means that $W_\ell(t_\ell)$ is quasi-convex, so it is maximized on $\{0, \infty\}$. Moreover, it is easily verified that

$$\lim_{t_\ell\to\infty} W_\ell(t_\ell) - W_\ell(0) = (1-\theta_0^*)\,(1-\Phi)\,\frac{\left(\Delta R_o\right)^2\theta_0^*}{\bar{\kappa}} - \Phi\frac{\rho}{2\rho+\phi}\frac{\Delta R_o\theta_0^*}{\bar{\kappa}}\,(B-b) > 0,$$

so $\arg\max_{\{t_\ell\geq 0\}} W_\ell(t_\ell) = \infty$. □

## Details for Minimum Retention

Next, we construct the equilibrium in the case of certification with minimum retention. First, let's consider the price of the loan once the bank hits the minimum retention level $\underline{\theta}$. Because $\theta = \underline{\theta}$ is an absorbing state, the price of the loan satisfies $q_c(\underline{\theta}) = d_c(\underline{\theta}) = \pi_c'(\underline{\theta})$. For any $\theta > \underline{\theta}$ with smooth trading, the HJB is (9) the loan price satisfies equation (12) are unchanged. Therefore, in any region with smooth-trading, the bank's value function, the equilibrium trading strategy, and the price of loans is unchanged. It remains to check if at some $\theta$, the bank has incentives to trade atomistically to $\theta^+ = \underline{\theta}$. The payoff of jumping from $\theta$ to $\underline{\theta}$ is $\Phi\pi_c(\underline{\theta}) + q_c(\underline{\theta})(\theta - \underline{\theta})$ while the payoff in the smooth trading region is $\Pi_c(\theta) = \Phi\pi_c(\theta)$. If $\Phi\pi_c(\underline{\theta}) + q_c(\underline{\theta})(1 - \underline{\theta}) > \Phi\pi_c(1)$, the certifying bank always sells immediately to $\theta^+ = \underline{\theta}$. Otherwise, as in the proof of the case without minimum requirements, there is a unique $\tilde{\theta}$ satisfying

$$\Phi\pi_c(\tilde{\theta}) = \Phi\pi_c(\underline{\theta}) + q(\underline{\theta})(\tilde{\theta} - \underline{\theta}) = \Phi\pi_c(\underline{\theta}) + \pi_c'(\underline{\theta})(\tilde{\theta} - \underline{\theta}). \tag{55}$$

By no-arbitrage, the price function must be upper semi-continuous (that is $q(\tilde{\theta}) = q(\tilde{\theta}+)$), which means that $q(\tilde{\theta}) = \Phi\pi_c'(\tilde{\theta})$.[24] At the same time, no-arbitrage also requires to be equal to the expected dividend $\mathbb{E}\left[d_c(\theta_{\tau_\phi})|\theta_t = \tilde{\theta}\right]$. If the bank stops trading at $\tilde{\theta}$, and remains there for an exponential time with mean arrival rate $\lambda$, at which time sells $\tilde{\theta} - \underline{\theta}$, the expected dividend is

$$\mathbb{E}\left[d_c(\theta_{\tau_\phi})|\theta_t = \tilde{\theta}\right] = \frac{\lambda}{\phi + \lambda} d_c(\underline{\theta}) + \frac{\phi}{\phi + \lambda} d_c(\tilde{\theta}).$$

Combining the previous conditions, we get that $\lambda$ is implicitly given bu

$$\Phi\pi_c'(\tilde{\theta}) = \frac{\lambda}{\phi + \lambda}\pi_c'(\underline{\theta}) + \frac{\phi}{\phi + \lambda}\pi_c'(\tilde{\theta}) \tag{56}$$

The next proposition, summarize the previous discussion and describe the equilibrium in the certification case with minimum retention requirements.

**Proposition 5** ( Equilibrium with Minimum Retention)**.** *There is a unique* ***Certification Equilibrium with Minimum Retention*** *. Given the bank's initial retention $\theta_0$, the bank sells its loans smoothly at a rate given by equation (13) until $\tilde{T} = \min\{t > 0 : \theta_t = \tilde{\theta}\}$, were $\tilde{\theta}$ is the unique solution to equation (55). After time $\tilde{T}$, the bank holds $\theta_t = \tilde{\theta}$ until an exponentially distributed random time $\tau_\lambda$, at which time it sells off $\tilde{\theta} - \underline{\theta}$. The exponential time $\tau_\lambda$ has a mean arrival rate $\lambda \mathbf{1}_{\{\theta_t = \tilde{\theta}\}}$, where $\lambda$ satisfies (56). After time $\tau_\lambda$, the bank holds $\underline{\theta}$ until the projects maturity. The equilibrium loan price is*

$$q_c(\theta_t) = \begin{cases} \Phi\Big(p_L + F\left(\Delta R_o \theta_t\right)\Delta\Big) R_o & t < \tau_\lambda \\ \left(p_L + \frac{\phi}{\lambda+\phi} F\left(\Delta R_o \tilde{\theta}\right)\Delta\right) R_o & T_* \leq t < \tau_\lambda \\ p_L + F\left(\Delta R_o \underline{\theta}\right)\Delta\Big) R_o & t \geq \tau_\lambda \end{cases} \tag{57}$$

We omit a formal proof of this proposition as it follows the steps of the proof of Proposition 1.

[24] Otherwise, there would be a deterministic downward jump in the price at the time $\theta_t$ reaches $\tilde{\theta}$ which would be inconsistent with no-arbitrage.

## Proof of Proposition 4

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

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

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

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

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

Recall that in Lemma 3, the constraint is slack whenever $\theta$ is sufficiently high.

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

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

Clearly, when $\xi = 1$, we get the results in subsection 3.2 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 $\xi < \xi_\dagger$. □