China's Financial System in Equilibrium*

Jie Luo[a], Cheng Wang[b,*]

[a] *NIFS, School of Economics and Management, Tsinghua University, China*
[b] *School of Economics, Fudan University, China*

Abstract

We present a macro view of China's financial system, where a state-owned monopolistic banking sector coexists endogenously with markets for corporate bonds and private loans. The source and size distributions of external finance are determined jointly in the model's equilibrium where, as in the data, smaller firms obtain finance from private lending, larger firms through bank loans, and the largest by way of corporate bonds. The model predicts that removing controls on the bank lending rates or tightening the supply of credit reduces bank loans but increases bond finance. The model is calibrated to China's financial market data to shed light on the expansion of China's bond market against banking over the recent years. The model estimates that the private loans market has varied in size between zero to ten percent of the economy's total credit and had been growing after 2002.

Keywords: China, financial system, monitoring, banking reforms

JEL: G18, G21, O16

*First version: May 2016. We thank Steve Williamson, Jun Qian, Xuewen Liu, Zheng Song and the audiences at Fudan University, Tsinghua PBC School of Finance, University of Goettingen, University of Hong Kong, SAIF, Shanghai Jiaotong University, 2017 Asian Meeting of the Econometric Society, 2017 China Meeting of the Econometric Society, 2017 North American Summer Meeting of the Econometric Society, and the 2017 China Financial Research Conference for helpful discussions.

*Corresponding author: Cheng Wang. Email address: wangcheng@fudan.edu.cn.

October 30, 2019

1. Introduction

We develop an equilibrium model of the financial market to capture the key features of China's financial system, and to understand its experiences related to the reforms in the banking sector over the past twenty years. In the model, private information and banking regulations give rise to the coexistence of bank loans, corporate debt, and private lending. The banking sector is monopolistic and subject to central bank regulations on the lending and deposit rates that it charges. Firms differ in net worth. As in the data, in equilibrium the source distribution of external finance across firms is such that firms that are smaller in net worth obtain finance from private lending, larger firms through bank loans, and the largest by way of corporate debt. As a theory for the composition of lending, the model suggests that a contraction in the supply of external finance cuts bank loans but increases the shares of corporate bonds and private lending. The model also implies that the reform to remove the controls over the bank lending rate should have reduced bank loans but increased bond finance. Calibrated to China's financial market data, the model accounts quantitatively for the observed decline in banking against the expansion of the corporate bond market, which was observed over the recent years. It also suggests that the private loans market has varied in size between zero to ten percent of the economy's total credit and had been growing after 2002.

China's financial system consists of a state-owned, tightly regulated, and (arguably) monopolistic banking sector, a less formal and decentralized direct lending market, an equity market, and a growing bond market. While there is no official data on the size of informal lending, Figure 1 shows how large and important each of the other three parts of China's financial system is, relative to total financing (i.e., excluding informal lending). Specifically, it depicts in time series the division between bank loans and the two other types of finance as a fraction of total lending (again excluding informal lending) over the period 2002-2015.[1] Notice the small equity market over the period. Notice more importantly the relative decline in banking and the rise of the market for bonds over the same period.[2]

[1]About two thirds of shadow banking in China result from regulatory arbitrages of banks (see Elliott, Kroeber and Qiao, 2015).

[2]A decline of banking has also been observed in other emerging economies after 2010, but to a much less

The largest banks in China are state owned and dominate the economy's banking sector. Figure A.1 in the appendix measures the degree of bank concentration in China, showing the time series of bank loans held by the largest five banks as a fraction of total bank loans in China, relative to the U.S.. Observe that bank concentration in China has been falling but remains high, much higher than that in the U.S.. Banks in China are subject to state regulations, but the past twenty years has seen reforms that gave commercial banks greater flexibility in decision making, especially with respect to the interest rates on loans and deposits. Before 2004, interest rates in the banking sector were tightly regulated by the People's Bank of China (PBC), by way of setting the policy interest rates (on bank loans and deposits) and interest rate ceilings and floors around the policy rates. The lending rate ceilings were removed in October 2004, and then the floors in July 2013, and by 2015 the PBC had also released its control over the deposit rates. Figure A.2 in the appendix depicts the time series of the policy rates on one year loans and one year saving deposits. Notice the greater variability in both rates after 2004.

The private lending market in China consists of non-delegated monitors, such as relatives, money lenders, and other less delegated monitors such as peer-to-peer platforms. This market is quite large according to some studies. Ayyagari et al. (2010) claim that private lending is at least one-quarter of all financial transactions, with an estimated size of CNY 740–830 billion at the end of 2003, equal to about 4.6% of total outstanding bank loans in 2003. Lu et al. (2015) estimate that in 2012, private lending totals $4,000$ in billions of RMB, about 6.4% of total outstanding bank loans in 2012. In the private lending market, there is much larger variability in the nominal interest rates charged, ranging from nearly zero on loans from relatives to more than 30% from money lenders. He et al. (2015) document that interest rates in the private credit markets are opaque and on average 2 to 3 times more than the bank lending rates.

A hallmark of China's financial system is the uneven distribution of bank loans between smaller and larger firms. There is wide documentation of the difficulties small firms face in

degree (see Figure A.5 in the appendix). Chang, Fernńdez and Gulan (2017) argue that for most emerging economies, the decline resulted from the drop in global interest rates which obviously could not explain China. Interest rates in China were stable over the period, as shown in Figure A.2 in the appendix.

obtaining bank loans. The Word Bank's enterprise surveys for China 2012 shows that the fraction of firms using bank loans for investment financing is on average much lower in China than in other countries. Specifically, for the small firms in their survey, it is 3.8% in China, 16.8% in East Asia and Pacific, and 21.5% across all countries.[3] Firms in the World Bank's surveys are on average much smaller in size – measured in total employment – than publicly listed firms (Table A.2 in the appendix). For a more comprehensive picture, we merge the listed firms with the firm in the World Bank's surveys, rank and divide them into 10 groups by size. We find an inverted-U relation between firm size and the fraction of firms using bank loans as their only source of external finance (Figure A.3 in the appendix).

One might suggest that bank loans are for some reason too expensive for smaller firms. This is not the case. As Table A.1 in the appendix shows, in the World Bank's Surveys data, among firms who need a loan but choose not to apply for one, for small firms the most important reason is that the application procedures were complex; while for larger firms, it is the unfavorable interest rates. The table also indicates that, relative to larger firms, a larger fraction of small firms would like to obtain a loan at the ongoing interest rate but could not. In addition, the fraction of firms who did not apply for a loan because they did not think it would be approved is much larger among smaller relative to larger firms.

China's bond market, where the majority of contracts traded are government and corporate bonds, has grown substantially over the last twenty years, from virtually nonexistent to the third biggest in the world, just behind the U.S. and Japan. From Figure A.4 in the appendix, although corporate bonds currently still account for a smaller part of the whole market, they have grown fast in relative size. Another feature of China's bond market, depicted in Figure 2, is that the firms who use bonds as a means of finance are much larger than those using bank loans who, in turn, are larger than those who use neither bonds nor bank loans.[4]

Major holders of China's corporate debt include (a) trust corporations, funds, and other non-financial institutions; (b) national commercial banks; (c) individuals and un-incorporated entities; and (d) other holders including for example online money market funds. Over about

[3]See https://www.enterprisesurveys.org/en/data/exploreeconomies/2012/china#finance.

[4]As is well known from the literature, that firms who use bonds for external finance are larger than those using bank loans is not just observed among Chinese firms.

the same period when the above mentioned banking reforms were taking place, several reforms that target the corporate bonds market also occurred. This includes the 2005 reform which gave businesses already in the bond market access to short-term commercial paper, and the 2007 reform which gave listed firms rights to issue bonds, and the the 2015 reform which made the bond market open to essentially all incorporations (Lin and Milhaupt, 2017).

To develop a theory for China's financial system, we take a standard approach in modeling lending and financial intermediation (banking). Specifically, following Diamond (1984) and Williamson (1986), lending is subject to costly state verification (CSV) and the bank is a delegated monitor. Firms (borrowers) differ in net worth, which is used as equity, as well as collateral for mitigating the effects of CSV and limited liability (Bernanke and Gertler, 1989). As delegated monitor, the bank is more efficient in lending than individual investors. Private lending coexists with bank lending because the low (regulated) deposit rate induces investors to participate in private lending for higher returns; or because a tight supply of credit dictates a sufficiently high interest rate on private lending to compete credit away from banking. On the other hand, that in equilibrium the bank lends to firms with larger net worth is because, relative to the bank, individual lenders have a comparative advantage in financing smaller than larger projects. Larger firms, with a larger net worth to support more investment, make the bank a more efficient delegated monitor. Meanwhile, financing a smaller project in non-delegated lending requires a fewer times of repetition in monitoring the firm's financial report.[5] Larger firms also find bonds a favorable means of finance, as the large net worth allows them to raise enough finance without resorting to costly monitoring.

The model provides a vehicle for evaluating the effects of the recent banking reforms that took place in China. In particular, the model suggests that removing the controls over the bank lending rate, which occurred in 2004, should have resulted in a decline in banking, while at the same time increasing bond finance but reducing private lending. The model also suggests that removing all interest rate controls would result in a higher interest rate, crowding out private lending. Calibrated to China's financial market data, the model suggests that a combination of several factors – related respectively to changes in the economy's technological

[5]An empirical literature relates bank loans with state ownership (e.g., Allen and Qian, 2014). Our model suggests that standard theories of banking are sufficient for explaining why banks prefer larger firms.

environment, the reforms on banking, and the government's policy choices – played essential roles in explaining the observed bond market expansion in China. In particular, all else equal, without the 2004 lending rate reform the model would only be able to explain about 55% of the observed expansion of the bond market against bank loans. The calibration also reveals that between 2002 and 2015 private lending in China had been increasing and its size varied in the range of 0 to 10 percent of total external finance.

This study is based on the works in the theory of costly state verification and its applications in financial contracting and intermediation, including, among others, Townsend (1979), Gale and Hellwig (1985), Boyd and Prescott (1986), Williamson (1986), Greenwood and Jovanovic (1990), Holmström and Tirole (1997) and Greenwood, Sanchez and Wang (2010). In modeling monitoring, we offer a novel specification which divides the total cost of monitoring between a fixed component that depends on the size of the investment, and a variable part that depends on the measure of lenders providing external finance.

Our work extends the existing studies of China's financial markets, much of which focuses on the roles of informal lending and shadow banking. Allen, Qian and Qian (2005) suggest that informal financial mechanisms played an important role in supporting the growth of China's private sector economy. In Wang et al. (2015), commercial banks use shadow banking to evade policy restrictions on deposit rates and loan quantities. Chen, Ren and Zha (2016) argue that the rising shadow banking in China results from small banks' incentives to fund risky industries while avoiding the loan-to-deposit ratio set by the regulator. Hachem and Song (2018) study an important component of shadow banking in China – the wealth management products (WMPs) of commercial banks – in a model of interbank competition. They show that small and medium-sized banks who are more constrained by the loan-to-deposit cap were more heavily involved in issuing off-balance-sheet WMPs.

Section 2 sets up the model and section 3 studies the optimal lending contracts. Section 4 defines and studies the model's equilibrium. Section 5 studies analytically the effects of the recent banking reforms. Section 6 calibrates the model to China's financial market data and uses the model to study, quantitatively, the composition of China's financial system, and the effects the banking reforms. Section 7 concludes the paper. Additional data and material, including the proofs of the theoretical results, are in the appendix.

2. Model

There are two time periods: $t = 0, 1$. In period 0 a financial market opens where lending and borrowing take place, and in period 1 production and consumption take place. There is a single good in the model that can be used as capital or consumption. There is a continuum of agents, among them M units are investors and μ units firms. Firms maximize expected profits in period 1. Investors maximize expected consumption in period 1. Profits and consumption must be non-negative. Each investor is endowed with 1 unit of the good in period 0. Firms differ in capital endowment, k, which is uniformly distributed over the interval $[0, \bar{k}]$ across individual firms, with $\bar{k} > 0$. Each firm is also endowed with an investment project with which any $X(\geq 0)$ units of capital invested in period 0 returns $\tilde{\theta}X$ units of output in period 1, where $\tilde{\theta}$ is a random variable that takes value θ_1 with probability π_1, and θ_2 with probability π_2, with $\theta_2 > \theta_1 > 0$ and $\pi_1 = 1 - \pi_2 \in (0, 1)$.

A bank in the model takes deposits from investors and offers loans to firms. This bank is "state owned" and subject to regulations. Let R_D denote the gross rate of return on deposits and R_L the gross interest rate charged on loans. The values of R_D and R_L are fixed by the state and are such that $0 < R_D < R_L$. Naturally, assume $R_D \in (\theta_1, E(\theta))$ and $R_L \in (R_D, \theta_2)$.[6]

Each investor could lend indirectly through the bank, at the fixed interest rate R_D, or directly to individual firms through a private lending market. Likewise, each firm can either borrow from the bank, or directly from individual investors in the private lending market. For convenience, assume firms cannot obtain finance simultaneously from both the bank and a set of individual investors, and investors cannot participate in both markets either.

The realization of $\tilde{\theta}$ is observed by the firm who runs the project. The same information can be revealed to any other party only if the firm incurs a cost to let that party monitor his report. This cost of monitoring equals

$$C(\Delta, X) = \gamma_0 X + \gamma \Delta X, \tag{1}$$

where X is the size of the project, Δ the measure of lenders who provide the external finance, and γ_0 and γ are positive constants. Assume $\gamma_0 < \theta_1$. Notice that equation (1) covers both

[6]Suppose $R_D \leq \theta_1$. Then as it will become clear as the analysis unfolds, the model would not have an equilibrium where bank loans are an active means of finance.

the case of delegated monitoring, with $\Delta = 0$, and that of non-delegated monitoring, with $\Delta > 0$. Observe that $C(\cdot, \cdot)$ is consistent with the very original idea of Diamond (1984) that delegation allows lenders to avoid the cost of repetition in monitoring, which is increasing in the degree of the repetition which, in turn, increases as the measure of lenders increases. Observe also that equation (1) implies that the bank as a lender is always more efficient than individual investors, as long as some monitoring is involved in the lending.

Monitoring, delegated or not, is deterministic: any lending contract can only specify to monitor the borrower in any given state of the world with probability one or zero. And lastly, lending is subject to a limited liability constraint: in no state of the world the firm be required to make a loan repayment that exceeds the output it produces.

3. Optimal Lending

Let r^* denote the market rate of (net expected) return on lending for individual investors. This is an endogenous variable whose value is determined in the equilibrium of the model, with $r^* \in [R_D, E(\theta))$. Obviously, if direct lending and bank loans are both active in equilibrium, then $r^* = R_D$. If direct lending is active but bank lending is not, then it must be that $r^* > R_D$. If there is no direct lending but there is active bank lending, then again $r^* = R_D$. All investors are lenders. Firms are free to participate in either side of the market. However, given $r^* < E(\theta)$, it is never optimal for any firm to lend any fraction of its net worth to the market, directly or indirectly. In the following analysis, therefore, we take as given that each firm is a borrower.

3.1. Direct Lending

Consider first the market where individual investors lend directly to firms, not through the bank. Consider an individual firm in this market, with net worth k. To obtain finance, it offers a contract to potential lenders, and the contract reads $\sigma_D(k) = \{L(k), S(k), r_1(k), r_2(k)\}$, where $L(k)$ is the size of the loan or $X(k) \equiv L(k) + k$ is the size of the project; $r_i(k)$ is the repayment per unit of the loan in output state θ_i, $i = 1, 2$; and $S(k)$ is the set of reported output states in which the lender monitors the borrower's report. It is straightforward to show that the optimal contract has either $S(k) = \emptyset$ or $S(k) = \{\theta_1\}$.[7]

3.1.1. Non-monitored Direct Lending

Suppose $S(k) = \emptyset$. To induce truth telling the firm's loan repayment must be constant across the states of output, or $r_1(k) = r_2(k) = r_{\mathrm{N}}(k)$, and the firm's value thus equals

$$V_{\mathrm{N}}(k) \equiv \max_{r_{\mathrm{N}};L\geq 0} \Big\{ \pi_1\theta_1(L+k) + \pi_2\theta_2(L+k) - r_{\mathrm{N}}L \Big\} \ s.t. \tag{2}$$

$$r_{\mathrm{N}}L \leq \theta_1(L+k), \tag{3}$$

$$r_{\mathrm{N}} \geq r^*, \tag{4}$$

where (3) is limited liability: repayment of the loan cannot exceed output, and (4) is individual rationality: the lender must get a rate of return that is not below the market rate.

Lemma 1. *Conditional on $S(k) = \emptyset$, for all $k \in [0, \overline{k}]$ the optimal contract has $r_{\mathrm{N}} = r^*$ and*

$$L_{\mathrm{N}}(k) = \frac{\theta_1 k}{r^* - \theta_1}, \quad X_{\mathrm{N}}(k) = \frac{r^* k}{r^* - \theta_1}. \tag{5}$$

So with no monitoring, the optimal way to raise finance is to issue a risk-free bond that pays the market interest rate r^*. At the optimum, constraint (3) binds. That is, in the low output state, loan repayment is just equal to the output produced, and this allows the firm to raise the maximum amount of finance that constraint (3) permits. With these, the firm's value is $V_{\mathrm{N}}(k) = \pi_2(\theta_2 - \theta_1)\frac{r^* k}{r^* - \theta_1}$. Notice that $L_{\mathrm{N}}(k)$, $X_{\mathrm{N}}(k)$ and $V_{\mathrm{N}}(k)$ are all linear and increasing in k. In other words, a larger firm net worth supports more finance, a larger project, and higher firm value.[8]

3.1.2. Monitored Direct Lending

Suppose lending involves monitoring: $S(k) = \{\theta_1\}$. Then optimal contracting solves

$$V_{\mathrm{M}}(k) \equiv \max_{\{r_1, r_2, L\geq 0\}} \Big\{ \pi_1 \Big[\theta_1(L+k) - r_1 L - \widetilde{C}(L,k) \Big] + \pi_2 \left[\theta_2(L+k) - r_2 L \right] \Big\} \ s.t.$$

$$0 \leq r_1 L \leq \theta_1(L+k) - \widetilde{C}(L,k), \tag{6}$$

[7] See Lemma 6 in the appendix for proof.

[8] Note that Lemma 1 is derived under the assumption of $R_D > \theta_1$ which implies $r^* > \theta_1$. Suppose $R_D \leq \theta_1$ and $r^* \leq \theta_1$. Then the optimal $L_{\mathrm{N}}(k)$ would be infinity for all $k \in [0, \bar{k}]$, which, given that M is finite, cannot be part of an equilibrium of the model.

$$0 \le r_2 L \le \theta_2(L+k), \tag{7}$$

$$\theta_1(L+k) - r_1 L - \widetilde{C}(L,k) \ge \theta_1(L+k) - r_2 L, \tag{8}$$

$$\pi_1 r_1 + \pi_2 r_2 \ge r^*, \tag{9}$$

where

$$\widetilde{C}(L,k) = \begin{cases} C(L, L+k) = \gamma_0(L+k) + \gamma L(L+k), & \text{if } L > 0 \\ 0, & \text{if } L = 0 \end{cases}. \tag{10}$$

In the above, (6) and (7) are non-negativity and limited liability, and (8) is incentive compatibility. Given $S(k) = \{\theta_1\}$, the contract must only ensure that the firm has no incentives to report θ_2 when output is θ_1. Equation (9) is a participation constraint and (10) says that the cost of monitoring is $C(L, L+k)$ if lending takes place, zero if not.

Monitoring affects the firm's value in two ways. First, monitoring enters the firm's objective function to reduce its value directly, and this effect is larger when k is larger. Second, monitoring enters the incentive constraint to affect the firm's value indirectly. To understand the latter, remember that with no monitoring, truth-telling imposes $r_1 = r_2$. With monitoring, truth-telling requires instead

$$r_2 - r_1 \ge \widetilde{C}(L,k)/L \ge 0, \tag{11}$$

or only a gap between r_1 and r_2. The size of this gap, however, is increasing in the cost of monitoring $\widetilde{C}(L,k)$ which, in turn, is increasing in k for any fixed L. With a smaller k (smaller $\widetilde{C}(L,k)$), a less tight incentive constraint (11) increases potentially the size of lending and thus the value of the firm. On the other hand, lending is more tightly constrained with a larger k. In particular, when k is sufficiently large to make $\widetilde{C}(L,k))$ sufficiently large, (11) is likely to bind, or even infeasible for the contract to implement (remember r_1 must be non-negative and r_2 must not exceed θ_2). This reduces the firm's value. To summarize, in monitored direct lending, monitoring goes better with a smaller rather than a larger k. Lastly, since each lender imposes on the firm a monitoring cost of $\gamma(L+k)$ to verify the report of θ_1, the repetition in monitoring implies that the total cost of monitoring incurred increases more than linearly in the size of the project, and this strengthens the effects discussed above.

3.1.3. Optimal Direct Lending

The firm's optimal finance is now determined, under

Assumption 1. *(i)* $r^* < E(\theta) - \pi_1\gamma_0 \equiv R_{\max}$. *(ii)* $R_D > \pi_2\theta_2 - \pi_1\theta_1 + \pi_1\gamma_0 \equiv R_{\min}$.

Part (i) ensures that the mean output of the project is sufficiently high so that once it is financed, on average the firm has enough to cover the reservation return of the lender plus the fixed cost in monitoring which is assumed to occur in the state of low output. Part (ii) assumes that the deposit rate is sufficiently high so that the non-negativity constraint $r_1 \geq 0$ in (6) dos not bind.[9]

Proposition 2. *(i) There is a cut-off level of k, $\tilde{k} \in (0, \bar{k})$, below which the optimal direct finance for firm k involves monitoring and above which the risk-free bond (described in Lemma 1) is optimal. (ii) For any $k \in [0, \tilde{k})$, the optimal contract, which prescribes $S(k) = \{\theta_1\}$, has:*

$$L_{\mathrm{M}}(k) = \frac{E(\theta) - \pi_1\gamma k - \pi_1\gamma_0 - r^*}{2\pi_1\gamma}, \tag{12}$$

$$X_{\mathrm{M}}(k) = \frac{E(\theta) + \pi_1\gamma k - \pi_1\gamma_0 - r^*}{2\pi_1\gamma}, \tag{13}$$

$$r_1(k) = \frac{(\theta_1 - \gamma_0)X_{\mathrm{M}}(k) - L_{\mathrm{M}}(k)\gamma X_{\mathrm{M}}(k)}{L_{\mathrm{M}}(k)}, \tag{14}$$

$$r_2(k) = \frac{r^* - \pi_1 r_1(k)}{\pi_2}, \tag{15}$$

$$V_{\mathrm{M}}(k) = \frac{[E(\theta) + \pi_1\gamma k - \pi_1\gamma_0 - r^*]^2}{4\pi_1\gamma} + kr^*, \tag{16}$$

where $\tilde{k}$ solves, uniquely, $V_{\mathrm{M}}(\tilde{k}) = V_{\mathrm{N}}(\tilde{k})$, as Figure 3 illustrates.

So the optimal contract has:

$$X(k) = X_{\mathrm{M}}(k),\ \forall k < \tilde{k},\ \textit{and}\ X(k) = X_{\mathrm{N}}(k), \forall k \geq \tilde{k}; \tag{17}$$

with

$$V(k) = V_{\mathrm{M}}(k), \forall k < \tilde{k},\ \textit{and}\ V(k) = V_{\mathrm{N}}(k), \forall k \geq \tilde{k}. \tag{18}$$

[9] Suppose (ii) is violated. Then the constraint $0 \leq r_1 L$ binds for all $k < \tilde{k}'$, where $\tilde{k}' = (\pi_2\theta_2 - \pi_1\theta_1 + \pi_1\gamma_0 - r^*)/(\pi_1\gamma)$. It then follows that $r_1(k) = 0$ and $X(k) = k + (\theta_1 - \gamma_0)/\gamma$, for all $k \in [0, \tilde{k}']$. This changes the analysis slightly but would not change the qualitative conclusions derived. Note also that the assumption $R_D > R_{\min}$ will not be maintained later when the model is calibrated.

That a larger k makes finance with monitoring less efficient relative to that with no monitoring was anticipated from earlier discussions.[10] Proposition 2 also suggests that conditional on monitoring, as conditional on no monitoring (Lemma 1), at the optimum a larger k supports a larger X and larger firm value. Remember, conditional on no monitoring, a larger k, used as collateral, increases the firm's ability to repay its debt in the state of low output. This same effect exists also in monitored private lending.[11]

Notice that the optimal size of the loan is increasing in k in the case of no monitoring, but decreasing in k in the case of monitoring. Conditional on no monitoring, a larger k supports larger repayments in the sate of low output and hence a larger loan. In monitored direct lending, however, a larger k also increases the cost of monitoring and at an increasing rate, resulting in a smaller loan.[12] In general, monitoring allows the contract to support more external finance and hence a larger project. In Appendix A.8, we show that the optimal size of the project conditional on monitoring, $X_{\mathrm{M}}(k)$, strictly exceeds that conditional on no monitoring, $X_{\mathrm{N}}(k)$, for all $k \in [0, \tilde{k}]$. Also, observe from Figure A.14, again in the appendix, the jump in the optimal size of the project as a function of k, $X(k)$, at $\tilde{k}$ which divides monitoring and no monitoring.[13]

3.2. Intermediated/Bank Finance

Let $D(\geq 0)$ denote the bank's total deposits raised, or the total supply of bank loans, an endogenous variable of the model whose value would depend on r^*, the market interest rate for all lenders. We focus on the case of $r^* = R_D$, for otherwise (i.e., $r^* > R_D$) no one lends through the bank and $D = 0$.

[10] In addition, note that conditional on monitoring, a larger k increases the cost of monitoring per unit of external finance, which is given by $C(L, X)/L = (\gamma_0 X + \gamma L X)/L = \gamma_0 X/L + \gamma X$, where L, the optimal amount of external finance raised, is decreasing in k.

[11] But not between monitoring and no monitoring, as shown in Appendix A.8 and Figure A.14.

[12] Because of the convex monitoring cost, the marginal net return on lending, which is $E(\theta) - r^* - \pi_1(\gamma_0 + \gamma k + 2\gamma L)$, is decreasing in both k and L.

[13] From Lemma 1, if the optimal contract prescribes no monitoring (i.e., $k \geq \tilde{k}$), then the interest rate must be constant and equal to r^* across the output states. In Appendix A.7, we show that with monitored direct lending, that is for all $k \in [0, \tilde{k})$, $r_1(k) < r^* < r_2(k)$ and $r_1'(k) > 0$, $r_2'(k) < 0$. That is, if the optimal contract prescribes monitoring, then there is an interest rate spread between the two output states, and the spread is smaller if the firm is larger in net worth. In addition, as is straightforward to show, for any fixed $k \in [0, \tilde{k}]$, $r_1(k)$ is larger if r^* is larger. This holds because a larger r^* reduces the optimal size of the investment and increases the efficiency in monitoring which, in turn, permits higher lender returns in the low output state.

As mentioned earlier, the bank lends out its funds through a standard loan contract which prescribes a fixed (gross) interest rate $R_L \in (R_D, \theta_2)$. The contract also prescribes that if the firm fails to make the required repayment, which would occur in the state of θ_1 given $\theta_1 < R_L$, it must submit all of its output to the bank. Given R_L, as part of the lending contract the bank then chooses the size of the loan $L(k)$, or equivalently the size of the firm's project $Z(k)(\equiv L(k)+k)$, and a policy for monitoring the firm's output. Now let $\mathbf{B}$, a subset of $[0, \bar{k}]$, denote the set of all firms whom the bank is willing to offer a loan to. For each $k \in \mathbf{B}$, the loan must ensure that the firm gets a value no less than $V(k)$ – the value the direct lending market could guarantee and thus the bank must take as the firm's reservation value.

It is straightforward to show that, with the optimal contract, monitoring occurs if and only if the lower output θ_1 is reported.[14] Given this, the bank's problem can be written as

$$\max_{\mathbf{B},\{L(k)\}_{k\in\mathbf{B}}} \mu \int_{\mathbf{B}} \Big[\pi_1 (\theta_1 - \gamma_0)(k + L(k)) + (\pi_2 R_L - 1)L(k)\Big] dG(k) + (1 - R_D)D \;\; s.t.$$

$$\mathbf{B} \subseteq [0, \bar{k}], \tag{19}$$

$$L(k) \geq 0, \quad \forall k \in \mathbf{B}, \tag{20}$$

$$\mu \int_{\mathbf{B}} L(k) dG(k) \leq D, \tag{21}$$

$$V_b(k, L(k)) \equiv \pi_2 \{\theta_2(k + L(k)) - R_L L(k)\} \geq V(k), \quad \forall k \in \mathbf{B}, \tag{22}$$

where equation (21) is a resource constraint: total loans made cannot exceed the total supply of bank credit; and (22) is a participation constraint: firms in $\mathbf{B}$ are better off obtaining finance from the bank than from individual lenders directly.[15]

[14] To see this, first it is straightforward to show that monitoring a report of θ_2 is never optimal. Next, monitoring must occur in some state of output. Suppose monitoring never occurs with the optimal contract. Then it must hold that $R_L L(k) \leq \theta_1(k + L(k))$, so the firm is able to repay the loan in the low output state. This in turn requires $L(k) \leq \theta_1 k/(R_L - \theta_1)$, where the right hand side gives the maximum size of the credit the firm could raise with the bank. Given this and $R_L > R_D$, the expected value of the firm, which equals $E(\theta)(k + L(k)) - R_L L(k)$, is strictly less than $V(k)$, as is easy to verify. In other words, if the bank never monitors the firm's report, it would not be able to induce the firm to participate – it could not offer a loan that is sufficiently large to make the firm better off than with direct lending.

[15] Note that although the bank is allowed to participate in the market for corporate bonds, but given that all parties are risk neutral and that the bank is given no cost advantage in bond purchasing over individual investors, the bank has no incentives to participate in the bond market. In practice however, banks do hold corporate bonds, but they do so for reasons that are not given by the model. In practice, the bank, in addition to being the delegated monitor as the model depicts, also plays the role as an intermediary for individual investors who participate in the market for corporate bonds. From the model's perspective, corporate bonds

Now rewrite (22) as

$$L(k) \geq L_0(k),\ \forall k \in \mathbf{B}, \tag{23}$$

where

$$L_0(k) \equiv \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\theta_2 - R_L)},\ \forall k \in [0, \bar{k}]. \tag{24}$$

In addition, let

$$Z_0(k) \equiv k + L_0(k) = \frac{V(k) - \pi_2 R_L k}{\pi_2(\theta_2 - R_L)},\ \forall k \in [0, \bar{k}], \tag{25}$$

where $L_0(k)$ is the firm's reservation loan size – the minimum size of the loan with which it is willing to participate in bank lending, and $Z_0(k)$ is the corresponding reservation size of the project. Given that the firm gets returns only in the state of high output, a larger loan always gives the firm a larger value, and only a sufficiently large loan (larger than $L_0(k)$) induces the firm to participate.

From (24), a larger k affects $L_0(k)$ in two ways. First, all else equal a larger k allows the firm to keep a larger share of the output θ_2 after repaying the bank, reducing $L_0(k)$. Second, a larger k increases the firm's outside value $V(k)$, requiring a lager loan for participation. Overall, however, it can be shown that $L_0(k)$ and $Z_0(k)$ are both increasing in k.[16] In addition, notice $V_b(k, L_0(k)) = V(k)$. That is, at $L_0(k)$, the firm is indifferent between raising finance from the bank and borrowing directly from individual lenders.

We now move on, letting

$$\bar{D} \equiv \mu \int_0^{\bar{k}} L_0(k) dG(k)\ and\ D_0 \equiv \mu \int_{\tilde{k}}^{\bar{k}} L_0(k) dG(k). \tag{26}$$

In words, $\bar{D}$ is the minimum total amount of loans the bank would make if it wishes to lend to all firms, and D_0 is the minimum total amount of loans made if it wishes to lend only to firms with $k \in [\tilde{k}, \bar{k}]$ – those who would issue bonds for finance if a bank loan were not available.

To characterize the bank's optimal policy, we assume its rate of return on lending to a firm is greater than what the storage technology guarantees and so the bank would lend out all of its deposits. More specifically,

---

held by banks in practice are essentially corporate bonds held by individual investors. They are no part of the bank's business in the model, which the above optimization describes.

[16] See Lemma 10 in the appendix for a proof.

Assumption 2. $\pi_2 R_L + \pi_1(\theta_1 - \gamma_0) > 1$.

Proposition 3. *The following holds under Assumption 2. (i) Suppose* $0 \le D < D_0$*. Then the bank's optimal plan has* $L_{\mathrm{B}}(k) = L_0(k)$, $\forall k \in \mathbf{B}$, *where* $\mathbf{B}$ *is any subset of* $[\tilde{k}, \bar{k}]$ *that solves*

$$\mu \int_{\mathbf{B}} L_0(k) dG(k) = D. \tag{27}$$

(ii) Suppose $D_0 \le D < \bar{D}$*. Then it is optimal for the bank to set* $\mathbf{B} = [\hat{k}, \bar{k}]$*, with* $L_{\mathrm{B}}(k) = L_0(k)$, $\forall k \in [\hat{k}, \bar{k}]$, *where* $\hat{k}$ *solves*

$$\mu \int_{\hat{k}}^{\bar{k}} L_0(k) dG(k) = D.$$

(iii) Suppose $D \ge \bar{D}$*. Then it is optimal for the bank to set* $\mathbf{B} = [0, \bar{k}]$*, and with* $\{L_{\mathrm{B}}(k), k \in \mathbf{B}\}$ *be any function that satisfies (20) and (21).*

To understand these results, consider the bank's rate of return on lending to a firm k with a loan of size L, with $L \ge L_0(k)$:

$$\begin{aligned} R_b(k, L) &\equiv \frac{\pi_1(\theta_1 - \gamma_0)(k + L) + \pi_2 R_L L}{L} - R_D \\ &= \pi_1(\theta_1 - \gamma_0)\frac{k}{L} + \pi_1(\theta_1 - \gamma_0) + \pi_2 R_L - R_D. \end{aligned} \tag{28}$$

Observe that the term $\pi_1(\theta_1 - \gamma_0)k/L$, which measures the returns from seizing the firm's output on its own capital k, is decreasing in L for fixed k, but increasing in k for fixed L. A larger k allows the bank to get a larger repayment in the state of low output, increasing its returns per unit of lending. A larger L, on the other hand, dilutes the gains from utilizing the firm's net worth as collateral (for enforcing repayments in the state of low output), reducing the bank's returns per unit of lending.

In (i) and (ii) where $D < \bar{D}$, the bank could not offer a loan to all firms, any capital above $L_0(k)$ could then be reallocated to a firm not yet receiving bank credit, and this gives extra returns to the bank.[17] In these cases, what the bank seeks, essentially, is to maximize the number of loans made, by making each loan as small as possible.[18] Under $D < \bar{D}$, equation

[17] For this, see the appendix (Step 3 in Appendix A.9) for related calculations.

[18] A key assumption that drives this result is that R_L is fixed. With a fixed loan rate, the bank's returns per unit of lending in the state of high output θ_2 is constant. This forces it to seek higher return rates on lending by focusing on what it could get from the low, not high, output state. Suppose R_L is set free – the case that will be analyzed later in the paper when banking reforms are discussed. Then the bank could shift to how to

(28) also indicates the bank should in general prefer larger to smaller firms. More specifically, given (24) and Corollary 7 in the appendix,

$$\frac{dR_b(k, L_0(k))}{dk} \begin{cases} > 0, & \text{for } k \in [0, \tilde{k}] \\ = 0, & \text{for } k \in [\tilde{k}, \bar{k}] \end{cases}.$$

That is, between firms with $k \in [0, \tilde{k}]$, the bank strictly prefers the larger; and between those with $k \in [\tilde{k}, \bar{k}]$, it is indifferent.

In case (i) with $0 \leq D < D_0$, the supply of bank credit is so tight that only a subset of the firms with $k \geq \tilde{k}$ are offered a loan. Remember these are the firms whose large net worth allows them to raise finance directly from the bond market at the market interest rate r^*. These firms, despite their differences in k, are equally attractive to the bank, as they all promise the same expected rate of return on a loan. To resolve the indeterminacy, and given the observation that firms who get finance from the bond market are on average larger than those from banks, we take the stand that $\mathbf{B} = [\tilde{k}, \hat{k}_2]$, where $\tilde{k} < \hat{k}_2 \leq \bar{k}$ (Figure 4).[19]

In case (ii) with $D_0 < D < \bar{D}$, the bank has more funds for firms with $k \geq \tilde{k}$ but not enough for all firms. What it does, optimally, is to lend to the larger firms (above a cutoff in k), by giving each of them a loan with their reservation size $L_0(k)$.

Lastly, in the case of $D \geq \bar{D}$, the bank has more than enough funds to lend to all firms to meet their minimum demand for bank lending. The proposition says that it is optimal for the bank to (a) meet the minimum demand for credit from each firm, and then (b) lend the remaining funds to an arbitrary set of firms, on top of their $L_0(k)$.[20]

get more in the high state of output, making simultaneously R_L higher and L larger.

[19] One way to justify this is the following. Suppose in order to raise finance in the bond market any firm must incur a fixed participation cost. Then no matter how small this cost is, the rate of return on a bank loan is decreasing in the borrower's net worth (i.e., k), and the desired result follows. Essentially, a fixed participation cost increases the per-unit-of-capital value of a larger firm (i.e., $V(k)/k$)) if it were to issue bonds instead of obtaining a bank loan. In turn, this increases the bank's cost per-unit-of-lending to a larger firm, which in turn makes smaller firms more desirable to the bank. A formal treatment of this idea is given in Appendix A.6. In addition, note that the *rationing* assumed here does not imply that those obtaining bank loans are better off than those who do not. In fact, the firms are indifferent between bank loans and bonds. The difference is: for any given k, bank finance, with the use of monitoring, is larger in size than bond finance (see discussion in the subsection to follow).

[20] Here (b) is optimal because, conditional on each individual firm getting its minimum external finance $L_0(k)$, the rate of return to the bank on any extra lending is constant (at $\pi_1(\theta_1 - \gamma_0) + \pi_2 R_L$), in k and in the amount of the extra lending.

Obviously, $\hat{k}_1(D)$ is decreasing and $\hat{k}_2(D)$ is increasing in D, as Figure 4 illustrates. To conclude this section then, we claim that as D increases, the use of bank loans relative to total finance increases monotonically, while the use of bond finance and monitored private lending decrease monotonically as a fraction of total external finance.

3.3. Direct vs. Bank Lending

Being more efficient in monitoring, what outcome, in particular in the size of the external finance it supports, would the bank achieve relative to direct lending? The answer is, if R_D is sufficiently low, bank lending always supports a larger investment relative to direct lending; If R_D is sufficiently high, however, direct lending would support a larger investment for firms with a sufficiently small net worth.[21]

The explanation for these results would touch the essence of the difference between the two lending mechanisms. On the one hand, while R_L is fixed for bank loans, interest rates on direct lending are freely adjustable to reflect market conditions, giving direct lending an upper hand. On the other hand, being more efficient in monitoring gives bank loans an advantage over direct lending.[22] And this advantage is greater when the size of the investment is larger, and the size of the investment is larger if k is larger, for a larger k implies not only larger internal finance, but also greater ability for the firm to borrow externally (the optimal $L(k)$ increases in k). In the model, for k sufficiently small and so the cost of duplication in monitoring is sufficiently low, it can be the case that direct lending supports a larger external finance than a bank loan, provided that R_D is sufficiently large.

A larger R_D increases the value of the individual investor but lowers that of the firm. This, given the fixed loan rate R_L, releases the pressure on the bank in increasing the size of the loan for inducing the firm's participate, reducing optimal loan size. A higher R_D also reduces the size of direct finance (i.e., $X(k) - k$). However, since interest rates are free to adjust in direct lending, the reduction in the size of direct finance would be less than that in the bank loan.[23] Overall, an increase in R_D would result in smaller bank loans relative to monitored

[21] See Lemma 11 in the appendix.

[22] Lending with the risk free bond could be viewed as an outcome under infinite monitoring costs.

[23] To see this more precisely, remember, for any fixed k, in order to induce the firm to participate, $Z_0(k)$ must satisfy $\pi_2 \{\theta_2 Z_0(k) - R_L [Z_0(k) - k]\} = V(k)$. A higher R_D decreases $V(k)$ which, given that R_L is fixed,

private loans.

4. Equilibrium

Definition 1. *A rational expectations equilibrium of the model consists of a market rate of return on lending for investors r^*, a quantity of deposits D^*, a set $\mathbf{B} \subseteq [0, \bar{k}]$ of firms whom the bank offers a loan to and the corresponding loan contracts $\{(Z(k), R_L) : k \in \mathbf{B}\}$, and the contracts $\{(X(k), r_1(k), r_2(k)) : k \in [0, \bar{k}]\}$ offered in the direct lending market, such that:*

1. *For all $k \geq 0$, the direct lending contract $(X(k), r_1(k), r_2(k))$ is optimal, as described in Section 3.*
2. *Suppose $r^* = R_D$. Then both the direct and indirect lending markets open, and*
   (a) *The set $\mathbf{B}$ and the loan contracts $\{(Z(k), R_L) : k \in \mathbf{B}\}$ solve the bank's optimization problem, as described in Section 3.*
   (b) *Entrepreneurs with net worth $k \in \mathbf{B}$ choose optimally to accept the loan the bank offers, those with $k \notin \mathbf{B}$ obtain finance from the direct lending.*
3. *Suppose $r^* > R_D$. Then there is no lending, with $D^* = 0$ and $\mathbf{B} = \emptyset$.*
4. *The demand for loans equals the supply of loans in the direct lending market:*

$$\mu \int_{[0,\bar{k}]\setminus\mathbf{B}} [X(k) - k] \, dG(k) = M - D^*. \tag{29}$$

The above defined equilibrium of the model is formulated more explicitly in a system of equations in Appendix A.12. We now characterize this equilibrium. To save space, we assume in the remainder of the paper $R_D < \hat{R}_D \equiv \pi_1\theta_1 + 2\pi_2 R_L - \pi_2\theta_2 - \pi_1\gamma_0$.[24]

The bank deposits D plays a key role in defining the model's equilibrium. To characterize the equilibrium, we solve for all other endogenous variables of the model as a function of D, and then let the equilibrium D, together with the equilibrium interest rate, r^*, clear the

forces the bank to decrease $Z_0(k)$ in order to decrease the firm's value on the left hand side of the equation to make it hold. On the other hand, for direct lending, from Lemma 1 and Proposition 2, $X(k)$ must satisfy $\pi_2 [\theta_2 X(k) - r_2(k)(X(k) - k)] = V(k)$. Now for the same decrease in $V(k)$ that results from the increase in R_D, in order to keep the equation hold the direct lender could optimize on two dimensions: $X(k)$ and $r_2(k)$, putting less pressure on the decrease in $X(k)$.

[24] A discussion of the two cases, $R_D < \hat{R}_D$ and $R_D \geq \hat{R}_D$, is given in Lemma 11 and Figure A.14 in the appendix. An earlier version of the paper, which is available by request, includes an analysis for the case of $R_D \geq \hat{R}_D$, which shows that similar outcomes arise between the two cases.

market.[25] Specifically, for any given $D \in [0, M]$ and $r^* \in [R_D, E(\theta))$, let $Q(D, r^*)$ denote the economy's total demand for external finance:

$$\begin{aligned} Q(D, r^*) &= \mu \int_0^{\hat{k}_1(D,r^*)} L_{\mathrm{M}}(k, r^*) dG(k) + \mu \int_{\hat{k}_1(D,r^*)}^{\hat{k}_2(D,r^*)} L_{\mathrm{B}}(k, r^*) dG(k) \\ &+ \mu \int_{\hat{k}_2(D,r^*)}^{\bar{k}} L_{\mathrm{N}}(k, r^*) dG(k), \end{aligned} \tag{30}$$

which is the sum of the demands for monitored direct finance, bank loans, and bond finance. Notice that the second part equals D, as the bank's resource constraint binds.

Figure 5 depicts $Q(D, r^*)$ as a function of D conditional on a given r^*, with $r^* \geq R_D$.[26] Consider first the case of $r^* > R_D$. In this case, in equilibrium there is no bank lending and thus the total demand for external finance, all from the market for direct lending, is

$$Q(0, r^*) = \mu \int_0^{\tilde{k}(r^*)} L_{\mathrm{M}}(k, r^*) dG(k) + \mu \int_{\tilde{k}(r^*)}^{\bar{k}} L_{\mathrm{N}}(k, r^*) dG(k),$$

where $L_{\mathrm{N}}(k, r^*)$ and $L_{\mathrm{M}}(k, r^*)$, given respectively in (5) and (12), are both decreasing in r^*. Depending on the value of r^* then, $Q(0, r^*)$ could take any value between 0 and $\underline{Q}$, where $\underline{Q}$ is the value of $Q(0, r^*)$ at $r^* = R_D$, with which the demand for external finance achieves its maximum conditional on $D = 0$.

What happens in the direct lending market in the case of $D = 0$ is depicted in Figure A.6 in the appendix, where a value of M below $\underline{Q}$ induces an equilibrium interest rate r^* to clear the market. There, for M sufficiently small, $M \leq \underline{M}$ specifically, r^* would be so high that $L_{\mathrm{M}}(k, r^*) = 0$ for all $k \in (0, \tilde{k})$, while $L_{\mathrm{N}}(k, r^*)$ remains positive for all $k \in [\tilde{k}, \bar{k}]$ (from equations (5) and (12)). That is, a sufficiently high interest rate, which results from a sufficiently small supply of external finance, would render monitoring being completely crowded out and the risk free bond being the only financial instrument used.[27]

[25] That the equilibrium quantity of deposits plays a key role in clearing the credit market is a somewhat unique feature of our model, resulting mainly from the fact that the price the bank offers for D, R_D, is fixed in this benchmark version of the model. The fixed R_D also puts a constraint on how effective the equilibrium interest rate, r^*, is in equalizing demand and supply for direct lending. Specifically, r^* is forced to be equal to R_D whenever bank loans are traded in equilibrium.

[26] This is the projection of $Q(D, r^*)$ on the D axis. Note that what the figure depicts is by no means holding r^* fixed. In particular, in the case of $D = 0$, r^* does move to change $Q(D, r^*)$ and clear the market.

[27] Bond finance survives higher interest rates better than monitored private loans. What's giving bond finance an upper hand is the cost of monitoring which occurs with monitored lending but is absent with bond finance. To see this more clearly, remember $L_{\mathrm{N}}(k, r^*) = \frac{\theta_1 k}{r^* - \theta_1}$ and $L_{\mathrm{M}}(k, r^*) = \max\left\{0, \ \frac{E(\theta) - \pi_1 \gamma k - \pi_1 \gamma_0 - r^*}{2\pi_1 \gamma}\right\}$

Consider next the case of $r^* = R_D$. In this case, D could take any value from $(0, M]$. In the appendix, Lemma 12, we show that $Q(D, R_D)$ is strictly increasing in D, as depicted in Figure 5, where $Q_0 \equiv Q(D_0, R_D)$ and $\overline{Q} \equiv Q(\bar{D}, R_D)$. If $D > \bar{D}$, all firms raise credit through the bank, with $Q(D, R_D) = D$. If $0 < D < \bar{D}$, lending takes place both directly and indirectly between firms and investors. In this case, the demand function $Q(D, R_D)$ is upward sloping in D. An increase in D, by taking firms away from direct lending and switching them to bank loans, increases $Q(D, R_D)$, the total demand for credit.[28]

Figure 5 thus gives four cases in terms of how the economy's total supply of external finance, M, is divided in equilibrium among the three different instruments for finance.

Case 1: $M \leq \underline{Q}$. All lending takes place directly between individual firms and investors, as depicted in Figure A.6 in the appendix.

Case 2: $\underline{Q} < M < Q_0$. Three markets open simultaneously in the unique equilibrium of the model, for bank loans, bond finance, and monitored direct finance respectively.

Case 3: $Q_0 < M < \overline{Q}$. Bank loans and monitored direct finance coexist in the unique equilibrium of the model.

Case 4: $M \geq \overline{Q}$. All lending takes place through the bank, as $D^* \geq \bar{D}$ (Proposition 3).

In Cases 2 and 3, where direct lending and bank loans coexist, a larger M implies a larger D^*, which, from Figure 4, implies an expanded set of firms obtaining bank loans but a smaller set of firms participating in direct lending. In other words, an increase in the total supply of external finance induces a crowding out of direct finance by bank loans.

So an increase in M reduces direct lending in both absolute and relative measures. Why is this? Imagine the economy is in an initial equilibrium where direct lending and bank loans coexist. Imagine M is increased by a small positive amount Δ. Any fraction of this Δ could not have flowed into the market for direct lending, for then the interest rate on direct lending

where $L_{\mathrm{N}}(k, r^*)$ is positive for all $r^* < E(\theta)$, whereas $L_{\mathrm{M}}(k, r^*)$ is zero for all $r^* > \bar{r}^* \equiv E(\theta) - \pi_1\gamma_0$. Notice that $\bar{r}^*$ is decreasing in $\pi_1\gamma_0$. That is, a larger expected cost of monitoring makes monitoring more vulnerable in the market for monitoring.

[28]Note that this is conditional on $R_D < \hat{R}_D$ and so bank loans are able to support larger finance relative to direct lending for all k, as depicted in Figure A.14 (a), and so the slope of Q in D is positive at all D. Obviously, if $R_D \geq \hat{R}_D$ and Figure A.14 (b) prevails, then the Q function would not be monotonic in D and that would give rise to multiplicity of the model's equilibrium at some levels of M – the case that is only briefly discussed in the paper, in Appendix A.13.

would fall and investors would flow into bank deposits, which now offer a higher interest rate. In other words, the new funds must become an addition to bank deposits, which now totals $D' \equiv D^* + \Delta$. Given D', however, the bank would re-optimize, to expand its $\mathbf{B}$ to $\mathbf{B}'$, with $\mathbf{B} \subset \mathbf{B}'$. This, in turn, would take firms away from direct lending, reducing the demand for direct lending, lowering the interest rate, and driving investors away from direct lending and into bank deposits, until the interest rate on direct lending is restored at R_D. This process increases the bank's deposits for the second time, say from D' to $D''(> D')$. But then the bank must again re-optimize, with the new D. And this continues until bank deposits settle at a new equilibrium level, which is strictly greater than that in the initial equilibrium. To summarize, the increase in M by Δ results in an increase in banking, D^*, by more than Δ.

Observe also that as bank loans crowd out direct lending following the increase in M, the composition of direct lending also shifts towards smaller shares of bond finance but larger shares of monitored private lending, from Figure 4.

4.1. Bank loans vs. direct lending: existence and co-existence

In addition to M, the deposit rate R_D also plays a key role in determining the model's outcome. Figure 6 shows the equilibrium composition of the market (the existence of each of the markets, for bank loans, bonds, and monitored private lending respectively) in a graph with two dimensions, M and R_D. Here, since Q_0, $\overline{Q}$, and $\underline{Q}$ are all functions of R_D, we write them explicitly as $Q_0(R_D)$, $\overline{Q}(R_D)$ and $\underline{Q}(R_D)$, respectively. These are all decreasing functions and are located relative to each other as the figure depicts.

Figure 6 shows, again, that for fixed R_D, increasing the supply of external finance M shifts the equilibrium composition of lending away from direct finance and towards bank loans; and tightening the supply of external finance squeezes bank lending but expands the market for direct finance. In particular, a sufficiently large M crowds out completely bond finance and monitored private lending to result in an equilibrium where bank loans is the only means of external finance; and a sufficiently small M gives rise to an equilibrium where bonds are the only source of external finance. The intuition, discussed earlier, is that a larger M puts downward pressure on the interest rate on direct lending, giving the bank, who is constrained by the fixed deposit rate, better ability in competing for deposits from the investors which, in turn, gives rise to a larger D and more bank lending, at the expense of direct finance.

The figure also shows that, fixing M, a higher R_D moves the market towards (weakly) more bank loans and less direct lending. On the one hand, a higher R_D gives the bank better ability in competing for deposits, increasing D^* and the loans made. On the other hand, within the direct lending market, a higher R_D dictates more repayments to the individual lender, putting more pressure on the contract in enforcing repayment incentives, making monitored finance more efficient than non-monitored lending (or bonds).

Naturally, the productivity parameters θ_1, θ_2, and π_1 affect the equilibrium composition of lending. A larger θ_1 increases the demand for finance in both the direct and the indirect lending markets, shifting the demand curve $Q(D)$ up and hence the equilibrium towards less bank lending (smaller D^*) but more direct lending, as Figure 5 indicates. On the other hand, a larger θ_2 or a lower π_1 increases the demand for private lending but has no effect on the demand for bond finance, with an ambiguous effect on the demand for bank lending. This leaves the effect on the equilibrium composition of lending ambiguous.

Lastly, the technology of monitoring also matters for the composition of lending. More efficient monitoring, by lowering the cost of monitoring in bank loans and private lending, increases the demand for monitored lending and, as with a larger θ_1, shifts the demand curve $Q(D)$ up in Figure 5, moving the markets towards less bank loans but more direct lending.

5. Banking Reforms

In this section, we use the model to evaluate analytically the effects of three recent reforms in China's banking sector. These reforms took place in a sequence of moves to lift interest rate controls over commercial bank loans and deposits.

Notice first that given the model's linear payoff and production functions, and the efficiency of delegated relative to individual monitoring, setting free the bank's lending rate would result in infinitely large bank loans. To avoid this, we modify the production function $f(\cdot)$ to make it weakly concave, by assuming

$$f(X) = \tilde{\theta}X, \text{ if } X \leq \bar{X} \quad \text{and} \quad f(X) = \tilde{\theta}\bar{X}, \text{ if } X > \bar{X}.$$

That is, there is a cutoff size of the project, $\bar{X}(> 0)$, beyond which any additional investment in the project is not productive.

Conceptually, what $\bar{X}$ captures is a limit to how large individual firms could make themselves be in the capital they employ, or how much output they are able to produce and sell, in the current state of technology and the market environment to which they are exposed. For example, a larger $\bar{X}$ may result from an expansion of the market for the firm's output – a larger market gives the firm an ability to sell more units of its output in a profitable way. Or, $\bar{X}$ may define the boundary to the machines and people that the firm can effectively manage (Lucas, 1978). Naturally, the value of $\bar{X}$ may grow over time, if new technologies or other improvements in the environment make larger firms possible.

Assume $\bar{X}$ is sufficiently large so that $\bar{X} > Z_0(\bar{k})$, and so the outcomes of the model in the prior section (essentially with $\bar{X} = \infty$) continue to hold under $\bar{X}(< \infty)$.[29] Assume $\underline{Q}(R_D) < M < Q_0(R_D)$ so that in equilibrium all three markets coexist prior to the reforms. And lastly, assume all other parts of the model, including the structure of the bond market, remain constant over the time when the reforms took place.

5.1. Removing the lending rate ceiling

In October 2004, the central bank removed the lending rate ceiling on commercial bank loans. This allows banks to set the lending rate on any individual loan anywhere above the floor rate, which continued to exist after the reform.[30] Specifically, after this reform, the bank is now able to choose, in additional to $\mathbf{B}$ (the set of firms in its loan portfolio) and $\{Z(k), k \in \mathbf{B}\}$ (the size of each loan), a lending rate for each loan to make: $\{R_L(k), k \in \mathbf{B}\}$. A full treatment of the bank's problem, which is parallel to that for the case with a fixed lending rate, is given in Appendix A.14.1. There, we show that for any $k \in \mathbf{B}$, it is optimal

[29] More precisely, we need for all $k \in [0, \bar{k}]$, $\bar{X} > \max\{X(k), Z_0(k)\}$.

[30] As part of the same reform, the central bank also replaced the fixed deposit rate with a deposit rate ceiling so the commercial banks were free to set their own rates on deposits subject to the ceiling. The data shows however that this ceiling had almost always been binding. That is, although the banks were allowed to set the deposit rate below the ceiling rate, they almost never chose to do so (see He et al., 2015). Given this, in this section we will confine our attention to cases where the deposit rate ceiling binds and the deposit rate, as in the prior section, is essentially fixed (at the ceiling rate). This however does not mean that the ceiling constraint binds in all cases, independent of how high the ceiling rate is and how large or small the value of M is. Indeed, in Appendix A.14.4, we show that if M and the ceiling rate $\bar{R}_D$ are sufficiently large, then the ceiling would not bind and equilibrium deposit rate is below the ceiling rate. On the other hand, existing research has paid more attention to how banks find ways of getting around the binding regulation. Hachem and Song (2018), for example, study a model of interbank competition where commercial banks issue wealth management products to avoid the rate ceiling.

to set the loan at its maximum size $L(k) = \bar{X} - k$, while the optimal lending rate $R_L(k)$ is set at $\bar{R}_L(k)$ – the maximum rate of return that a loan is able to charge on firm k that must go together with a loan with the maximum size. Remember with a fixed R_L the bank wants the loans to be of the minimum possible size subject to the firm's participation. There, by keeping the loans small, the bank lends to more firms, maximizing the use of firm net worth as collateral for enforcing credit repayments (in the low output state). Here, with a flexible R_L, the bank is able to make loans larger so to charge the highest possible interest rate on each loan, or to maximize its returns per unit of lending.

Being able to choose optimally the loan rate, the bank's optimal plan has

$$\mathbf{B} = \{k : \lambda(k) \geq \lambda^*\}, \tag{31}$$

where

$$\lambda(k) = \frac{(E(\theta) - \pi_1\gamma_0)\bar{X} - V(k)}{\bar{X} - k} - R_D \tag{32}$$

is the bank's expected rate of return on lending to firm k, and λ^* solves

$$\mu \int_{\{k:\, \lambda(k)\geq\lambda^*\}} (\bar{X} - k)dG(k) = D.$$

That is, to maximize profits, the bank includes in its portfolio firms with the largest $\lambda(k)$s subject to the total funds available, as depicted in Figure 7.

A larger k has two effects on $\lambda(k)$. First, it implies a larger $V(k)$ which reduces the returns on lending to firm k. Second, it implies a smaller loan $(\bar{X} - k)$, which results in a higher average return on lending, increasing $\lambda(k)$. In Lemma 13 (in the appendix) we show that $\lambda(k)$ is increasing in k for $k \in [0, \tilde{k}]$, and decreasing in k for $k \geq \tilde{k}$, as in Figure 7, where $\mathbf{B} = [\tilde{k}_1, \tilde{k}_2]$, with $0 < \tilde{k}_1 < \tilde{k} < \tilde{k}_2 < \bar{k}$. Moreover, it follows from Proposition 2 that firms with $k \in [0, \tilde{k}_1)$ seek monitored private finance, and those with $k \in (\tilde{k}_2, \bar{k}]$ obtain credit by way of issuing bonds. So removing the lending rate ceiling does not change the model's prediction that small firms use private loans, medium sized firms are financed with bank loans, and large firms issue bonds. And, from Figure A.7 in the appendix, a larger D, by giving a lower λ^*, results in a lower $\tilde{k}_1$ but a larger $\tilde{k}_2$, implying less bond finance and less private lending.

Proposition 4. *(i) Fixing M and R_D, removing the lending rate ceiling results in a decline in banking and private lending, but an increase in bond finance. (ii) After the removal of the* *lending rate ceiling, an increase in M increases the equilibrium bank deposits and loans, but squeezes bond finance and monitored private lending, as in the case of fixed bank lending rate.*

That is, the 2004 reform should not have altered the direction in which a change in M affects banking relative to direct lending. Consider the story behind (i) of the proposition. After removing the rate ceiling, the bank would want its loans to be larger and charge a higher rate (the $\bar{R}_L(k)$). With the given D, it must then take out from its initial portfolio $\mathbf{B}$ a set of larger firms and replace them with a group of smaller firms. This immediately expands the market for bonds – the larger firms, upon leaving the bank, would get finance by way of issuing bonds – and at the same time reduces private lending. The story continues. The adjustment in the bank's portfolio would result in a net increase in the demand for direct finance, pushing up the interest rate on direct lending. This, in turn, would induce individual investors to substitute bank deposits for direct lending, cutting D and lowering the interest rate on direct lending. With the decreased D, the bank must again adjust its loan portfolio to make $\mathbf{B}$ even smaller, moving more (large) firms into direct lending, pushing up again the interest rate on direct lending, inducing more investors to leave bank deposits and join direct lending. And this goes on, until the market settles at a new and lower equilibrium D, the $\widetilde{D}^*$ in Figure 8, together with an expanded bond market but a smaller market for private lending.

One might think that the 2004 reform, by giving the banks larger flexibilities in commercial lending, should have made them more competitive and therefore larger players in China's financial market. The model suggests, however, that the reform should weaken rather than strengthen the bank's position in the financial system, in terms of how large banking is relative to other types of lending. Obviously, the fixed deposit rate R_D plays an important role in the story. What it does, essentially, is to force the bank to choose larger profits on individual contracts at the expense of the total amount of loans made. Suppose the bank is free to choose optimally both R_L and R_D. Then it may raise R_D at the same time as it increases R_L to at least partially offset the above effect.

To end the discussion in this section, notice that $\bar{X}$, which defines the maximum size of the project, affects the equilibrium composition of lending. A larger $\bar{X}$ makes loans larger ($\bar{X} - k$ larger). All else equal this increases the total demand for bank lending and shifts the demand curve $Q(D)$ up in Figure 5. In turn, this results in a smaller amount of bank lending (small D^*) but a larger amount of direct lending in the model's equilibrium.[31]

5.2. All lending rate controls removed

In July 2013, the central bank also scraped the interest rate floors on commercial bank loans. The effects of this reform depends of course on whether the floor, $\underline{R}_L$, binds prior to the reform. If the floor was below the equilibrium lending rate prior to the reform, then removing it should have no effect on the outcome of the model. If the floor was sufficiently high prior to the reform, then removing it would shift the bank's loan portfolio towards larger firms, by giving away smaller firms to private lending (and hence increase private lending) and taking in larger firms from bond finance (and hence cut bond finance). An illustration of this is in Appendix A.14.2.

5.3. Removing all deposit rate controls

In October 2015 the central bank also lifted its ceiling on the interest rate that commercial banks were able to offer for deposits. With this, all of the central bank's restrictions on interest rates had been removed. To model the effects of this reform, the bank is free in this section to choose the deposit rate R_D, the lending rates $\{R_L(k)\}$, as well as its loan portfolio $\mathbf{B}$ and the loan sizes $\{Z(k)\}_{k\in\mathbf{B}}$ to maximize profits. It then follows, as is shown in Appendix A.14.3, that again the bank's optimal lending portfolio is an interval $\mathbf{B} = [\tilde{k}_1, \tilde{k}_2]$, with $0 < \tilde{k}_1 < \tilde{k} < \tilde{k}_2 < \bar{k}$, except that the values of $\tilde{k}_1$ and $\tilde{k}_2$ differ from those in the case where R_D is fixed.

Naturally, removing the ceiling on R_D tightens the competition between the bank and the firms in the market for direct lending, and this bids up the returns for individual investors. Removing the ceiling on R_D also enables the bank to expand banking at the expense of direct lending, as the following proposition suggests.

Proposition 5. *Let direct lending and bank loans coexist before the reform. Removing the ceiling on R_D results in a higher equilibrium interest rate on direct lending and deposits (r^**

[31] Notice that $\bar{X}$ differs from θ_1, θ_2, and π_1 in affecting the equilibrium composition of lending. A larger $\bar{X}$ makes larger loans feasible/profitable, while the values of θ_1, θ_2, and π_1 govern the economy's average rate of return on capital investment and how risky it is – given that it's below $\bar{X}$ – and, through that, affect the lender's incentives to fund investments. This distinction between a returns rate effect and the firm-size effect has not been captured in earlier models of the financial market.

and R_D higher). It also squeezes the market for direct lending while expanding the market for bank loans (D^* larger). With a higher interest rate, firms in the private lending market each raise a smaller amount of finance ($X(k) - k$ smaller) and operate a smaller project.

A full treatment of the above discussion, including a proof of Proposition 5, is given in Appendix A.14.3.

6. Quantitative Analysis

In this section, we first calibrate the model to China's financial market data to show that it makes sense quantitatively. We then use the calibrated model as a vehicle for interpreting the observed expansion of China's corporate bond market versus the relative decline in banking over the recent years. The calibration also allows us to obtain an estimate for the size of private lending in China. Private lending is missing from existing data but has inspired speculations and concerns from researchers and policy makers.

6.1. The data

The National Bureau of Statistics (NBS) provides annual data on China's total fixed asset investment – which we denote as FA_t for its time t value – and divides that into five categories by the source of the investment: (a) state budget, (b) domestic loans, (c) foreign investment, (d) self-raised funds, and (e) others. Figure A.9 in the appendix depicts the shares of the individual categories in 2002 and 2015, it also defines what is included in each category.

We take FA_t as the data counterpart of the total capital investment in the model, or $M_t + K_t$, the sum of internally and externally raised finance. We take the sum of "domestic loans", which includes "loans of various forms to investing units, including those from banks and non-bank financial institutions, and loans arranged by local governments in the form of special funds", to be the data counterpart of the D in the model.[32] The "others" category of FA_t consists of capital raised through private lending (e.g., P2P loans) – interpreted interpret as the data counterpart of the private lending (or P) in the model, and corporate bond issuance

[32]In the model, D is defined as the funds that the bank raises to make bank loans. So in the calibration, the value of D is calculated as the sum of loans to businesses made from banks and non-bank financial institutions, not including the banks' holdings of corporate bonds and their own funds, or any other assets.

– interpreted as the B in the model, and part of the investment that is funded by the firm's net worth (through for example donations from various sources). Since we do not model the use of "foreign investment" and "state budget" against loans raised in the domestic capital market, we take the stand that the data counterpart of the K in the model includes capital investment in (a), (c), (d) and part of (e).

The NBS does not offer information on what fraction of the "others" category is raised through private lending and how much of it is from the issuance of corporate debt. Information on corporate bonds, as depicted in Figure 1, is obtained from the NBS reports on "aggregate financing to the real economy".

We use the firm's fixed assets as a proxy for its net worth and compute the distribution of firm net worth (together with its support $[0, \bar{k}]$) using data from the 2008 China Economic Census. Specifically, we group firms in the dataset into 1000 percentiles by fixed assets. We then take the median value of each percentile to represent the values of net worth in the whole percentile in the calibration. The data and the calibrated distributions of firm net worth are shown in Figure A.10 in the appendix.

Other data we use for calibrating the model includes the time series of the economy's GDP (Y in the model), and the policy deposit and lending rates (R_D and R_L respectively in the model). All measures are in nominal terms and the CEIC offers direct information on them. The CEIC also provides time series information on the observed rate of non-performing bank loans, which we take to be the calibration of π_1 in the model, as shown in Figure A.11 in the appendix. Remember, in the model, all firms in the bank's portfolio default with the same probability π_1. Notice the decline in that rate, from 26% in 2002 to 1.67% in 2015.

6.2. Firms and the production function

For any period t, given $FA_t \equiv M_t + K_t$, $M_t \equiv P_t + D_t + B_t$, and

$$K_t \equiv \mu_t \int_{k_0}^{\bar{k}} k dG(k), \tag{33}$$

the measure of firms can be computed as

$$\mu_t = \frac{FA_t - D_t - B_t - P_t}{\int_{k_0}^{\bar{k}} k dG(k)}, \tag{34}$$

where P_t, the amount of private lending that actually occurs, will be calibrated along with a set of the model's unknown exogenous values, as to be discussed shortly.

In the calibration, we set the value of $\pi_{1,t}$ to be the share of non-performing bank loans in year t. We use the identity $E(\theta_t)FA_t = Y_t$ to obtain a calibration of the mean of θ_t: $E(\theta_t) = Y_t/FA_t$. Next, taking $\theta_{1,t}$ as a parameter whose value is to be calibrated in the last stage of the calibration, for each t the value of the firm's $\theta_{2,t}$ is given by

$$\theta_{2,t} = \frac{E(\theta_t) - \pi_{1,t}\theta_{1,t}}{1 - \pi_{1,t}}. \tag{35}$$

Lastly, to calibrate the maximum project size $\bar{X}_t$, we use $\bar{k}$ as a benchmark and write $\bar{X}_t$ as $\bar{X}_t = \kappa_t \bar{k}$, and calibrate the parameter κ_t instead of $\bar{X}_t$. Remember $\bar{k}$ is the maximum firm net worth observed. We assume $\kappa_t \geq 1$ so that potentially all firms have a demand for external finance in all periods.[33]

6.3. Additional model elements

In the model, after removing the lending rate ceiling, each firm in the bank's portfolio would attain $\bar{X}$, its maximum size. To deviate from this obviously unrealistic feature, we assume that bank lending is subject to an additional borrowing constraint,

$$L(k) \leq \alpha k, \forall k \tag{36}$$

so that loans cannot exceed a fraction α of the firm's net worth, as in Kiyotaki and Moore (1997). We use this to capture forces that also affect how much lending is possible, in addition to the problems of costly state verification and limited commitment that were already explicitly modeled. These forces may include moral hazard in the commonly understood sense, and competition in the banking sector that gives firms extra ability in extracting surplus from financial contracting but lowers the bank's incentives to lend. They may also include policy related incentives or disincentives that impose restrictions on how much the bank is willing to lend to a specific firm. As can be shown analytically and as depicted in Figure A.8 in the appendix, this constraint lowers the rate of return on lending, but does not change the fact that in equilibrium, smaller firms get finance from private lending market, large firms from

[33] In general, $\bar{X}_t$ can be calibrated as a free parameter for each t, or as a function of t and a small number of the firm's other states.

bank and the largest form bond market.[34]

A second new element we add to the model for calibration is government backed lending, which in China often serves as a means of funding policy related objectives (Song and Xiong, 2018). For example, a large part of the four-trillion-yuan stimulus package after 2008 was implemented through local governments for infrastructure financing, in the form of bank loans to target firms and through government sponsored finance vehicles. Bai et al. (2016) estimate that about 24% of the total assets of China's banking system went to the local government finance vehicles (LGFVs) in 2014. There is, however, no systematic information about specifically how large government backed lending is, where exactly it went and at what prices.

To model government backed lending, we assume that in any period an exogenous amount $G(>0)$ of external finance is channeled to a government backed sector in the forms of bank loans and bonds. This sector consists of a separate collection of enterprises, state owned presumably, who otherwise would not participate in the financial market.

Obviously, introducing government backed lending changes the equilibrium of the model, including in particular the equilibrium composition of lending.[35] In terms of the calibration, the value of G must be chosen to meet the condition $M - G \in (\underline{Q}, \overline{Q})$ so the markets for both direct lending and indirect lending exist in equilibrium, as in the data. Given this, instead of choosing directly the value of G subject to $M - G \in (\underline{Q}, \overline{Q})$, we pick equivalently the value of a parameter ξ so that $M - G = \underline{Q} + \xi(\overline{Q} - \underline{Q})$, where $\xi \in (0, 1)$.[36]

In the calibration, we set the rate of return on government backed lending at R_D to reflect the fact that these funds were priced at below market rates. The composition of G – how it is divided between bank loans and bonds – also matters for the calibration. We assume that in all periods, G is divided between loans and bonds at an exogenous ratio. This ratio is then obtained by way of linear interpolation, using information from the National Audit Office's

[34]Technical details in Appendix A.15.

[35]Specifically, the M in the market clearing condition (32) must now be replaced by $M - G$, and $M - G$ must now replace M in Figures 5 and 6 to determine which markets exist in equilibrium.

[36]Remember from Section 4 that $\underline{Q}$ ($\overline{Q}$) is the total demand for external finance if all firms obtain external finance through direct (indirect) lending. It is straightforward to show that $\underline{Q}$ and $\overline{Q}$ do not depend on G. Technical details on these are in Appendix A.16.

reports on the bonds/(bonds + bank loans) ratio in government backed lending.[37]

6.4. Calibration: procedure and outcome

To compute an equilibrium of the model, information on M – an exogenous part of the model and the sum of bank loans (D), bond finance (B), private lending (P), and government back lending (G) – must be used as an input. As discussed earlier however, information on P and G is not available from published sources. In the calibration, the value of G will be chosen optimally, along with the model's unknown parameters, to let the model's outcome match as well as possible the data targets. But this cannot be applied to P, whose value must be used both in calculating M as an input of the model, and as a target that the model seeks to match. To resolve this difficulty, our idea is to "estimate" the size of private lending and to compute the equilibrium of the model simultaneously in an integrated program, as the following describes.

Let ν denote the set of the model's parameters and exogenous variables that we seek to calibrate. Let the values of the elements in ν be given. For any given value of P (and hence the M obtained from summing up P and values of loans, bonds, and government backed lending), let the amount of private lending that the model generates in its equilibrium be denoted $T(P;\nu)$. We say that $\hat{P}$ is a calibration/estimate for P if it solves

$$P = T(P;\nu), \tag{37}$$

where T being the mapping that generates the model's equilibrium private lending. What (37) requires is that the calibration be internally consistent: the amount of private lending that the model takes as a part of the economy's supply of external finance must be consistent with the amount of private lending that the model produces.[38] Computationally, $\hat{P}$ is obtained simultaneously with the rest of the model's equilibrium outcome, in an iterative procedure.

The model thus leaves us for each period t six unknown parameters, $\{\gamma_{0,t}, \gamma_t, \alpha_t, \theta_{1,t}, \kappa_t, \xi_t\}$, whose values are then chosen optimally to match a set of calibration targets. These targets

[37] The reports cover two data points: December 2010 and June 2013. The interpolation results are shown in Figure A.12 in the appendix.

[38] Standard calibrations have no use for such a fixed point argument each time an equilibrium of the model is computed.

include the observed amounts of bank loans, bond finance, and the average bank lending rate across all loans, over the period 2002-2015. A metric on the variability of the bank lending rate across loans and time could also be used as a target for the model to match, but we save that as a test for whether the calibration does well on at least one un-targeted moment. Note that these targets, for the calibration and for testing the calibration, have included the model's major outputs with observed moments.

To reduce dimensionality, for each $x \in \{\gamma_0, \gamma, \alpha, \theta_1, \kappa, \xi\}$ we put the following restriction on its time series: $x_t = x_1 + (x_T - x_1)(t-1)/(T-1)$, where x_1 and x_T are the initial and end values of x, respectively.[39]

Table 1 gives the values of the model's calibrated parameters. Notice that γ_0, which captures how large the fixed part is in the total cost of monitoring, is almost constant across the sample periods, while the parameter that represents the variable part, γ_1, went up slightly over the same time. With the increase in α, constraint (36), which captures the role of the lending frictions in addition to costly state verification and limited commitment, was tighter in 2002 than in 2015. The value of θ_1 is almost constant over the sample periods. Remember $E(\theta_t)$ has declined over the sample years. The calibrated value of κ_1 has gone up between 2002-2015 to indicate an increasing $\bar{X}_t$ in t or increased production capabilities. Lastly, the calibrated government backed lending has gone down significantly over the sample time.

Figure 9 shows how the calibrated model performs in matching the data targets. To show how the model does on an un-targeted metric, Figure A.13 in the appendix shows the fraction of bank loans that charged a rate that exceeds the benchmark rate (determined by the PBC) in the calibrated model versus in the data. This un-targeted metric reflects the variability of the lending rate charged by the bank, and the model does well on that.

Panels III and IV of Figure 9 display the shares of private loans and government backed lending in total external finance, respectively. Note that the share of government backed lending had been fallen over the sample period, from 58% in 2002 to 38% in 2015. In addition, the time series of the $E(\theta_t)$, $\theta_{1,t}$ and $\theta_{2,t}$ in the calibrated model are plotted in Figure A.11 in the appendix. Notice the secular and significant decline in $E(\theta_t)$ which, in the calibrated model,

[39] The calibration outcomes are sensitive to the starting points chosen. For robustness, we used the method of "MultiStart" in the Matlab to get the best local solution for the calibration.

measures the average capital productivity (rate of return on capital) of Chinese firms.[40] Also, the calibrated $\theta_{1,t}$, $\theta_{2,t}$ and their difference had all been declining, suggesting also a decline in the variability in the productivity of capital among Chinese firms over the sample period – because the low output occurred with lower probabilities.

6.5. The size of private lending

How large is China's private lending market? The calibration allows us to obtain, for each t, an estimate for the measure of private lending as a fraction of total external finance. Remember private lending consists of non-delegated monitors, such as relatives, money lenders, and other less delegated monitors including peer-to-peer platforms, and that this market was estimated to be quite large according to some studies. Panel III of Figure 9, which plots the time series of P_t/M_t in the calibrated model, shows that over the sample periods the size of private lending has varied in the range of 0 to 10 percent of the economy's total external finance and had been growing after 2002.

6.6. The bond market expansion

What does the calibrated model say about the observed bond market expansion that Figure 1 depicts? We answer this question with a series of counterfactual experiments whose outcomes are reported in Table 2. In the table, column (1) reports the fraction of bonds in the sum of bonds and bank loans (the bonds/(bonds+loans) ratio) in the calibrated model. Columns (2)-(5) answer the following question: if the value of the parameter considered stayed constant at its 2002 level, what would the bonds/(bonds+loans) ratio be at 2015? Column (6) reports what the bonds/(bonds+loans) ratio would be at 2015 if the interest rate reform in 2004 did not take place. Lastly, column (7) reports what the 2015 bonds/(bonds+loans) ratio would be if government backed lending stayed at its 2002 level and composition.

Obviously, γ_0 and γ did not matter much for explaining the rising bond market. On the other hand, if the values of θ_1, θ_2 and π_1 stayed constant over time, the model would have predicted a (slightly) higher (not lower) bonds/(bonds+bank-loans) ratio than in the data, so these parameters should not matter for interpreting the observed bond market expansion

[40] The decline in the return of capital in China over the recent years has been observed in the literature, by Bai and Zhang (2015) for example.

either. Table 2 however does suggest that the changes in the values of κ and α, the 2004 lending rate reform, and the movements in the quantity and composition of government backed lending might play essential roles in explaining the bond market expansion in the calibrated model. Specifically, suppose either the value of κ did not move up as in the calibration, or the value of α stayed constant at its 2002 level, or the reform did not occur in and after 2004, or the quantity and composition of the government backed lending went invariant over the time period. Then the bonds/(bonds+bank-loans) ratio would have gone up by a magnitude that is far short of what is in the calibration or in the data.

To conclude, the model suggests that several factors – related respectively to the changes in the economy's physical environment, the reforms on banking, and the actions of the government – might have played essential roles in explaining the observed bond market expansion in China. Note that the ways in which these factors affect bond finance and bank loans are all consistent with the logic of the model. For example, holding government backed lending constant at its 2002 level essentially means less government backed lending in the years after 2002 in the calibrated model which, in turn, means a larger supply of credit in the markets for external finance. But this expansion in external finance lowers the market interest rate and should then favor bank loans, at the expense of bond finance and private lending.

7. Concluding Remarks

In this paper, we have studied a model of the financial market where bank loans, corporate bonds and private lending compete and coexist as alternative means of external finance. The paper shows that the model can be used for evaluating, both qualitatively and quantitatively, the impact of the recent financial reforms in China. The model is "small", and may be extended in potentially many ways to capture more features of China's financial system. For example, in the model, all investment projects are identical and face the same probabilities of failure and success and, because of this, firms are treated differently in the financial market only because their net worth supports more or less efficient lending, not because some projects are fundamentally safer or riskier than others. Obviously, one way to enrich the current setup is to let firms face heterogeneous productivity risks, and to let the equilibrium of the model decide what risk types get allocated what types of finance. The current model also abstracts from interbank competition. As China's banking environment moves towards less restrictions on entry and more competition between commercial banks (e.g., Gao et al., 2019), how would the financial system evolve in the way it supports external finance? To answer a question like this, the model need be modified to bring competition into the banking sector. The model may also be used as a vehicle for evaluating the effects of a specific financial instrument, the wealth management products for example, on the performance of the financial system. We leave interesting and important topics like these as possibilities for future research.

Reference

Allen, F. and J. Qian (2014) "China's Financial System and the Law," *Cornell International Law Journal*, Vol. 47, pp. 499–553.

Allen, F., J. Qian, and M. Qian (2005) "Law, Finance, and Economic Growth in China," *Journal of Financial Economics*, Vol. 77, pp. 57–116.

Ayyagari, M., A. Demirgüç-Kunt, and V. Maksimovic (2010) "Formal versus Informal Finance: Evidence from China," *Review of Financial Studies*, Vol. 23, pp. 3048–3097.

Bai, C., C. Hsieh, and Z. Song (2016) "The Long Shadow of a Fiscal Expansion," *NBER Working Paper No. 22801.*

Bai, Chong-En. and Q. Zhang (2015) "The Return on Capital in China and Its Determinants," *China Economist*, Vol. 10(3), pp. 20–37.

Bernanke, B. and M. Gertler (1989) "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, Vol. 79(1), pp. 14–31.

Boyd, J. H. and E. C. Prescott (1986) "Financial Intermediary-coalitions," *Journal of Economic Theory*, Vol. 38(2), pp. 211–232.

Chang, R., A. Fernéndez and A. Gulan (2017) "Bond Finance, Bank Credit, and Aggregate Fluctuations in an Open Economy," *Journal of Monetary Economics*, Vol. 85, pp. 90–109.

Chen, K., J. Ren, and T. Zha (2016) "What We Learn from China's Rising Shadow Banking: Exploring the Nexus of Monetary Tightening and Banks' Role in Entrusted Lending," *NBER Working Paper No. 21890*.

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

Duan, C. and C. Wang (2014) "China's Financial Markets in a CSV Model," *Manuscript*.

Elliott, D., A. Kroeber, and Y. Qiao (2015) "Shadow Banking in China: A Primer," *the Brookings Institution*.

Gale, D. and M. Hellwig (1985) "Incentive-Compatible Debt Contracts: The One-Period Problem," *Review of Economic Studies*, Vol. 52(4), pp. 647–663.

Gao, H., H. Ru, R. Townsend, and X. Yang (2019) "Rise of Bank Competition: Evidence from Banking Deregulation in China," *NBER Working Paper No. 25795*.

Greenwood, J. and B. Jovanovic (1990) "Financial Development, Growth, and the Distribution of Income," *Journal of Political Economy*, Vol. 98, pp. 1076–1107.

Greenwood, J., J. M. Sanchez, and C. Wang (2010) "Financing Development: The Role of Information Costs," *American Economic Review*, Vol. 100, pp. 1875–1891.

Hachem, K. C. and Z. M. Song (2018) "Liquidity Rules and Credit Booms," *NBER Working Paper No. 21880*.

He, D., H. Wang, and X. Yu (2015) "Interest Rate Determination in China: Past, Present, and Future," *International Journal of Central Banking*, Vol. 11, pp. 255–277.

Holmström, B. and J. Tirole (1997) "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics*, Vol. 112(3), pp. 663–691.

Kiyotaki, N. and J. Moore (1997) "Credit Cycles," *Journal of Political Economy*, Vol. 105(2), pp. 211–248.

Lin, L. and C. J. Milhaupt (2017) "Bonded to the State: A Network Perspective on China's Corporate Debt Market," *Journal of Financial Regulation*, Vol. 3(1), pp. 1–39.

Lu, Y., H. Guo, E. Kao, and H. Fung (2015) "Shadow banking and firm financing in China," *International Review of Economics & Finance*, Vol. 36, pp. 40–53.

Lucas, R. E. (1978) "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, Vol. 9(2), pp. 508–523.

Song, Z, and W. Xiong (2018) "Risks in China's Financial System,"*NBER Working Paper No. 24230*.

Townsend, R. M. (1979) "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, Vol. 21(2), pp. 265–293.

Wang, C. and S. D. Williamson (1998) "Debt Contracts and Financial Intermediation with Costly Screening," *The Canadian Journal of Economics/Revue canadienne d'Economique*, Vol. 31(3), pp. 573–595.

Wang, H, H. Wang, L. Wang, and H. Zhou (2015) "Shadow Banking: China's Dual-Track Interest Rate Liberalization," *Manuscript*.

Williamson, S. D. (1986) "Costly Monitoring, Financial Intermediation, and Equilibrium Credit Rationing," *Journal of Monetary Economics*, Vol. 18, pp. 159–179.

8. Tables and Figures

Table 1: Calibrated parameters

Parameter	Symbol	Initial value	End value	Source
Distribution of firm net worth	$G(\cdot)$	constant over time		2008 China Economic Census
Fixed assets investment	FA	4.350E+06	5.562E+07	CEIC, million yuan
GDP	Y	1.217E+07	6.891E+07	CEIC, million yuan
Probability of low productivity	π_1	0.26	0.017	CEIC
Total domestic loans in FA	D	8.555E+05	5.873E+06	CEIC, million yuan
Total corporate bond issuance in FA	B	3.670E+04	2.939E+06	CEIC, million yuan
The share of bonds in GBL	ω	0	22.7%	NAO's reports
Fixed cost of monitoring parameter	γ_0	1.018E-02	1.015E-02	Calibrated
Variable cost of monitoring parameter	γ	0.995E-04	1.018E-04	Calibrated
Collateral constraint parameter	α	1.121	1.138	Calibrated
Level of low productivity	θ_1	0.103	0.100	Calibrated
The maximum project size parameter	κ_1	1.066	1.105	Calibrated
Government backed lending parameter	ξ	0.787	0.735	Calibrated
Total private lending in FA	P	1.834E+03	7.928E+05	Calibrated, million yuan
Total government backed lending in FA	GBL	5.224E+05	3.617E+06	Calibrated, million yuan
Level of high productivity	θ_2	3.745	1.245	Calculated as $(Y/FA - \pi_1\theta_1)/(1-\pi_1)$
Total external credit in FA	M	8.940E+05	9.605E+06	Calculated as $P + D + B$, million yuan
The measure of firms	μ	0.491	6.626	Calculated as $(FA - M)/\int_0^{\bar{k}} kdG(\cdot)$

Table 2: Bonds/(bonds+bank loans): counterfactual outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
year	calibration	γ_0, γ	θ_1, θ_2, π_1	κ	α	2004 reform	GBL
2002	3.45%	3.45%	3.45%	3.45%	3.45%	3.45%	3.45%
2015	27.40%	27.44%	29.86%	14.02%	16.68%	16.41%	10.14%

Note: “2004 reform” denotes the lending rate reform in 2004, “GBL” denotes government backed lending.

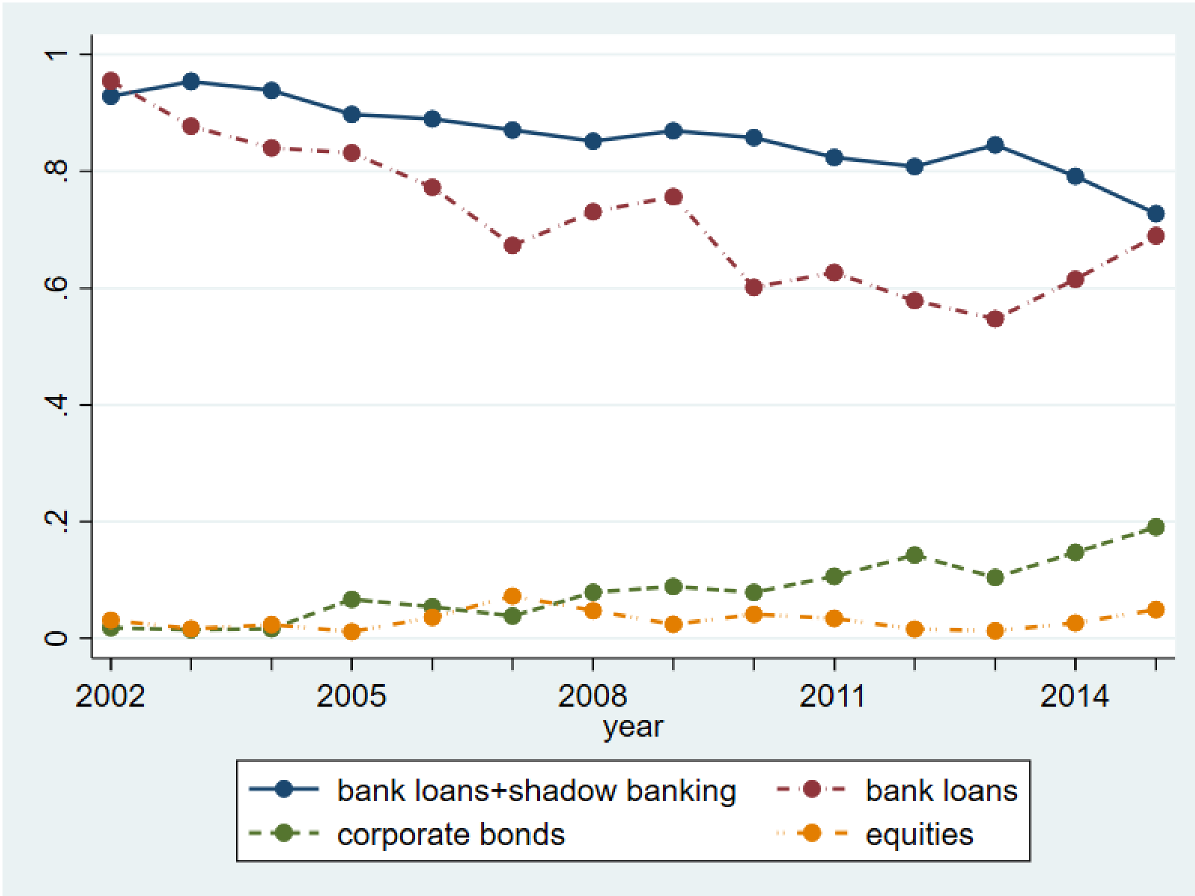


Figure 1: Composition of aggregate financing in China

Source: CEIC.[41]

Note: "Bank loans" is defined as (loans in local currency + loans in foreign currency)/aggregate financing; "shadow banking" is defined as (trust loans + entrusted loans + banker's acceptance bills)/aggregate financing; "corporate bonds" is defined as corporate bond financing/aggregate financing; and "equities" is defined as non-financial enterprise equity financing/aggregate financing.

[41] The CEIC Database, created by the Euromoney Institutional Investor, provides expansive macro data for a large set of developed and developing economies around the world. We draw information from this database multiple times in this paper.

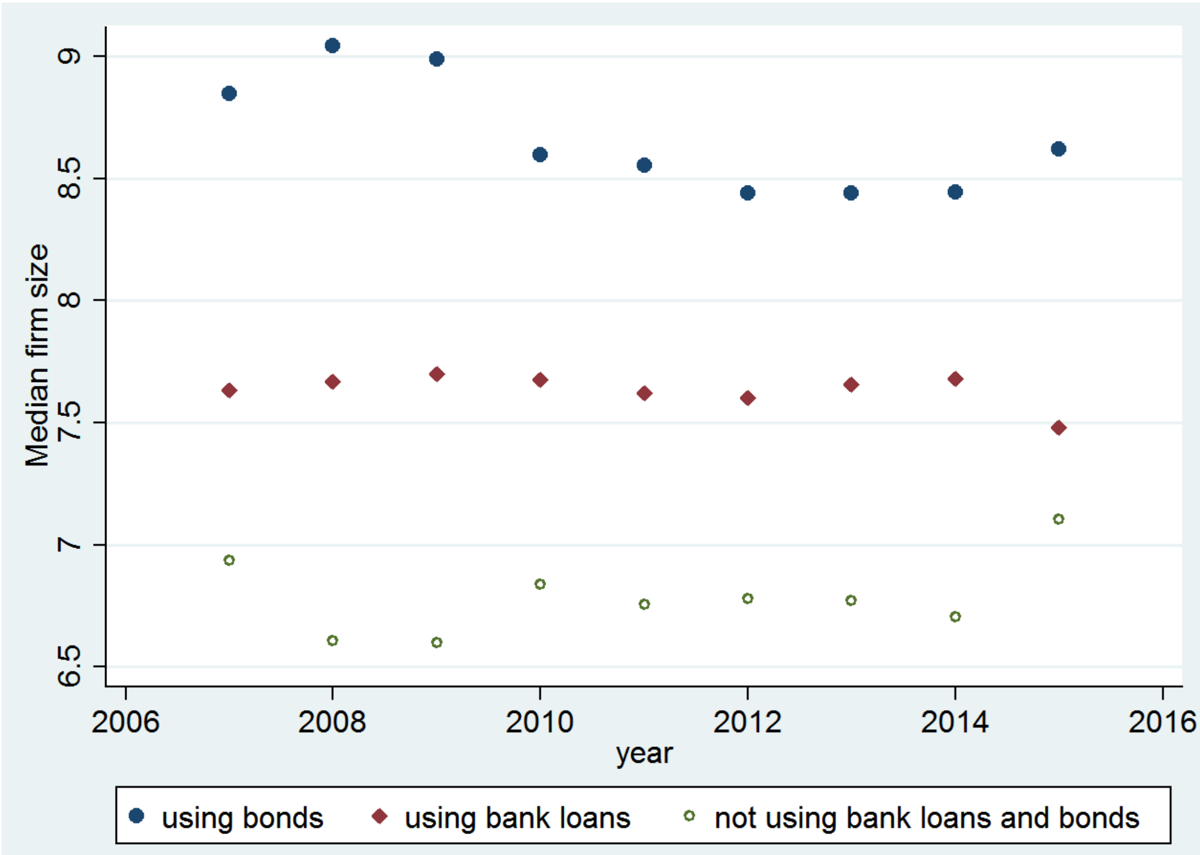


Figure 2: Using versus not using bonds: the median size of listed firms in China, 2007-2015

Source: CSMAR.[42]

Note: Values on the vertical axis are in logarithm. The solid dots represent the median of employment in firms that use bonds (and possibly other instruments) for external finance. The solid squares represent the median of employment in firms that use bank loans (and possibly other instruments) for external finance. The hollow dots represent the median measure of employment of all other firms.

[42]CSMAR (China Stock Market & Accounting Research) Database, developed by GTA Information Technology, covers data on the Chinese stock market, financial statements and China Corporate Governance of Chinese Listed Firms.

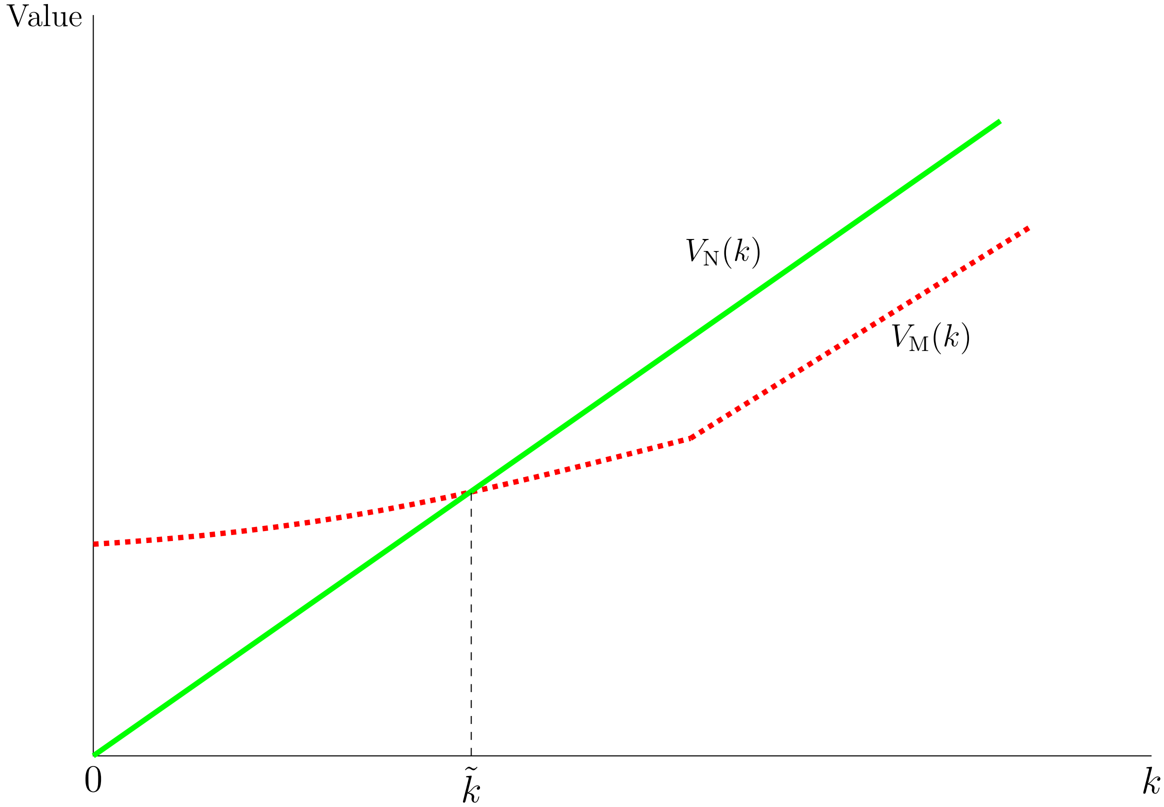


Figure 3: Lender's value functions in direct lending

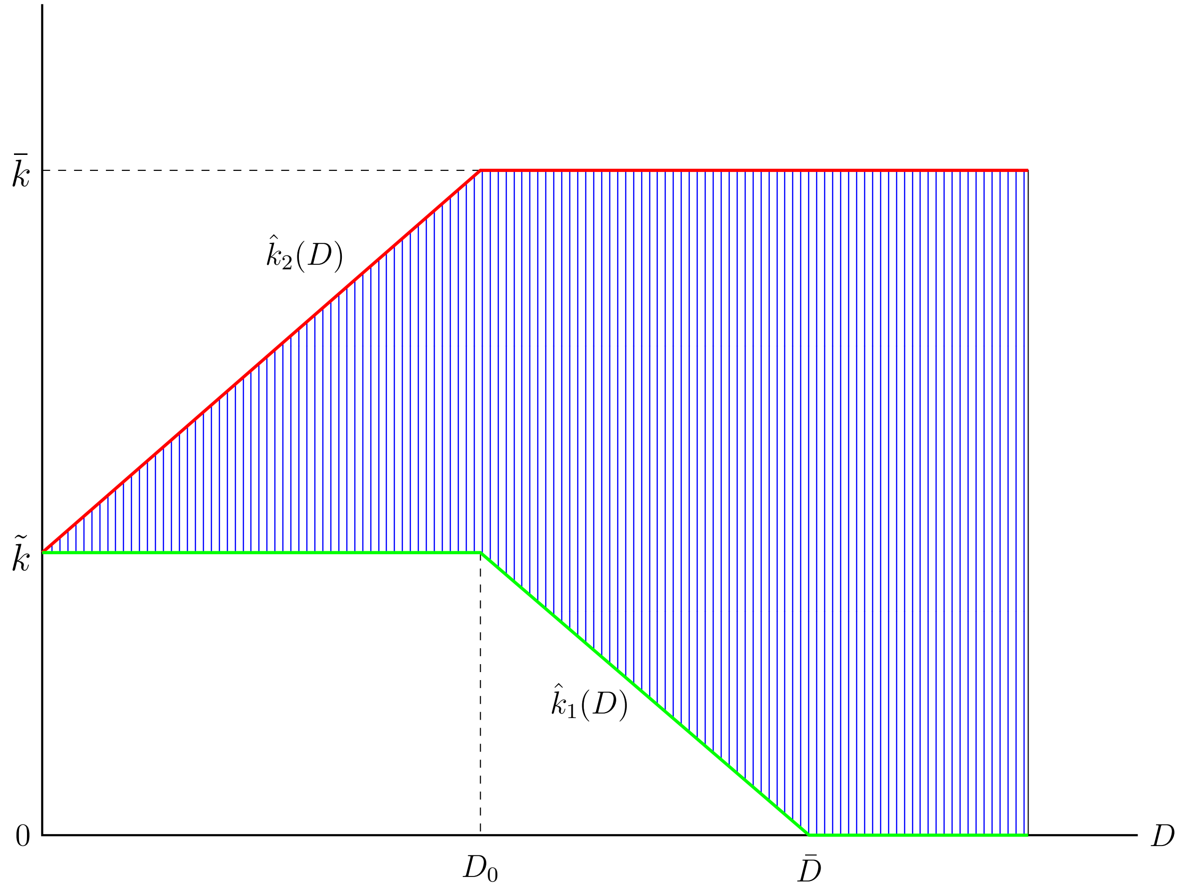


Figure 4: The bank's optimal loan portfolio **B** as a function of D

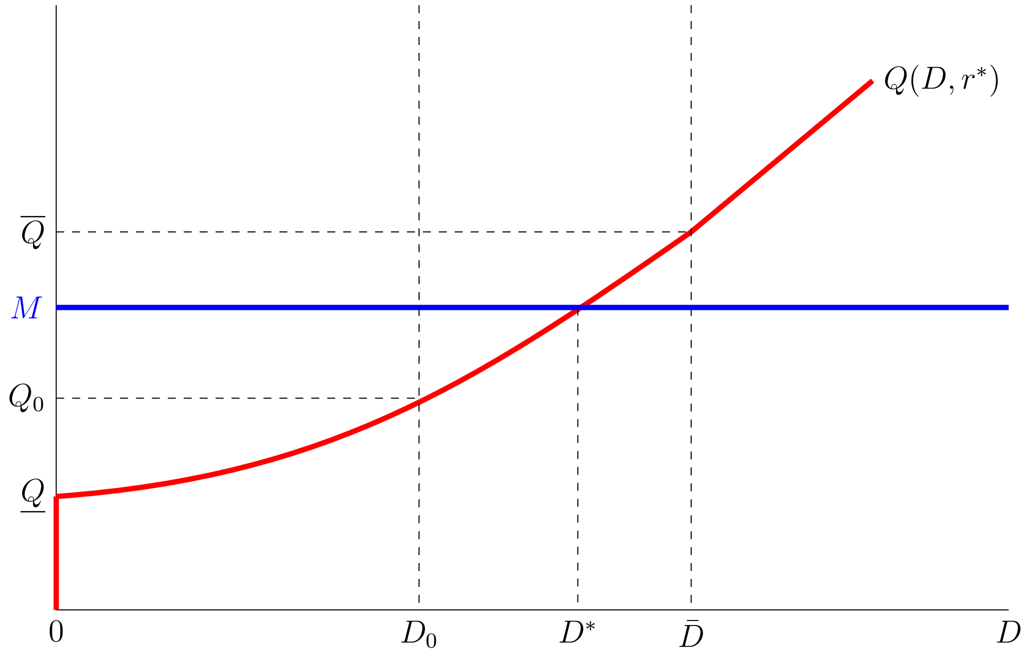


Figure 5: Equilibrium when $0 < M < \underline{Q}$

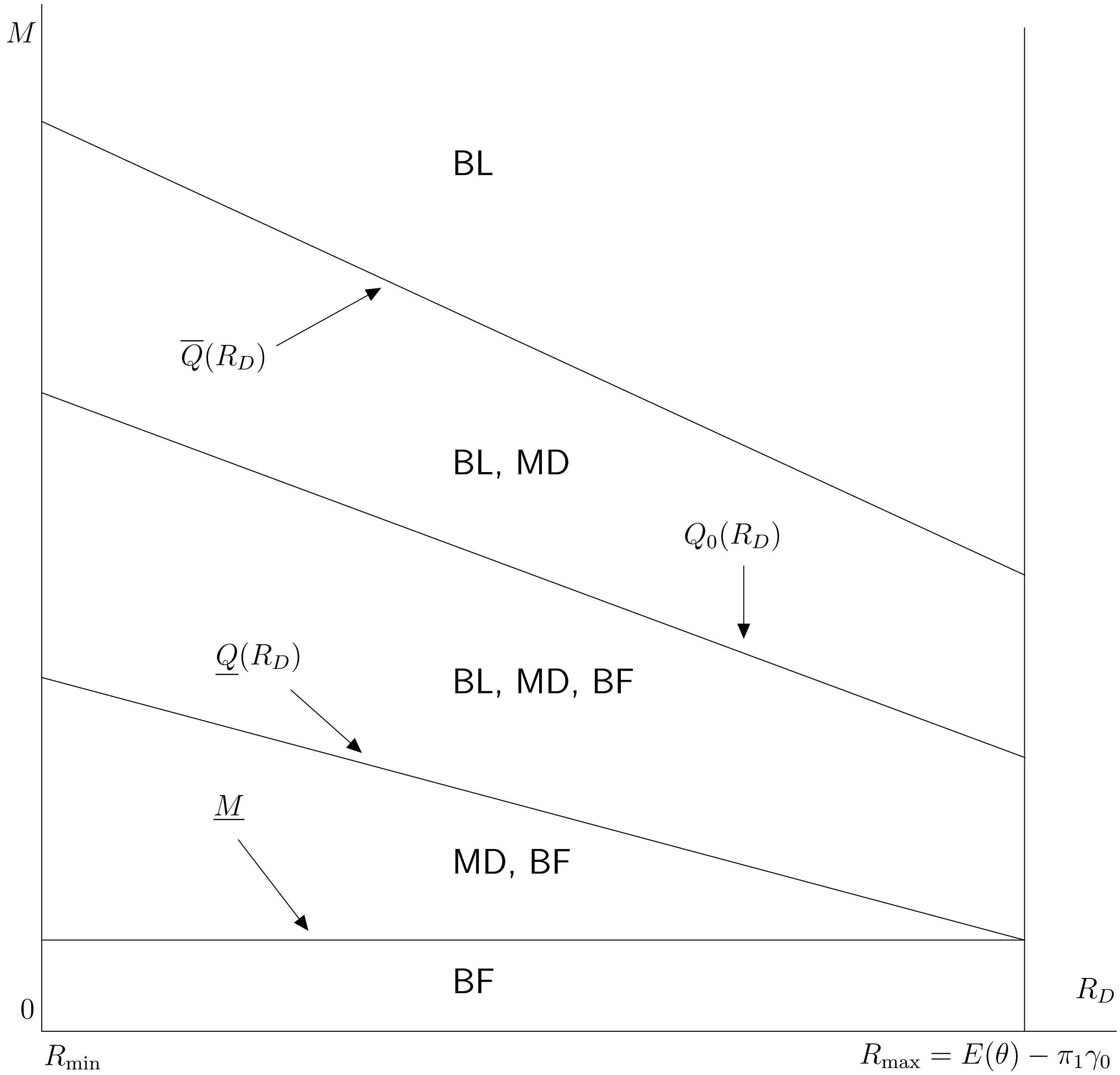


Figure 6: Equilibria with respect to R_D and M

Note: This figure shows the existence and coexistence of the three distinctive markets for finance (bank loans, corporate bond, and monitored direct finance) in the equilibrium of the model with any given R_D and M. Here BL denotes bank loans, MD denotes monitored directed finance, BF denotes bond finance. The area (BL, MD), for example, includes all pairs of (R_D, M) with which bank loans (i.e., BL) and monitored direct finance (i.e., MD) coexist in the model's equilibrium.

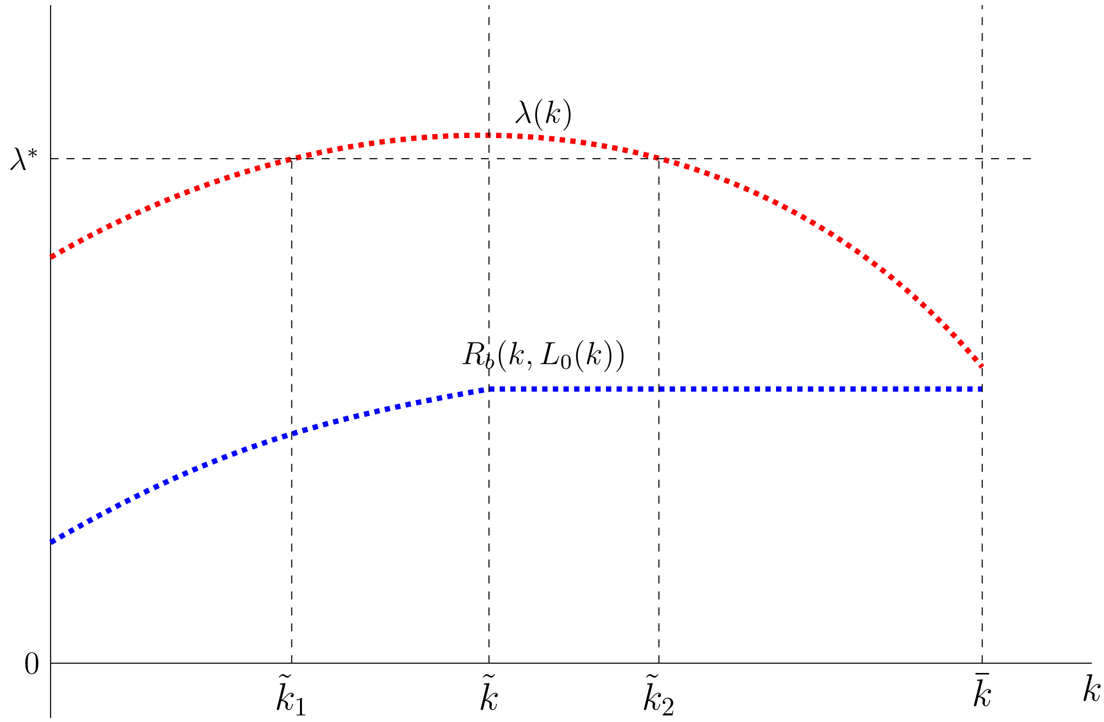


Figure 7: The bank's optimal loan portfolio: $\mathbf{B} = [\tilde{k}_1, \tilde{k}_2]$

Note: This figure compares $\lambda(k)$ with $R_b(k, L_0(k))$ (which is the bank's expected rate of return on lending to firm k with the fixed lending rate). The bank's expected rate of return on lending to firm k is higher after the removal of the lending rate ceiling, for any $k \in [0, \bar{k}]$.

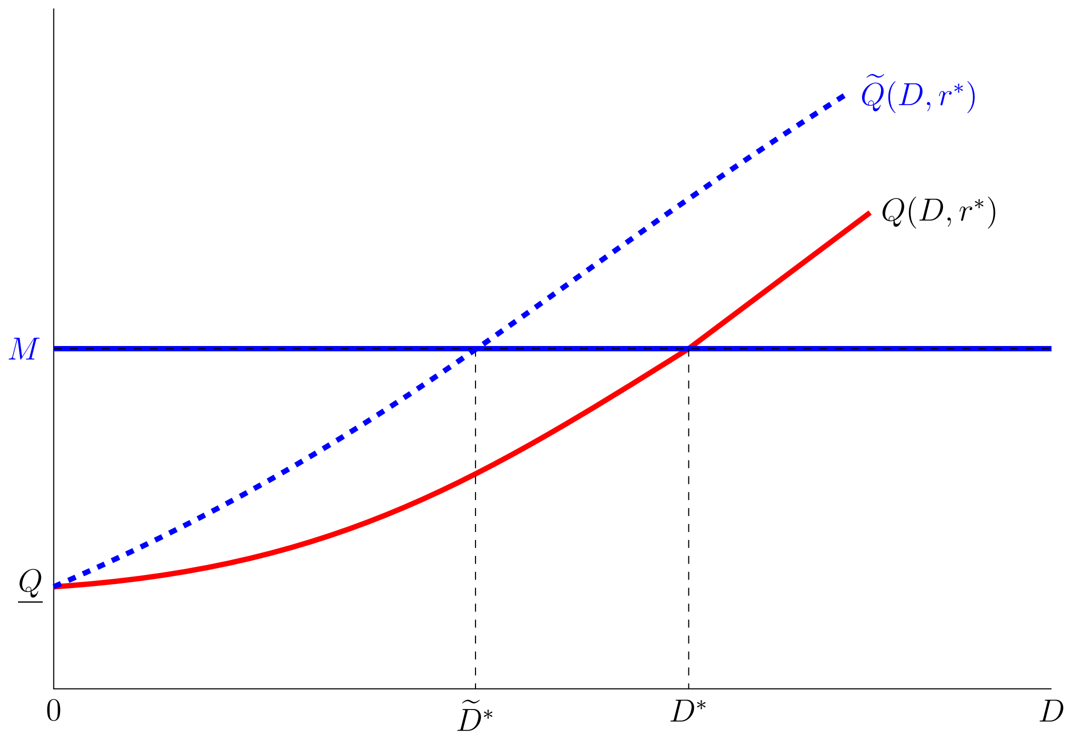


Figure 8: Equilibrium after removing the lending rate ceiling

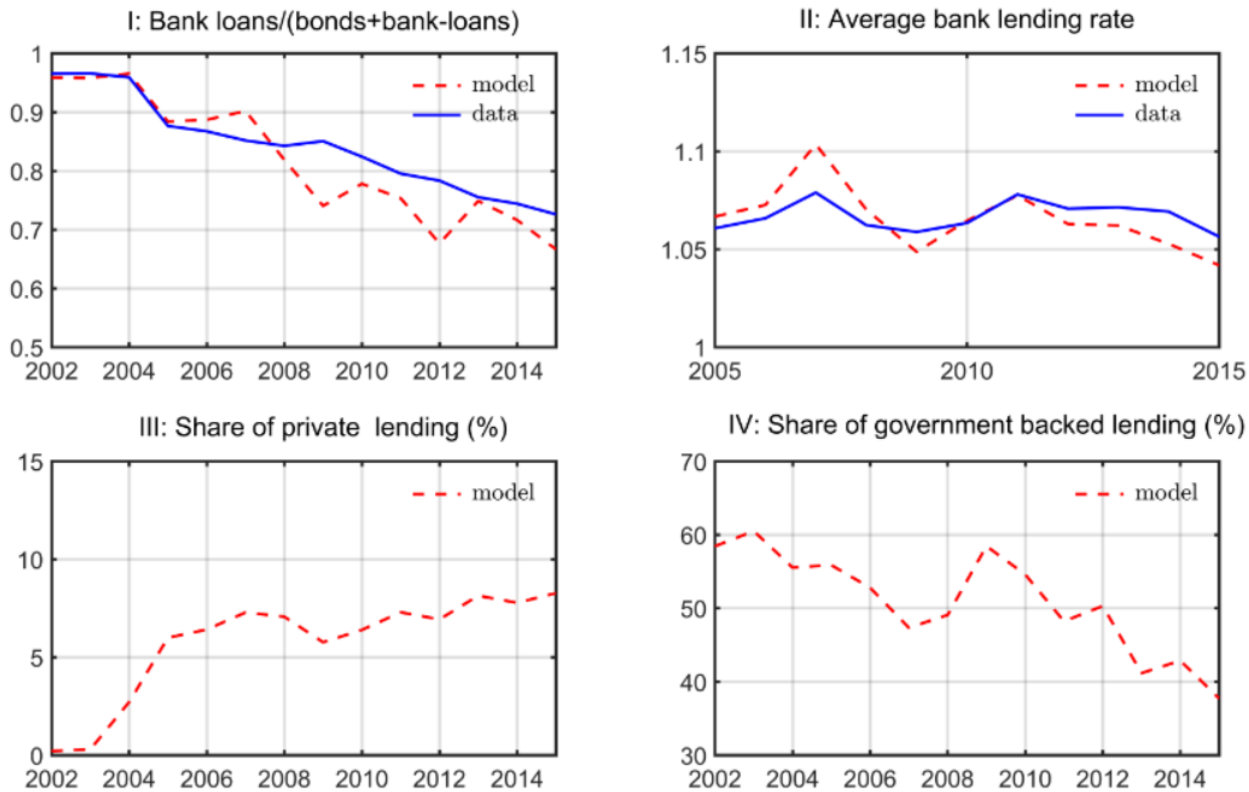


Figure 9: Calibration: data and model

Note: The observed average lending rate and information on the fraction of loans that charge a rate higher than the policy lending rate is available only from 2005, after the lending rate ceiling was removed.

Appendix A.

Appendix A.1. Additional tables and figures

Table A.1: Percent of firms who did not apply for a loan for a given reason, across size groups

	Small (5-19)	Medium (20-99)	Large (100+)
No need for a loan	53.5	56.1	64.9
Application procedures were complex	13.8	9.5	8.5
Interest rates were not favorable	6.6	12.8	11.5
Collateral requirements were too high	8.7	9.8	6.3
Size of loan and maturity were insufficient	9.2	5.8	3.0
Did not think it would be approved	6.2	3.4	2.2
Other	2.0	2.7	3.7

Source: World Bank's Enterprise Surveys data for China 2012.

Table A.2: Number of firms in China, by firm size and sources of finance

(a) Within manufacturing firms in the World Bank's Enterprise Surveys for China, 2011

Employment	Total number	No external finance	Only bank finance	Both bank and other finances	Only other finances
6 – 40	190	153	13	4	20
40 – 80	189	140	17	15	17
80 – 120	189	141	18	12	18
120 – 272	189	142	19	14	14
272+	189	123	30	12	24

(b) Within listed manufacturing firms in China, 2011

Employment	Total number	No external finance	Only bank finance	Both bank and other finances	Only other finances
3 – 714	275	15	66	160	34
714 – 1401	274	21	69	171	13
1401 – 2522	274	12	77	178	7
2522 – 5254	274	4	75	189	6
5254+	274	4	55	208	7

Source: Self-calculated using World Bank's Enterprise Surveys data for China 2012 and the CSMAR.

Note: Other instruments of finance include equity, bond and trade credit, et al.

Figure A.1: 5-bank loans concentration in commercial banks, China and U.S.

Source: Bankscope, self-calculations.

Note: In 2015, the 5 largest commercial banks in China are Industrial & Commercial Bank of China, China Construction Bank, Bank of China, Agricultural Bank of China, Bank of Communications. In the U.S., they are Wells Fargo Bank, Bank of America, JPMorgan Chase, Citibank, and U.S. Bank National Association. The 5-bank concentration within bank holding companies in the U.S. is similar to that within Commercial banks.

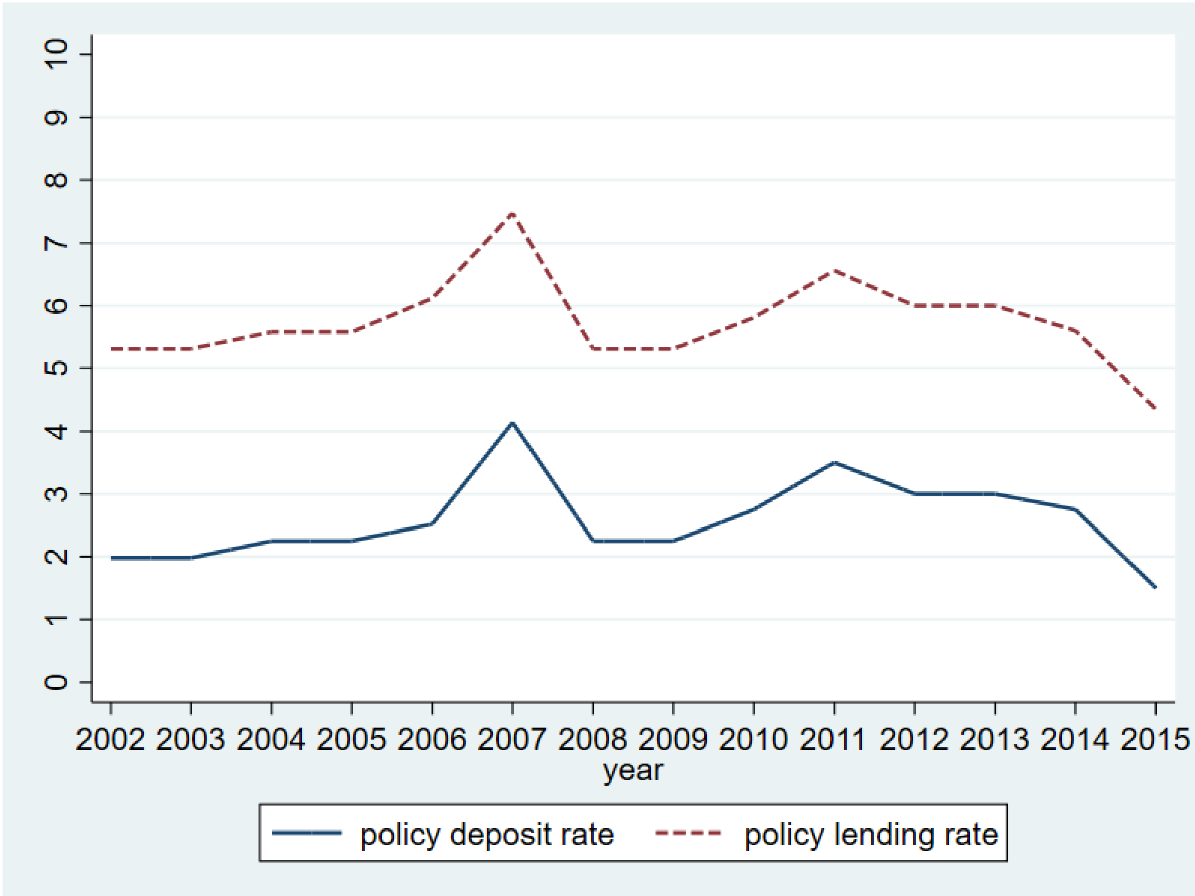


Figure A.2: Yearly policy lending and deposit rates in China

Source: CEIC.

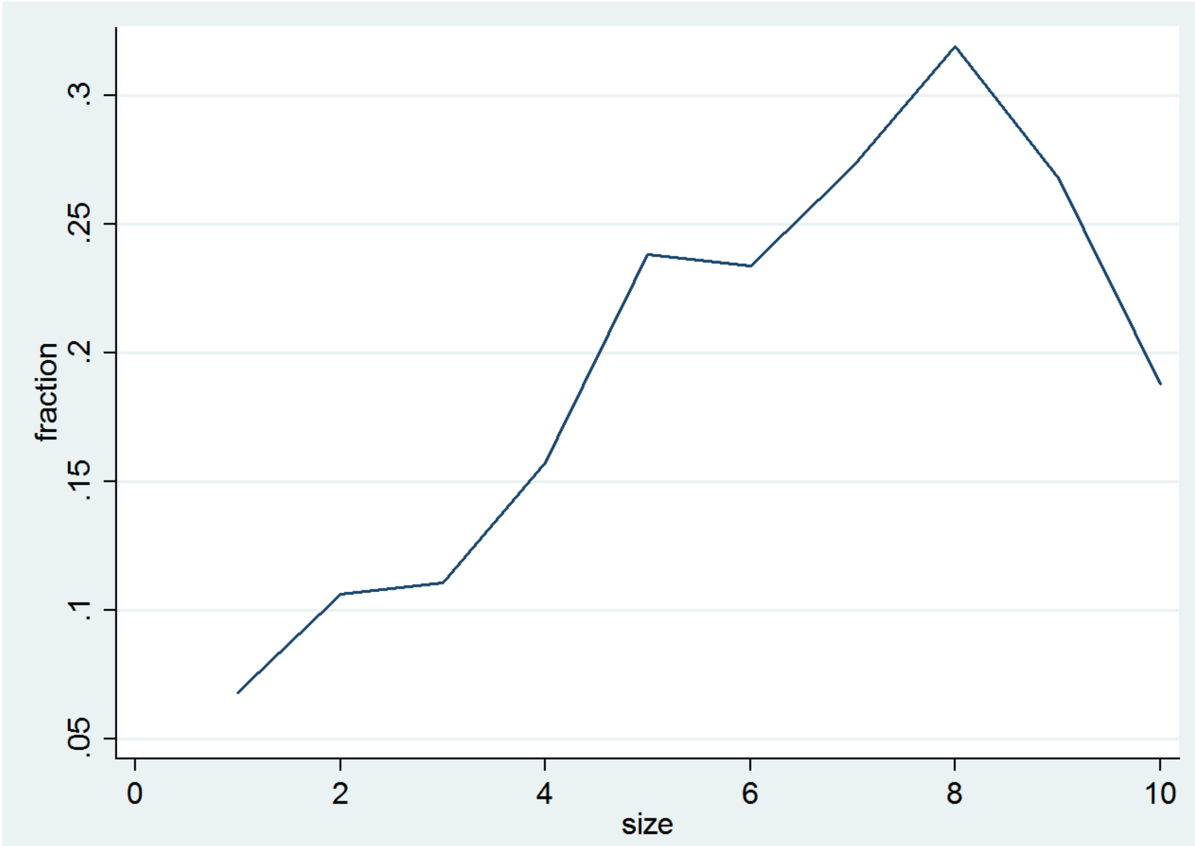


Figure A.3: Fraction of firms in China with only bank finance, 2011

Source: World Bank's Enterprise Surveys data for China 2012 and CSMAR.

Note: The X-axis represents the firm's group number, with a larger value indicating a larger firm size.

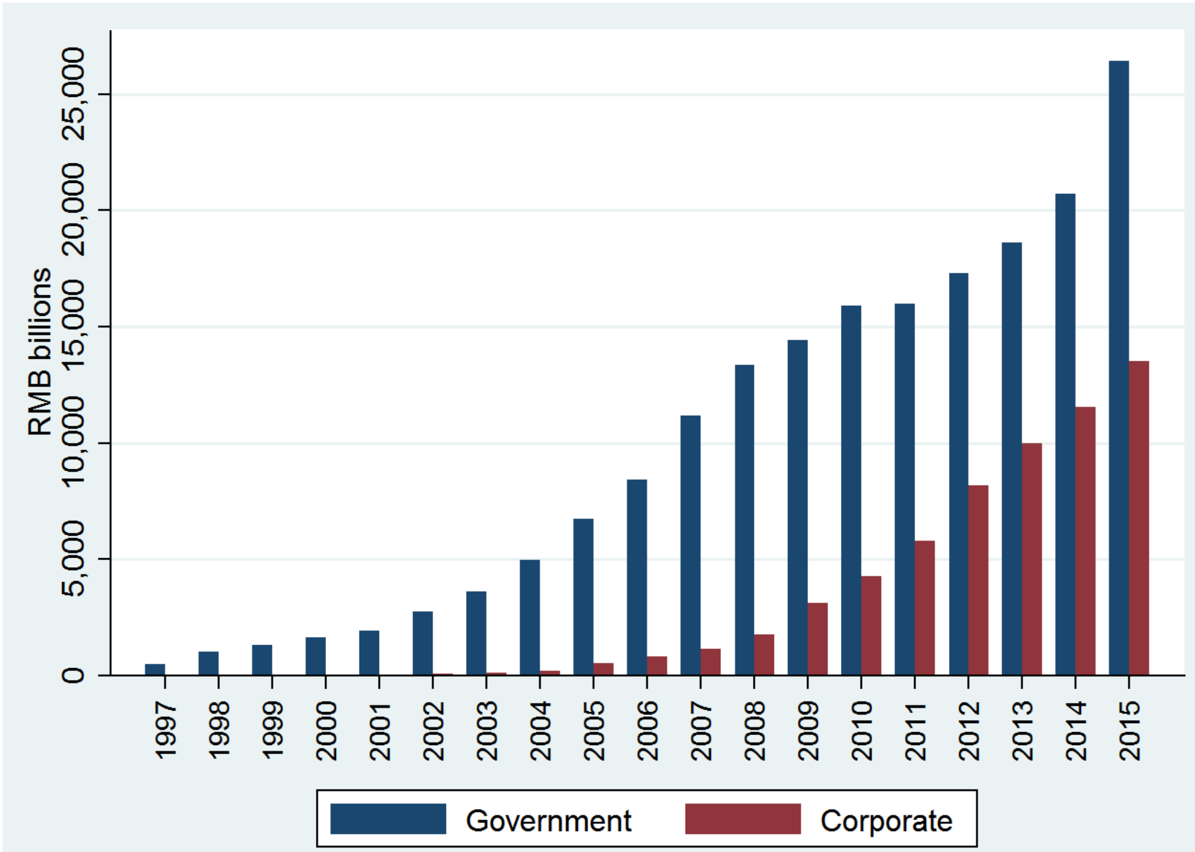


Figure A.4: Size of local currency bonds in China

Source: AsianBondsOnline.

Note: Government bonds include obligations of the central government, local governments, and the central bank. Corporate bonds include debt instruments issued by both public and private companies.

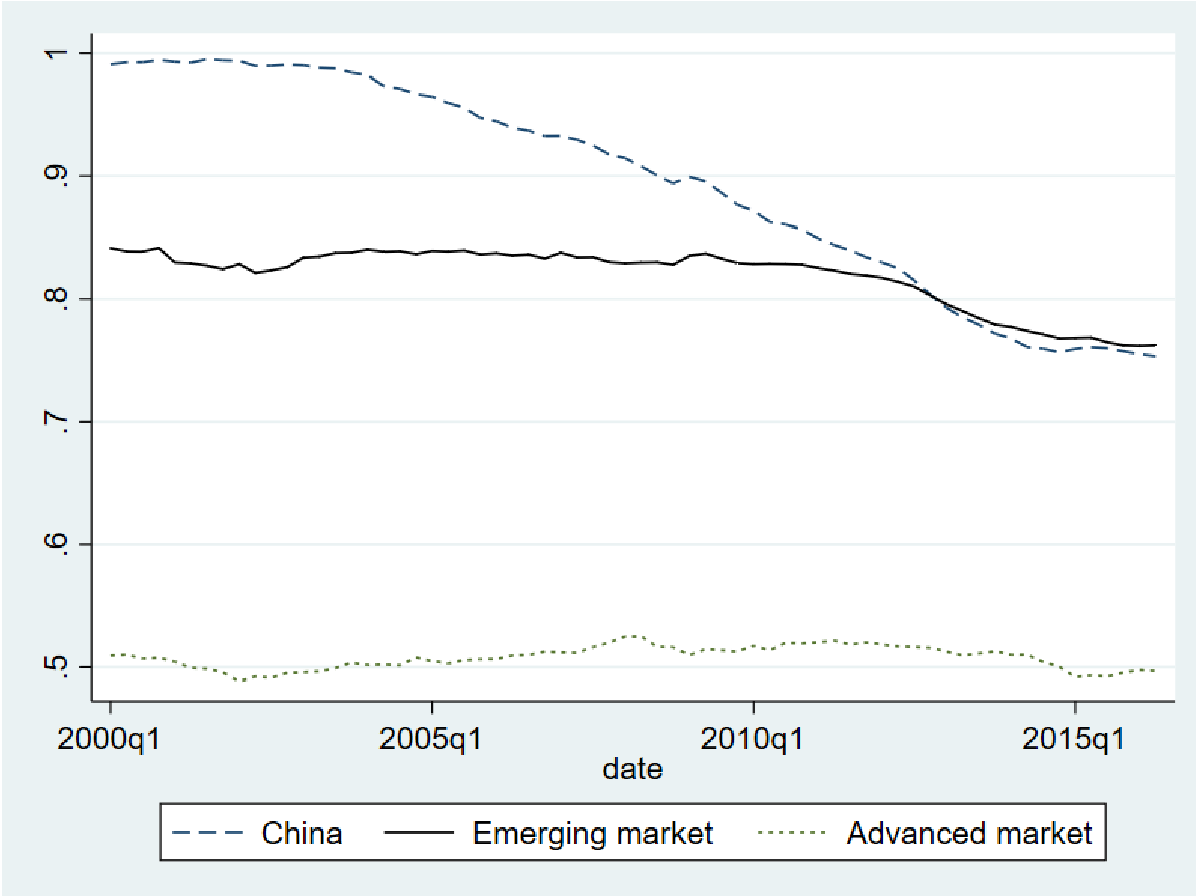


Figure A.5: Share of banking in markets around the world

Source: BIS (Bank for International Settlements).

Note: This figure displays the end-of-quarter outstanding bank loans as a fraction of total credit to the private non-financial sector, in China, in the emerging market economies, and in the advanced market economies.

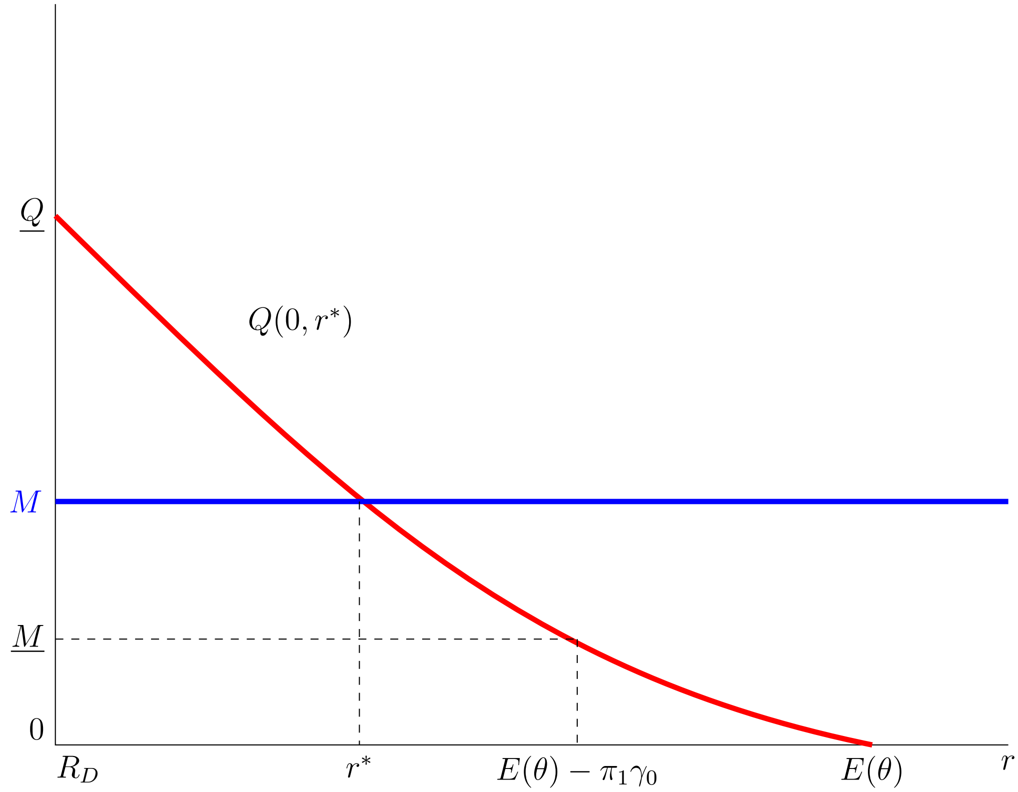


Figure A.6: Equilibrium with $D^* = 0$

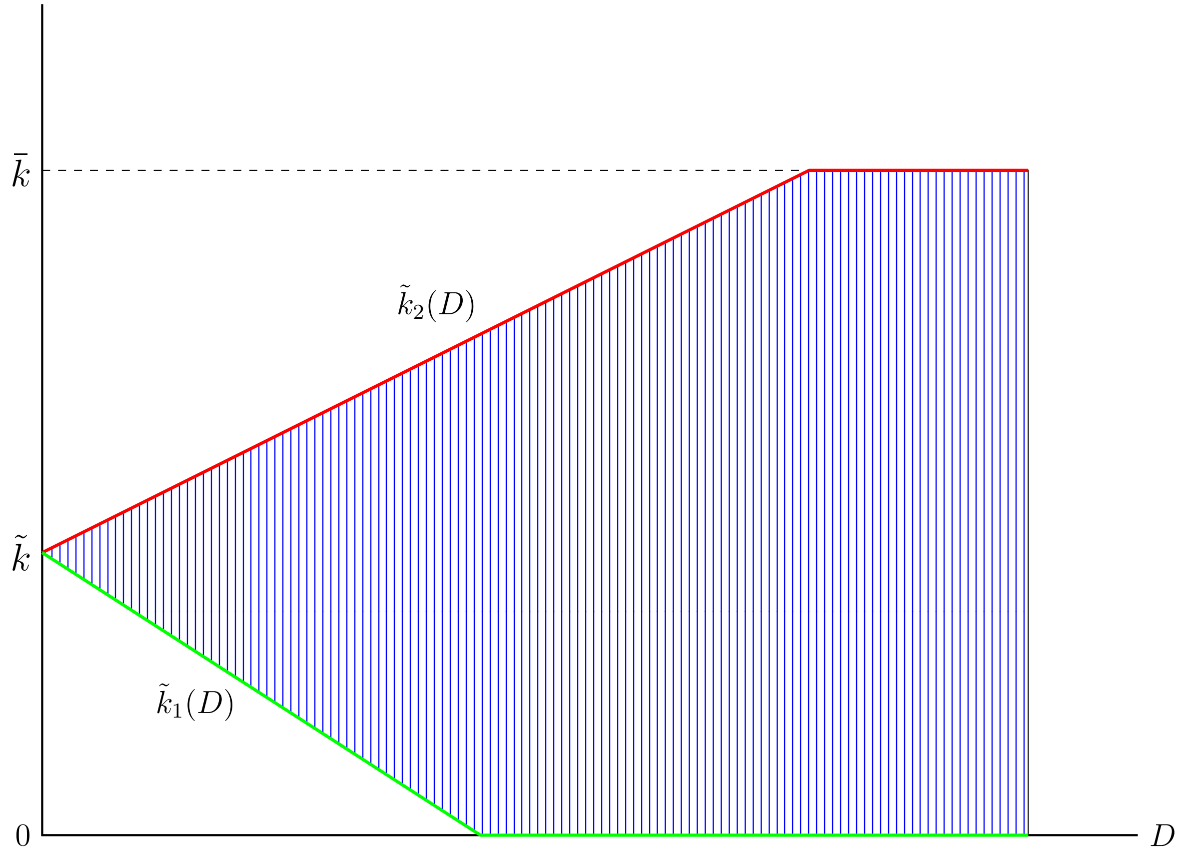


Figure A.7: The bank's optimal lending portfolio **B** as a function of D, after removing the lending rate ceiling

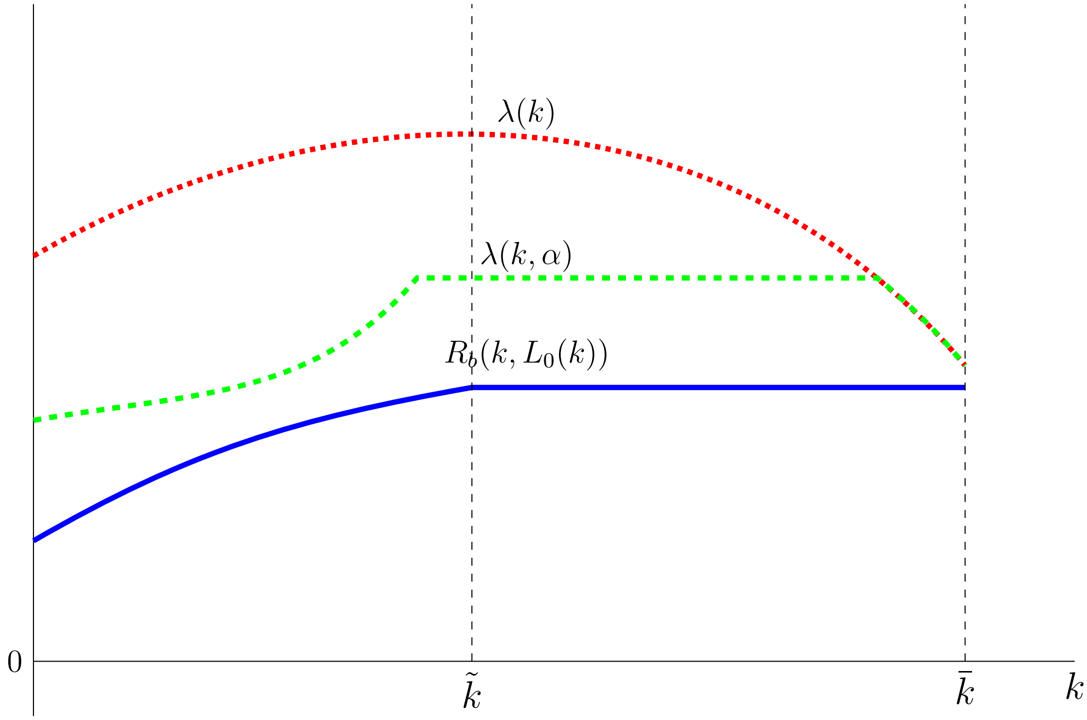


Figure A.8: The bank's expected rate of return on lending to firm k

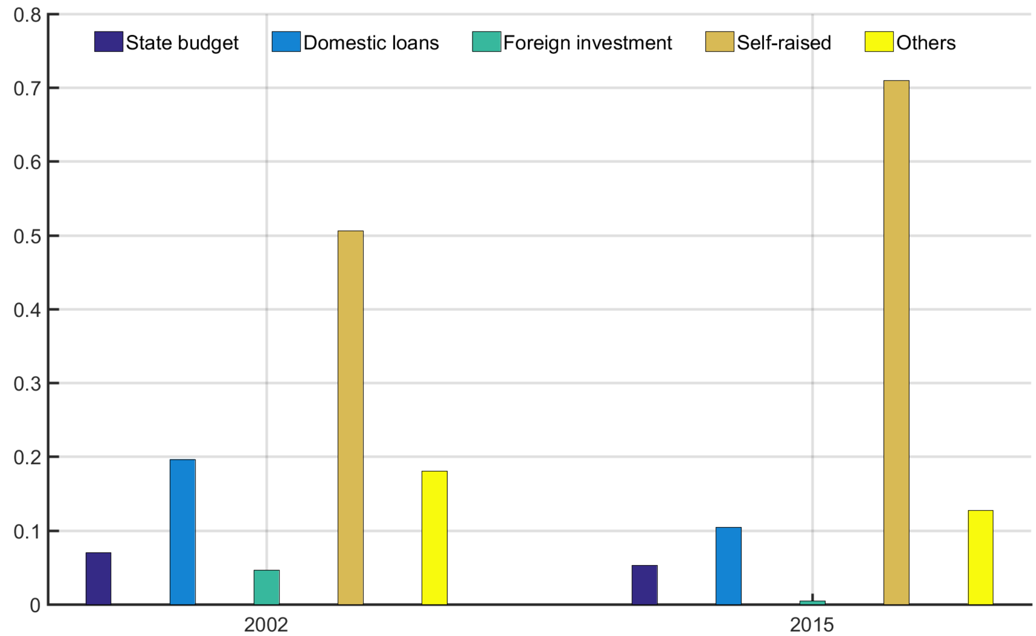


Figure A.9: Source shares of financing in fixed asset investments

Source: CEIC.

Note: (a) “State budget” consists of funds for investment in fixed assets from various state budgets, government bonds at all levels are also included. (b) “Domestic loans” refers to loans of various forms borrowed by investing units from banks and non-bank financial institutions for the purpose of investment in fixed assets. (c) “Foreign investment” refers to overseas funds received, foreign direct investment and other foreign investments. (d) “Self-raised” funds refers to investment in fixed assets using own funds of various enterprises and institutions or funds raised from other units other than financial funds, funds borrowed from financial institutions and overseas funds. Lastly, (e) “Others” refers to funds for investment in fixed assets received from sources other than those listed above, including funds raised from individuals and through donations, and funds transferred from other units. See China statistical yearbook for more details.

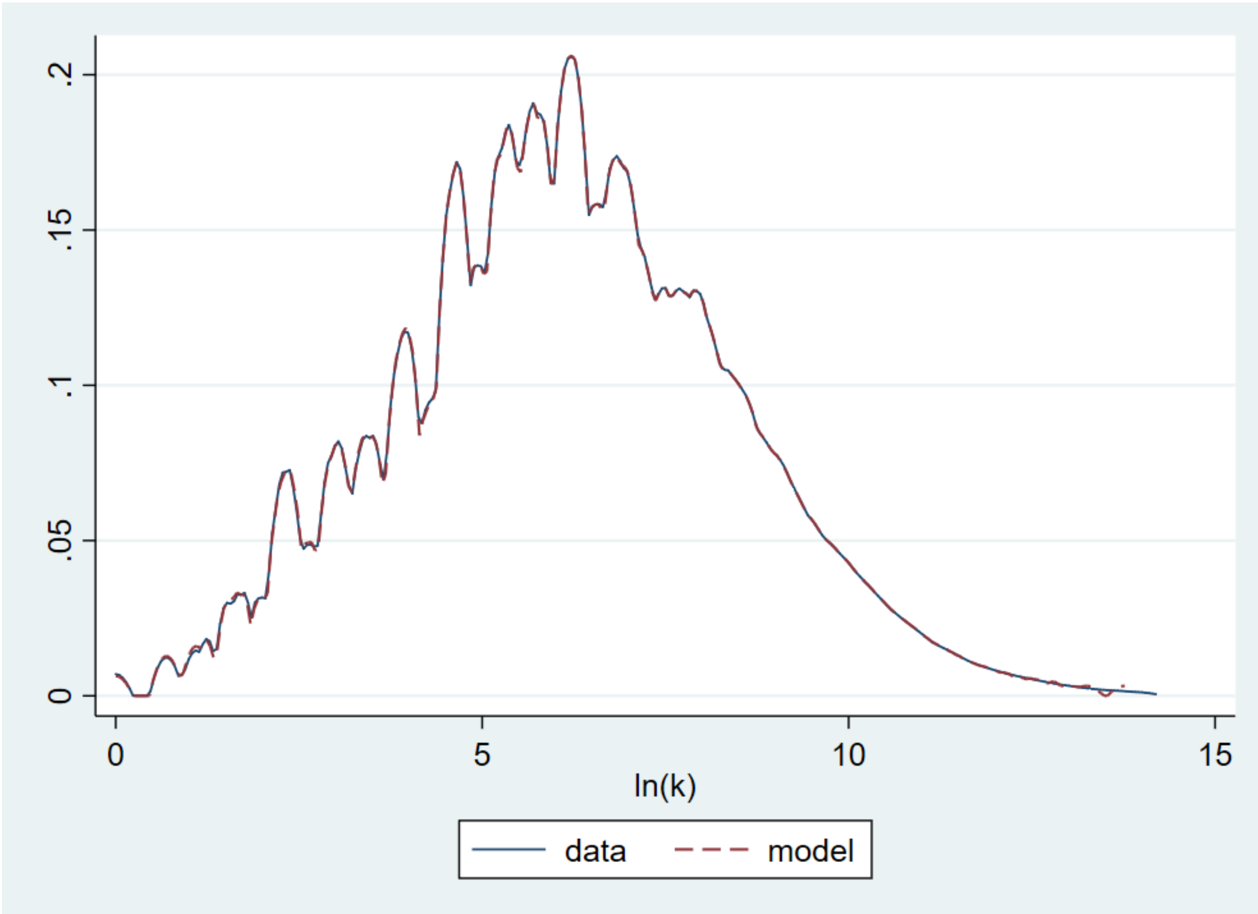
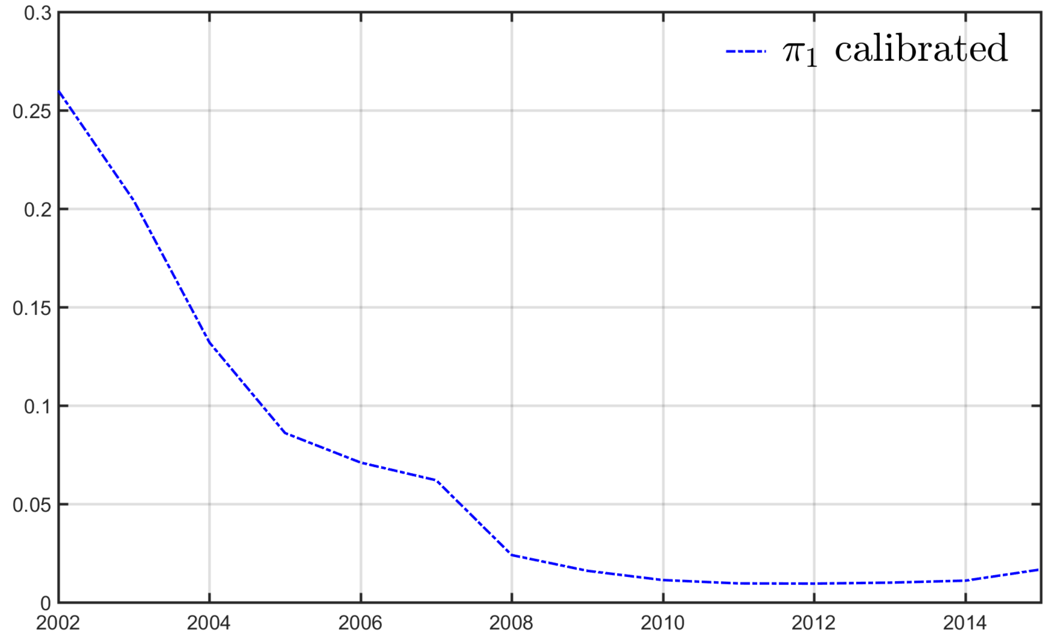


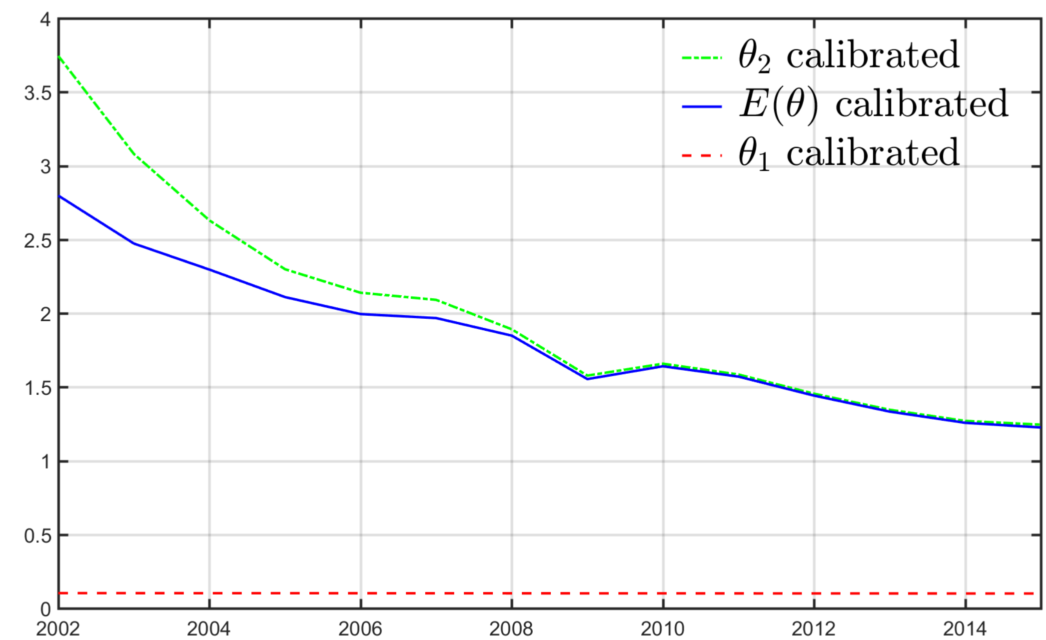
Figure A.10: The distribution of firms in capital

Source: China 2008 Economic Census Data.

Note: "data" denotes the kernel density constructed from the data, and "model" denotes the kernel density used in the calibration. We excluded the largest 0.1% of firms as outliers. The number of firms in the dataset, with a non-missing value in firm fixed assets, is about 2.9 million.



(a)



(b)

Figure A.11: Calibration of π_1, θ_1, θ_2, and $E(\theta)$.

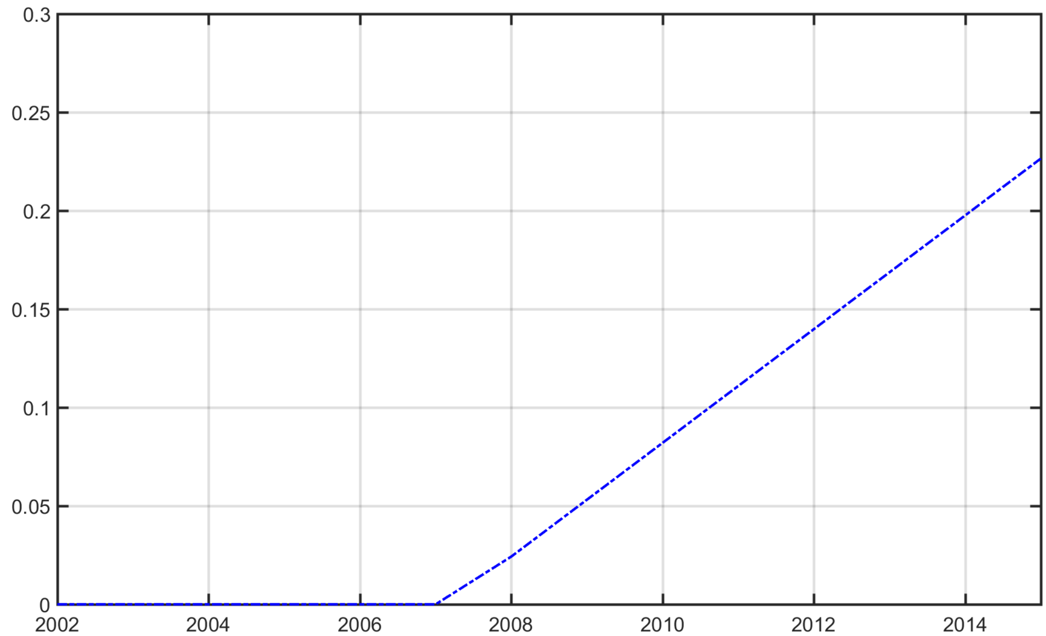


Figure A.12: Share of bond issuance in local government debt balance

Source: National Audit Office's reports and self-calculations.

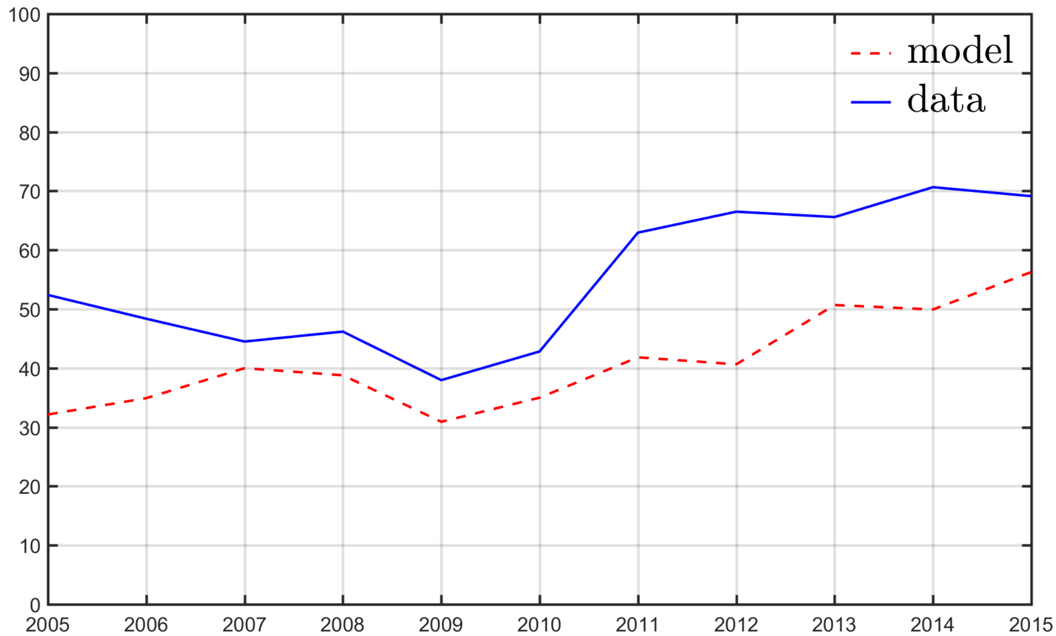


Figure A.13: Share of bank loans with above the policy lending rates

Appendix A.2. Lemma 6 and proof

Lemma 6. *Optimal direct lending has either $S^*(k) = \emptyset$ or $S^*(k) = \{\theta_1\}$.*

Proof. Fix $k \in [0, \bar{k}]$. Suppose the optimal monitoring policy has $S^*(k) = \{\theta_2\}$. Then the remainder of the optimal contract $\{r_1^*, r_2^*, X^*\}$ solves

$$\max_{\{r_1, r_2, X \geq k\}} \left\{ \pi_1 \left[\theta_1 X - r_1(X-k)\right] + \pi_2 \left[\theta_2 X - r_2(X-k) - \widetilde{C}(X-k,k)\right] \right\}$$

subject to

$$0 \leq r_1(X-k) \leq \theta_1 X,$$

$$0 \leq r_2(X-k) \leq \theta_2 X - \widetilde{C}(X-k,k),$$

$$\theta_2 X - r_2(X-k) - \widetilde{C}(X-k,k) \geq \theta_2 X - r_1(X-k), \tag{A.1}$$

$$\pi_1 r_1 + \pi_2 r_2 \geq r^*.$$

Notice that the incentive constraint (A.1) implies $r_2^* < r_1^*$, as $\widetilde{C}(X-k,k) > 0$. With these, consider now an alternative contract $\{S'(k), r_1', r_2', X'\}$, which specifies $S'(k) = \emptyset$, $r_1' = r_2' = \pi_1 r_1^* + \pi_2 r_2^* \leq r_1^*$, and $X' = X^*$. It is straightforward to verify that this new contract satisfies all the constraints on non-monitored direct lending, (3) and (4) specifically, but gives the firm an extra value of $\pi_2 \widetilde{C}(X^* - k, k) > 0$. A contradiction.

Suppose next the optimal contract has $S^*(k) = \{\theta_1, \theta_2\}$. Then the remainder of the optimal contract, $\{r_1^*, r_2^*, X^*\}$, must solve

$$\max_{\{r_1, r_2, X \geq k\}} \left\{ \pi_1 \left[\theta_1 X - r_1(X-k)\right] + \pi_2 \left[\theta_2 X - r_2(X-k)\right] - \widetilde{C}(X-k,k) \right\} \tag{A.2}$$

subject to

$$0 \leq r_1(X-k) \leq \theta_1 X - \widetilde{C}(X-k,k),$$

$$0 \leq r_2(X-k) \leq \theta_2 X - \widetilde{C}(X-k,k),$$

$$\pi_1 r_1 + \pi_2 r_2 \geq r^*.$$

Suppose $r_2^* > r_1^*$. Consider an new contract $\{S'(k), r_1', r_2', X^*\}$ with $S'(k) = \{\theta_1\}$ and

$$r_1' = r_1^*, r_2' = r_2^* - \varepsilon + \frac{\widetilde{C}(X^* - k, k)}{X^* - k},$$

where ε is positive but sufficiently small so that

$$\begin{aligned} r_2' - r_1' &= r_2^* - r_1^* - \varepsilon + \frac{\widetilde{C}(X^* - k, k)}{X^* - k} \\ &> \frac{\widetilde{C}(X^* - k, k)}{X^* - k}. \end{aligned}$$

This new contract satisfies all the constraints on monitored direct lending (6) - (9) and gives the firm an extra value of $\pi_2 \varepsilon (X^* - k) > 0$, which is a contradiction.

Suppose $r_2^* \leq r_1^*$. Then the plan $\{S'(k), r_1', r_2', X^*\}$ with $S'(k) = \emptyset$ and $r_1' = r_2' = \pi_1 r_1^* + \pi_2 r_2^* \leq r_1^* \leq \theta_1$ would satisfy all the constraints on non-monitored direct lending (3) and (4), and give the firm an extra value of $\widetilde{C}(X^* - k, k) > 0$. Again a contradiction. □

Appendix A.3. Proof of Lemma 1

Fix $k \in [0, \bar{k}]$. Notice that the participation constraint is binding: $r_{\mathrm{N}} = r^*$, for otherwise r_{N} can be reduced to make the firm strictly better off. With this, the firm's optimization can be rewritten as

$$\max_X \big\{ (E(\theta) - r^*) X + r^* k \big\}$$

subject to

$$k \leq X \leq \frac{r^* k}{r^* - \theta_1}, \tag{A.3}$$

where (A.3) is from (3). Clearly, the optimal X has $X_N^* = r^* k / (r^* - \theta_1)$. That is, it is optimal to maximize the size of the lending. Substituting the optimal solution into the firm's objective delivers the desired results.

Appendix A.4. Proof of Proposition 2

Let $\Phi \equiv \{k \in [0, \bar{k}] \mid V_{\mathrm{M}}(k) > V_{\mathrm{N}}(k)\}$. This is the set of firms who prefer monitored direct lending to bond finance. Let To prove the proposition we need only show $\Phi = [0, \tilde{k})$ and that for all $k \in \Phi$, equations (12) - (16) hold at the optimum.

Step 1 Fix any $k \in \Phi$. Let the optimal contract conditional on $S(k) = \{\theta_1\}$ be $\{r_1, r_2, X\}$. Notice first that if $X = k$, then $V_{\mathrm{M}}(k) = E(\theta) k \leq V_{\mathrm{N}}(k)$, a contradiction to $k \in \Phi$. Thus the optimal contract must have $X > k$ and so $\widetilde{C}(X - k, k) = C(X - k, X)$.

Observe next that the incentive constraint (8) does not bind. Suppose otherwise, or

$$\theta_1 X - r_1 (X - k) - C(X - k, X) = \theta_1 X - r_2 (X - k).$$

Plugging this into (6) gives

$$r_2(X-k) = r_1(X-k) + C(X-k, X) \leq \theta_1 X,$$

and

$$(\pi_1 r_1 + \pi_2 r_2)(X-k) \leq \theta_1 X.$$

Now consider an alternative plan at k, $\{S'(k), r_1', r_2', X'\}$, with $S'(k) = \emptyset$, $r_1' = r_2' = \pi_1 r_1 + \pi_2 r_2$ and $X' = X$. This plan is feasible (satisfying all the constraints on non-monitored direct lending, (3) and (4) specifically), implying

$$\begin{aligned} V_{\mathrm{N}}(k) &\geq E(\theta)X - (\pi_1 r_1 + \pi_2 r_2)(X-k) \\ &\geq E(\theta)X - (\pi_1 r_1 + \pi_2 r_2)(X-k) - \pi_1 C(X-k, X) \\ &= V_{\mathrm{M}}(k), \end{aligned}$$

contradicting $k \in \Phi$.

Notice also that the participation constraint (9) binds, or $\pi_1 r_1 + \pi_2 r_2 = r^*$. For otherwise r_2 can be reduced to make the firm strictly better off. To see this, remember the incentive constraint (8) does not bind and $r_2 > r_1 \geq 0$. So the plan $\{r_1, r_2', X\}$ with $r_2' = r_2 - \varepsilon$ would satisfy all the constraints on monitored direct lending (6) - (9) (given ε is positive but sufficiently small), and give the firm an extra value of $\pi_2 \varepsilon (X-k) > 0$.

Given the above, the firm's problem is now rewritten as

$$\max_{\{r_1, r_2, X \geq k\}} \{r^* k + (E(\theta) + \pi_1 \gamma k - \pi_1 \gamma_0 - r^*)X - \pi_1 \gamma X^2\}$$

subject to

$$\begin{gathered} 0 \leq r_1(X-k) \leq \theta_1 X - \gamma_0 X - \gamma X(X-k), \\ 0 \leq r_2(X-k) \leq \theta_2 X, \qquad \text{(A.4)} \\ \theta_1(L+k) - r_1 L - \widetilde{C}(L,k) > \theta_1(L+k) - r_2 L, \\ \pi_1 r_1 + \pi_2 r_2 = r^*. \end{gathered}$$

Notice that the objective does not depend on r_1 and r_2 directly.

Step 2 Before moving on to characterize the solution to the above problem, consider the auxiliary problem of maximizing the objective subject only to the constraint $X \geq k$. The solution to this problem has

$$X_{UC}(k) = \begin{cases} [E(\theta) + \pi_1\gamma k - \pi_1\gamma_0 - r^*]/(2\pi_1\gamma), & \text{if } k < k' \\ k, & \text{if } k \geq k' \end{cases},$$

where

$$k' \equiv \frac{E(\theta) - \pi_1\gamma_0 - r^*}{\pi_1\gamma} > 0,$$

and with the firm's value being

$$V_{UC}(k) = \begin{cases} [E(\theta) + \pi_1\gamma k - \pi_1\gamma_0 - r^*]^2/(4\pi_1\gamma) + kr^*, & \text{if } k < k' \\ (E(\theta) - \pi_1\gamma_0)k < E(\theta)k, & \text{if } k \geq k' \end{cases}.$$

Given the above, it can then be shown that there exists a unique $\tilde{k} < k'$ such that

$$V_{UC}(k) - V_{\mathrm{N}}(k) \begin{cases} > 0, & k \in [0, \tilde{k}) \\ = 0, & k = \tilde{k} \\ < 0, & k \in (\tilde{k}, \bar{k}] \end{cases}. \tag{A.5}$$

Specifically, to prove the above, let $f(k) \equiv V_{UC}(k) - V_{\mathrm{N}}(k)$ for all $k \in [0, \bar{k}]$. Notice first that for all $k \in [0, \bar{k}]$,

$$f'(k) < E(\theta) - \pi_2(\theta_2 - \theta_1)\frac{r^*}{r^* - \theta_1} < 0.$$

Notice next that from Lemma 1 and Step 1, $f(0) = V_{UC}(0) - V_{\mathrm{N}}(0) > 0$ and $f(k') = V_{UC}(k') - V_{\mathrm{N}}(k') < 0$. And the desired result follows because $\tilde{k} < k' < \bar{k}$.

Step 3 To prove part *(ii)* of the proposition, we show that under $S(k) = \{\theta_1\}$, the contract with

$$X_{\mathrm{M}}(k) = X_{UC}(k) = \frac{E(\theta) + \pi_1\gamma k - \pi_1\gamma_0 - r^*}{2\pi_1\gamma},$$

$$r_1(k) = \frac{\theta_1 X_{\mathrm{M}}(k) - \gamma_0 X_{\mathrm{M}}(k) - (X_{\mathrm{M}}(k) - k)\gamma X_{\mathrm{M}}(k)}{X_{\mathrm{M}}(k) - k}, \tag{A.6}$$

and

$$r_2(k) = \frac{r^* - \pi_1 r_1(k)}{\pi_2} \tag{A.7}$$

is optimal conditional for all $k \in [0, \tilde{k})$, and so $V_{\mathrm{M}}(k) = V_{UC}(k)$ for all $k \in [0, \tilde{k}]$.

From Step 1, the above specified contract attains the "unconstrained" value $V_{UC}(k)$, so to show that it is optimal we need only show that it is feasible (i.e., satisfying (6) - (9)).

First, equation (A.5) implies $k \in \Phi \subseteq [0, \tilde{k})$, and so $k < \tilde{k} < k'$ and hence $X_{\mathrm{M}}(k) > k$.

Second, from (A.7), the participation constraint (9) is satisfied.

Third, notice that given

$$V_{UC}(k) = \pi_2[\theta_2 X_{\mathrm{M}}(k) - r_2(k)(X_{\mathrm{M}}(k) - k)] > V_{\mathrm{N}}(k) > 0,$$

constraint (7) is satisfied.

Fourth, from Assumption 1 we have

$$\begin{aligned}\theta_1 - \gamma_0 - (X_{\mathrm{M}}(k) - k)\gamma &= \theta_1 - \frac{E(\theta) - \pi_1\gamma k + \pi_1\gamma_0 - r^*}{2\pi_1} \\ &> \frac{r^* - (\pi_2\theta_2 - \pi_1\theta_1 + \pi_1\gamma_0)}{2\pi_1} \\ &> 0,\end{aligned}$$

and so $r_1(k) > 0$ which, together with (A.6), implies that constraint (6) is satisfied.

Fifth, we show the incentive constraint (8) is also satisfied. Given $V_{UC}(k) > V_{\mathrm{N}}(k)$, we have

$$E(\theta)X_{\mathrm{M}}(k) - r^*(k)(X_{\mathrm{M}}(k) - k) > E(\theta)X_{\mathrm{N}}(k) - r^*(k)(X_{\mathrm{N}}(k) - k)$$

which implies

$$X_{\mathrm{M}}(k) > X_{\mathrm{N}}(k).$$

This, together with Lemma 1 where $r^*(X_{\mathrm{N}}(k) - k) = \theta_1 X_{\mathrm{N}}(k)$, gives

$$r^*(X_{\mathrm{M}}(k) - k) > \theta_1 X_{\mathrm{M}}(k),$$

which, together with (A.6) and (A.7), gives $r_2(X_{\mathrm{M}}(k) - k) > \theta_1 X_{\mathrm{M}}(k)$ and that (8) is satisfied. This proves part (ii) of the proposition.

Step 4 The above step, together with equation (A.5), which implies $\Phi \subseteq [0, \tilde{k})$, gives $\Phi = [0, \tilde{k})$, or part (i) of the proposition. This concludes the proof of Proposition 2.

Appendix A.5. Corollary 7 and proof

Corollary 7. *With the optimal contract, the firm's gross rate of return on equity, $V(k)/k$ is strictly decreasing in k for $k \in [0, \tilde{k}]$ and constant in k for $k \in (\tilde{k}, \bar{k}]$.*

Proof. From (17) and (18), for all $k \in [\tilde{k}, \bar{k}]$ we have

$$\frac{V(k)}{k} = \frac{\pi_2(\theta_2 - \theta_1)r^*}{r^* - \theta_1}$$

and

$$\frac{X(k)}{k} = \frac{r^*}{r^* - \theta_1},$$

both constant in k. Next, for all $k \in [0, \tilde{k})$, we have

$$\frac{V(k)}{k} = \frac{\left[(E(\theta) - r^* - \pi_1\gamma_0)/\sqrt{k} + \pi_1\gamma\sqrt{k}\right]^2}{4\pi_1\gamma} + r^*$$

and

$$\frac{X(k)}{k} = \frac{(E(\theta) - r^* - \pi_1\gamma_0)/k + \pi_1\gamma}{2\pi_1\gamma},$$

both strictly decreasing in k for $k \leq \tilde{k} < k'$. □

Appendix A.6. Costly participation in the bond market

Throughout the paper, it is assumed that all firms are free to raise funds in the bond market, and there are no fixed or variable costs associated with bond issuance. Suppose participating in the bond market is costly, and the cost is a constant $c_b(> 0)$ for any firm who wishes to raise any amount of finance. Then for all $k \in [\tilde{k}, \bar{k}]$,

$$V(k) = \frac{\pi_2(\theta_2 - \theta_1)r^*}{r^* - \theta_1}k - c_b,$$

and

$$\frac{V(k)}{k} = \frac{\pi_2(\theta_2 - \theta_1)r^*}{r^* - \theta_1} - \frac{c_b}{k},$$

which is strictly increasing in k for $k \in [\tilde{k}, \bar{k}]$. This, in turn, implies that the bank's rate of return on lending is strictly decreasing in firm size k, that is,

$$\frac{dR_b(k, L_0(k))}{dk} < 0, \ k \in [\tilde{k}, \bar{k}].$$

It then follows that the bank would prefer smaller firms over larger firms in the interval $[\tilde{k}, \bar{k}]$.

Appendix A.7. Corollary 8, proof and intuition

Corollary 8. *The optimal contract for monitored direct lending has: for all $k \in [0, \tilde{k})$, $r_1(k) < r^* < r_2(k)$ and $r_1'(k) > 0$, $r_2'(k) < 0$.*

Proof. The participation constraint (9) binds for all $k \in [0, \tilde{k}]$. The incentive constraint (8) gives $r_1(k) < r_2(k)$. Combining these gives $r_1(k) < r^* < r_2(k)$ for all $k \in [0, \tilde{k})$. Next, from Proposition 2,

$$\begin{aligned}
r_1'(k) &= \frac{(\theta_1 - \gamma_0)(X_{\mathrm{M}}(k) - 1/2k)}{(X_{\mathrm{M}}(k) - k)^2} - \frac{1}{2}\gamma \\
&= \frac{2\pi_1\gamma(\theta_1 - \gamma_0)(E(\theta) - \pi_1\gamma_0 - r^*)}{(E(\theta) - \pi_1\gamma k - \pi_1\gamma_0 - r^*)^2} - \frac{1}{2}\gamma \\
&\geq \frac{2\pi_1\gamma(\theta_1 - \gamma_0)}{E(\theta) - \pi_1\gamma_0 - r^*} - \frac{1}{2}\gamma \\
&= \frac{\gamma}{2[E(\theta) - \pi_1\gamma_0 - r^*]}(2\pi_1\theta_1 - 2\pi_1\gamma_0 + \pi_1\theta_1 + r^* - \pi_2\theta_2 - \pi_1\gamma_0) \\
&> 0,
\end{aligned}$$

where the last inequality is from Assumption 1. Lastly,

$$r_2'(k) = -\frac{\pi_1}{\pi_2}r_1'(k) < 0. \qquad \square$$

The intuition for the above proof is as follows. As k increases, $r_1(k)$ increases, as a larger firm net worth allows the contract to pay the investor more in the state of low output. How a larger k would affect $r_2(k)$ is less obvious. From equation (15), a larger k affects the sign of $r_2'(k)$ in two ways. A larger k allows the investor be paid more in the state of low output, this lowers $r_2(k)$. A larger k also implies a larger project and a larger total and per-unit-of-investment cost of monitoring, which must be compensated by a larger $r_1(k)$, as well as a larger $r_2(k)$.

Appendix A.8. Corollary 9 and proof

Corollary 9. *With the optimal direct lending contract, $X_{\mathrm{M}}(k) > X_{\mathrm{N}}(k)$, for all $k \in [0, \tilde{k})$.*

Proof. It follows from Lemma 1 and Proposition 2 that for all $k \in [0, \tilde{k})$, $V_{\mathrm{M}}(k) > V_{\mathrm{N}}(k)$, or

$$\begin{aligned}
& E(\theta)X_{\mathrm{M}}(k) - r^*(X_{\mathrm{M}}(k) - k) - \pi_1\gamma_0 X_{\mathrm{M}}(k) - \pi_1\gamma X_{\mathrm{M}}(k)(X_{\mathrm{M}}(k) - k) \\
> \; & E(\theta)X_{\mathrm{N}}(k) - r^*(X_{\mathrm{N}}(k) - k),
\end{aligned}$$

which in turn gives $X_{\mathrm{M}}(k) > X_{\mathrm{N}}(k)$. $\square$

Appendix A.9. Proof for Proposition 3

The proof is carried out in 5 steps, using an approach developed in Wang and Williamson (1998) for optimally determining a set as a choice variable.

Step 1 We show that the budget constraint (21) binds. Suppose at the optimum

$$\mu \int_{\mathbf{B}} [Z(k) - k]\, dG(k) < D.$$

Rewriting the bank's net profits as

$$\mu \int_{\mathbf{B}} \left\{ \pi_1 (\theta_1 - \gamma_0) + \pi_2 R_L - 1 \right\} [Z(k) - k]\, dG(k) + \mu \int_{\mathbf{B}} \pi_1 (\theta_1 - \gamma_0)\, k dG(k) - (R_D - 1)D.$$

By Assumption 2 we have $\pi_2 R_L + \pi_1 (\theta_1 - \gamma_0) - 1 > 0$. Then $Z(k)$ can be increased for a positive measure of $k \in \mathbf{B}$ to make the bank strictly better off. A contradiction.

Step 2 Let $L(k) = Z(k) - k$ for all $k \in \mathbf{B}$, the optimization problem can be written as

$$\max_{\mathbf{B}; L(k), k \in \mathbf{B}} C_1 \int_{\mathbf{B}} k dG(k) + C_2 \tag{A.8}$$

subject to

$$\mathbf{B} \subseteq [0, \bar{k}],$$

$$L(k) \geq 0,\ \forall k \in \mathbf{B},$$

$$\mu \int_{\mathbf{B}} L(k) dG(k) = D, \tag{A.9}$$

$$\pi_2 \theta_2 k + \pi_2 (\theta_2 - R_L) L(k) \geq V(k),\ \forall k \in \mathbf{B}, \tag{A.10}$$

where $C_1 \equiv \pi_1 (\theta_1 - \gamma_0)\, \mu$, and $C_2 \equiv [\pi_1 (\theta_1 - \gamma_0) + \pi_2 R_L - R_D]\, D$.

Step 3 We show that either the participation constraint (A.10) binds for all $k \in \mathbf{B}$, or the bank provides loans to all firms, $\mathbf{B} = [0, \bar{k}]$. Suppose not. Suppose the bank's optimal plan is $\{Z(k) : k \in \mathbf{B}\}$ and $\mathbf{B} \subset [0, \bar{k}]$, and suppose a subset $H \subseteq \mathbf{B}$ of the firms get higher values than their reservation values through bank loans, where $H \neq \emptyset$. From (22), $Z(k) - k > L_0(k), \forall k \in H$. Then suppose the bank lends $L_0(k)$ units of funds to the firms $k \in H$ instead, and lends the extra funds $\int_H Z(k) - k - L_0(k) dG(k)$ to a set of firms $F \subseteq [0, \bar{k}] \setminus \mathbf{B}$ with size of loans $\{L_0(k) : k \in F\}$ such that

$$\mu \int_F L_0(k) dG(k) = \mu \int_H [Z(k) - k - L_0(k)]\, dG(k).$$

This way, the bank would get a strictly positive extra value which, specifically, equals

$$
\begin{aligned}
& \mu \int_{F\cup H} \left[\pi_1(\theta_1-\gamma_0)(k+L_0(k)) + \pi_2 R_L L_0(k)\right] dG(k) \\
& \qquad\qquad\qquad -\mu \int_H \left[\pi_1(\theta_1-\gamma_0)Z(k) + \pi_2 R_L(Z(k)-k)\right] dG(k) \\
= \; & \mu \int_F \pi_1(\theta_1-\gamma_0)\left[k+L_0(k)\right] dG(k) - \mu \int_H \pi_1(\theta_1-\gamma_0)\left[Z(k)-k-L_0(k)\right] dG(k) \\
= \; & \mu\pi_1(\theta_1-\gamma_0)\left\{\int_F k dG(k) + \int_F L_0(k) dG(k) - \int_H \left[Z(k)-k-L_0(k)\right] dG(k)\right\} \\
= \; & \mu\pi_1(\theta_1-\gamma_0)\int_F k dG(k),
\end{aligned}
$$

which is strictly positive given $\theta_1 > \gamma_0$. A contradiction.

Step 4 Consider the case where $D \geq \bar{D}$. Suppose $\mathbf{B} \subset [0, \bar{k}]$. From **Step 3**, the participation constraint (A.10) binds for all $k \in \mathbf{B}$. So

$$L(k) = \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\theta_2 - R_L)}.$$

From equation (A.9) we have

$$D = \mu \int_{\mathbf{B}} \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\theta_2 - R_L)} dG(k) < \mu \int_0^{\bar{k}} \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\theta_2 - R_L)} dG(k) = \bar{D}.$$

A contradiction. So when $D \geq \bar{D}$, we have $\mathbf{B} = [0, \bar{k}]$.

Now the total net worth of the firms $\mu \int_{\mathbf{B}} k dG(k)$ is constant. From (A.8) we know any feasible allocation is optimal. Thus any contract $\{\mathbf{B} = [0, \bar{k}]; L(k), k \in \mathbf{B}\}$ is feasible and optimal when

$$L(k) \geq \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\theta_2 - R_L)} \quad \forall k \in [0, \bar{k}],$$

and

$$\mu \int_0^{\bar{k}} L(k) dG(k) = D.$$

This proves part (iii) of the proposition.

Step 5 Consider the case where $D < \bar{D}$. From (A.9) and (A.10) we have

$$D = \mu \int_{\mathbf{B}} L(k) dG(k) \geq \mu \int_{\mathbf{B}} \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\theta_2 - R_L)} dG(k).$$

Thus $\mathbf{B} \subset [0, \bar{k}]$, which implies resource constraint (A.10) binds for all $k \in \mathbf{B}$. So

$$L(k) = \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\theta_2 - R_L)} \quad \forall k \in \mathbf{B},$$

or

$$Z(k) = \frac{V(k) - \pi_2 R_L k}{\pi_2(\theta_2 - R_L)} \quad \forall k \in \mathbf{B}.$$

Now the optimal $\mathbf{B}$ solves the problem

$$\max_{\mathbf{B} \subseteq [0,\bar{k}]} \int_{\mathbf{B}} k dG(k) \tag{A.11}$$

subject to

$$\int_{\mathbf{B}} L(k) dG(k) = \frac{D}{\mu}, \tag{A.12}$$

where

$$L(k) = \frac{V(k) - \pi_2 \theta_2 k}{\pi_2(\theta_2 - R_L)}.$$

Let λ be the Lagrange multiplier of the constraint (A.12). The Lagrangian for the above problem is

$$L = \int_{\mathbf{B}} [k - \lambda L(k)]\, dG(k) + \frac{\lambda D}{\mu}.$$

Thus L is maximized when $\mathbf{B}$ includes all the ks that

$$\frac{k}{L(k)} > \lambda,$$

and part or all of the ks that

$$\frac{k}{L(k)} = \lambda.$$

By Corollary 7,

$$\frac{k}{L(k)} = \frac{\pi_2(\theta_2 - R_L)}{V(k)/k - \pi_2 \theta_2}$$

is strictly increasing with k for $k \in [0, \tilde{k}]$ and constant for $k \in [\tilde{k}, \bar{k}]$.

So $\mathbf{B} = [\hat{k}, \bar{k}]$ when $D \in [D_0, \bar{D})$, where $\hat{k}$ satisfies

$$\mu \int_{\hat{k}}^{\bar{k}} L(k) dG(k) = D,$$

and $\mathbf{B} \subset [\tilde{k}, \bar{k}]$ when $D \in (0, D_0)$. Parts (i) and (ii) of the proposition are now proved.

Appendix A.10. Lemma 10 and proof

Lemma 10. *$L_0(k)$ and $Z_0(k)$ are increasing in k.*

Proof. From (18),

$$V'(k) = \begin{cases} V'_{\mathrm{M}}(k) = [E(\theta) + \pi_1\gamma k - r^* - \pi_1\gamma_0]/2 + r^*, & \forall k < \tilde{k} \\ V'_{\mathrm{N}}(k) = \pi_2(\theta_2 - \theta_1)r^*/(r^* - \theta_1), & \forall k \geq \tilde{k} \end{cases},$$

which is increasing in k for $k \in [0, \tilde{k})$ and constant in k for $k \in [\tilde{k}, \bar{k}]$. Moreover,

$$V'(k) \geq V'(0) = r^* + [E(\theta) - r^* - \pi_1\gamma_0]/2, \forall k < \tilde{k}.$$

These, with Assumption 1, then imply $V'(k) > \pi_2\theta_2 > \pi_2 R_L$ for all $k \in [0, \bar{k}]$, and hence

$$Z'_0(k) = \frac{V'(k) - \pi_2 R_L}{\pi_2(\theta_2 - R_L)} > 0, \forall k \in [0, \bar{k}],$$

and

$$L'_0(k) = \frac{V'(k) - \pi_2\theta_2}{\pi_2(\theta_2 - R_L)} > 0, \forall k \in [0, \bar{k}].$$ □

Appendix A.11. Lemma 11 and proof, and Figure A.14

Lemma 11. *Let $\hat{R}_D \equiv \pi_1\theta_1 + 2\pi_2 R_L - \pi_2\theta_2 - \pi_1\gamma_0$. Then the optimal direct and bank lending contracts have*

(i) If $R_D < \hat{R}_D$, then $Z_0(k) > X(k)$ for all $k \in [0, \bar{k}]$.

(ii) If $R_D \geq \hat{R}_D$, then $Z_0(k) < X(k)$ for all $k \in [0, k^*)$ and $Z_0(k) > X(k)$ for all $k \in (k^*, \bar{k}]$, where $k^* \in (0, \tilde{k})$ and solves $Z_0(k^*) = X(k^*)$.

Proof. Notice from equation (25) that $Z_0(k)$ satisfies

$$\pi_2\left\{\theta_2 Z_0(k) - R_L\left[Z_0(k) - k\right]\right\} = V(k), \quad \forall k \in [0, \bar{k}], \tag{A.13}$$

and from Lemma 1 and Proposition 2 it holds that

$$\pi_2\left[\theta_2 X(k) - r_2(k)(X(k) - k)\right] = V(k), \quad \forall k \in [0, \bar{k}], \tag{A.14}$$

where $r_1(k) = r_2(k) = r^* = R_D$, $\forall k \in [\tilde{k}, \bar{k}]$, and remember that for all $k \in [0, \tilde{k})$, $r_1(k)$ and $r_2(k)$ were defined in Proposition 2. Next, following from (A.13) and (A.14) it holds that

$$Z_0(k) \geq X(k) \Leftrightarrow R_L \geq r_2(k). \tag{A.15}$$

And lastly, from Proposition 2 and Corollary 8, $r_2(k)$ is decreasing in k for all $k \in [0, \tilde{k})$ and

$$r_2(0) = \frac{R_D + \pi_2\theta_2 - \pi_1\theta_1 + \pi_1\gamma_0}{2\pi_2}.$$

Given the above, suppose $R_D < \hat{R}_D = \pi_1\theta_1 + 2\pi_2 R_L - \pi_2\theta_2 - \pi_1\gamma_0$. Then

$$R_L > \frac{R_D + \pi_2\theta_2 - \pi_1\theta_1 + \pi_1\gamma_0}{2\pi_2} = r_2(0) > r_2(k),$$

which, together (A.15), gives $Z_0(k) > X(k)$ for all $k \in [0, \bar{k}]$, and so (i) holds.

Suppose $R_D \geq \hat{R}_D$. Then

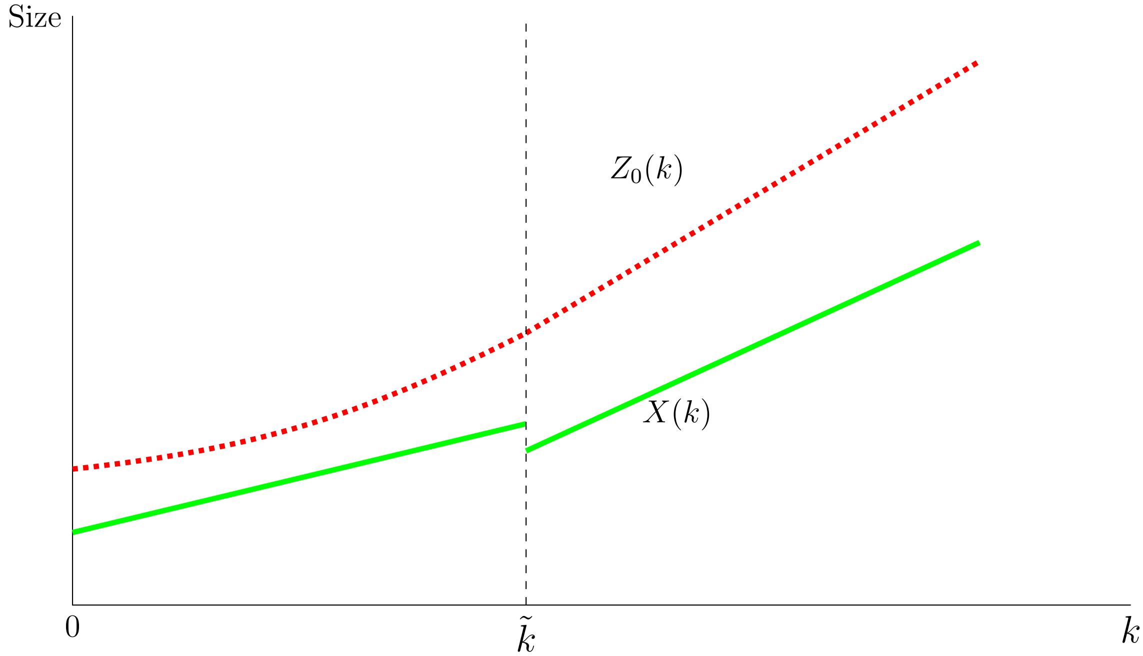
$$r_2(\tilde{k}) = R_D < R_L < r_2(0),$$

and so there exists some $k^* \in [0, \tilde{k})$ such that

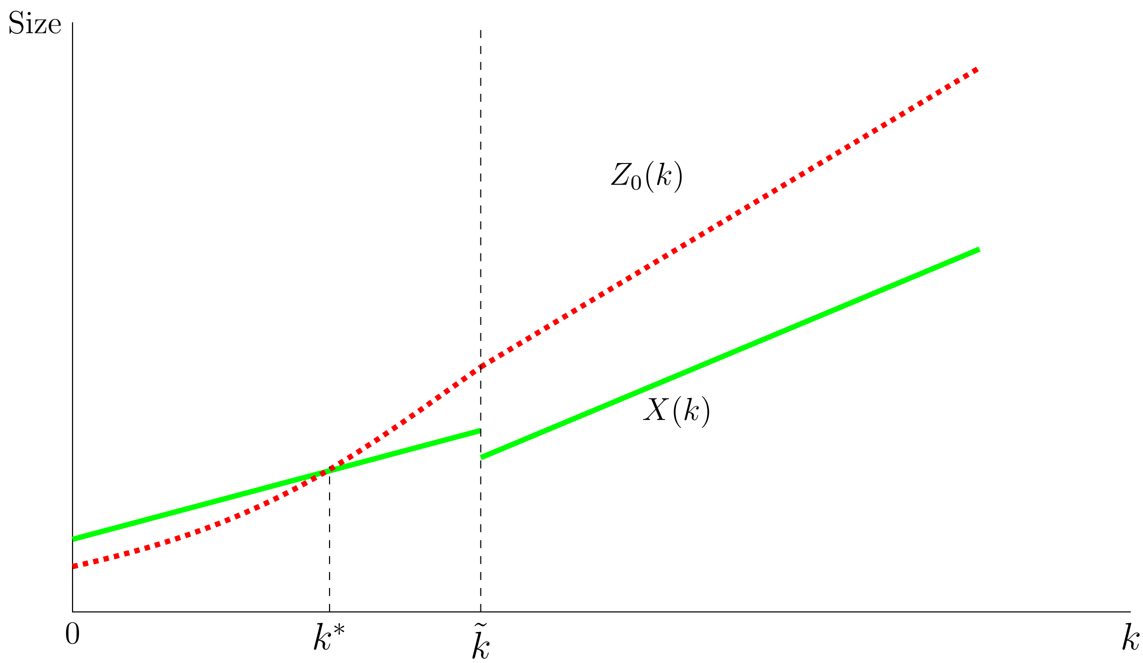
$$R_L \begin{cases} < r_2(k), & \text{if} \quad k \in [0, k^*) \\ = r_2(k), & \text{if} \quad k = k^* \\ > r_2(k), & \text{if} \quad k \in (k^*, \bar{k}] \end{cases}.$$

This, given (A.15), proves part (ii) of the lemma. □

Figure A.14 depicts what Lemma 11 states for the cases $R_D < \hat{R}_D$ and $R_D \geq \hat{R}_D$, respectively.



(a) Case i: $R_D < \hat{R}_D$



(b) Case ii: $R_D \geq \hat{R}_D$

Figure A.14: The optimal size of the project: direct and bank lending

Appendix A.12. Formulating the equilibrium in a system of equations

Following from Definition 1, an equilibrium of the model is characterized by a tuple

$$\Big\{(r^*, D^*);\ (\tilde{k}, V(k), X(k)) : k \in [0, \bar{k}];\ (\mathbf{B}, Z(k) : k \in \mathbf{B})\Big\}$$

that solves the following system of equations:

(I) (r^*, D^*) satisfies:

$$r^* \geq R_D, \text{ and } D^* = 0 \text{ if } r^* > R_D.$$

(II) $\tilde{k}$ solves

$$\frac{\Big(E(\theta) + \pi_1\gamma\tilde{k} - r^* - \pi_1\gamma_0\Big)^2}{4\pi_1\gamma} + \tilde{k}r^* = \pi_2(\theta_2 - \theta_1)\frac{\tilde{k}r^*}{r^* - \theta_1},$$

and $X(k)$ and $V(k)$ are given by

$$X(k) = \begin{cases} X_{\mathrm{M}}(k) = [E(\theta) + \pi_1\gamma k - r^* - \pi_1\gamma_0]/(2\pi_1\gamma), & \forall k < \tilde{k} \\ X_{\mathrm{N}}(k) = kr^*/(r^* - \theta_1), & \forall k \geq \tilde{k} \end{cases}$$

and

$$V(k) = \begin{cases} V_{\mathrm{M}}(k) = [E(\theta) + \pi_1\gamma k - r^* - \pi_1\gamma_0]^2/(4\pi_1\gamma) + kr^*, & \forall k < \tilde{k} \\ V_{\mathrm{N}}(k) = \pi_2(\theta_2 - \theta_1)kr^*/(r^* - \theta_1), & \forall k \geq \tilde{k} \end{cases}.$$

(III) The set of firms to receive bank lending $\mathbf{B}$ and the size of the project that receives bank finance $Z(k)$ satisfy:

(a) $\mathbf{B} = [0, \bar{k}]$ if $D^* \geq \bar{D}$. In this case,

$$Z(k) \geq \frac{V(k) - \pi_2 R_L k}{\pi_2(\theta_2 - R_L)},\ \forall k \in [0, \bar{k}],$$

and

$$\mu\int_0^{\bar{k}} Z(k)dG(k) = D^* + \mu\int_0^{\bar{k}} kdG(k).$$

(b) $\mathbf{B} = [\hat{k}, \bar{k}]$ where $\hat{k} \in (0, \bar{k})$ if $D^* \in (D_0, \bar{D})$. In this case,

$$Z(k) = \frac{V(k) - \pi_2 R_L k}{\pi_2(\theta_2 - R_L)}, \quad \forall k \in [\hat{k}, \bar{k}],$$

and

$$D^* = \mu \int_{\hat{k}}^{\bar{k}} \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\theta_2 - R_L)} dG(k).$$

(c) $\mathbf{B} \subset [\tilde{k}, \bar{k}]$ if $0 \leq D^* < D_0$. In this case,

$$Z(k) = \frac{V(k) - \pi_2 R_L k}{\pi_2(\theta_2 - R_L)}, \ \forall k \in \mathbf{B},$$

and

$$\mu \int_{\mathbf{B}} \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\theta_2 - R_L)} dG(k) = D^*.$$

(IV) The market for finance clears:

$$\mu \int_{k \in [0,\bar{k}] \setminus \mathbf{B}} [X(k) - k] dG(k) + \mu \int_{k \in \mathbf{B}} [Z(k) - k] dG(k) = M.$$

Appendix A.13. Lemma 12 and proof

In this section we take $r^* = R_D$ as given.

Lemma 12. *With the optimal contracts, $Q(D)$ is strictly increasing in D at all $D \in (0, M]$.*

Proof. Given $R_D < \hat{R}_D$, it follows from Lemma 11 that $Z(k) \geq Z_0(k) > X(k), \forall k \in [0, \bar{k}]$. And from Proposition 3, $\hat{k}_1(D)$ is weakly decreasing in D and $\hat{k}_2(D)$ weakly increasing in D. Given equation (30), the lemma is proved. □

Suppose $R_D \geq \hat{R}_D$. Then $Q(D)$ is decreasing in D over the interval $[D_1', \bar{D}]$, where $D_1' = \mu \int_{k^*}^{\bar{k}} L_B(k) dG(k)$ and k^* is defined in part (ii) of Lemma 11. This case is depicted in Figure A.15 below.

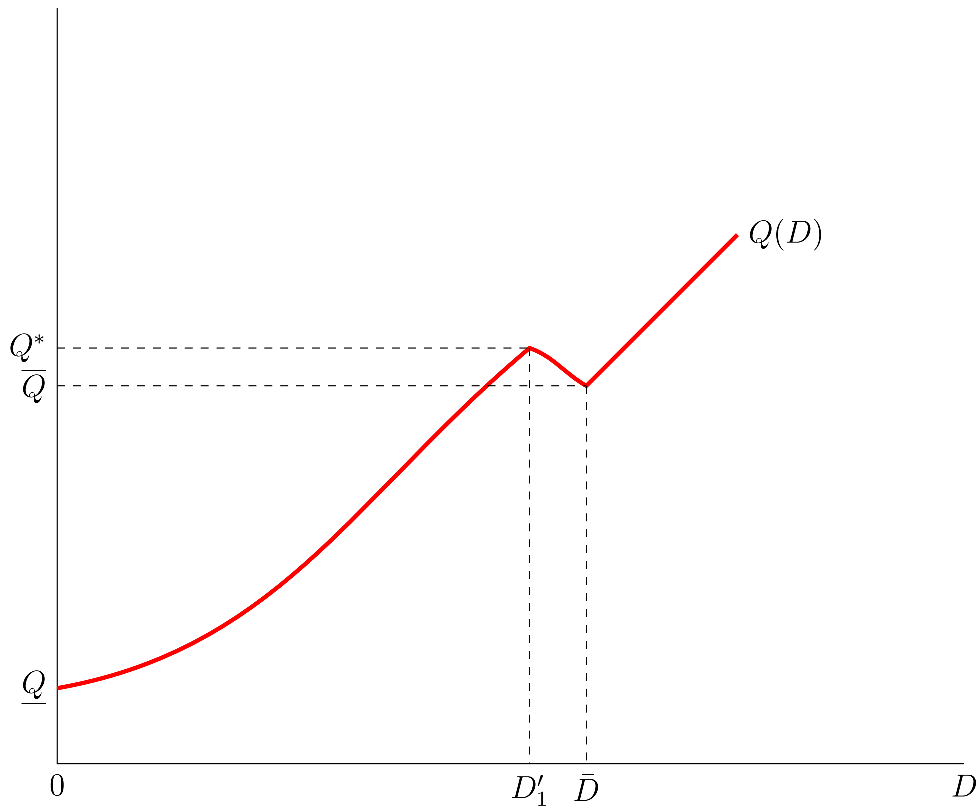


Figure A.15: The demand function under $R_D \geq \hat{R}_D$

Appendix A.14. Banking reforms: a full treatment

Appendix A.14.1. Removing the lending rate ceiling

With the 2004 reform to remove the lending rate ceiling, the bank's optimization becomes

$$\max_{\mathbf{B},\{Z(k),R_L(k)\}_{k\in\mathbf{B}}} \mu \int_{\mathbf{B}} \Big\{ \pi_1 (\theta_1 - \gamma_0) Z(k) + \pi_2 R_L(k) [Z(k) - k] \Big\} dG(k)$$
$$+ D - \mu \int_{\mathbf{B}} [Z(k) - k] \, dG(k) - R_D D \tag{A.16}$$

subject to (19), (21) and

$$k \leq Z(k) \leq \bar{X}, \quad \forall k \in \mathbf{B}, \tag{A.17}$$

$$R_L(k) \geq \underline{R}_L, \quad \forall k \in \mathbf{B}, \tag{A.18}$$

$$\pi_2 \left\{ \theta_2 Z(k) - R_L(k) [Z(k) - k] \right\} \geq V(k), \quad \forall k \in \mathbf{B}. \tag{A.19}$$

As in the case where the lending rate is fixed, the participation constraint (A.19) dictates a relationship between the lending rate charged, $R_L(k)$, and the size of the loan, $Z(k) - k$: a larger loan allows for a higher lending rate $R_L(k)$ that the bank can charge on it. This, given (A.17), implies

$$R_L(k) \leq \theta_2 - \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\bar{X} - k)} \equiv \bar{R}_L(k), \forall k \in \mathbf{B}, \tag{A.20}$$

where $\bar{R}_L(k)$ is the maximum lending rate the bank is able to charge on firm k, subject to (A.17) and (A.19). Naturally, $\bar{R}_L(k)$ is decreasing in k. With a larger k, the firm's reservation value $V(k)$ is higher and the demand for external finance, $\bar{X} - k$, is smaller, both implying a lower maximum lending rate – the size of the firm imposes a constraint on what the bank can charge on the loan.

Parallel to Assumption 2 in the benchmark environment, we make

Assumption 3. $\pi_2\bar{R}_L(k) + \pi_1(\theta_1 - \gamma_0) > 1, \ \forall k.$

That is, for any k, the bank is better off lending to the firm at the maximum loan rate $\bar{R}_L(k)$, which implies an average rate of return on lending of $\pi_2\bar{R}_L(k) + \pi_1(\theta_1 - \gamma_0)$, than putting the funds on storage.

Under Assumption 3, the bank's rate of return on lending to firm k is

$$R_b(k) = (E(\theta) - \pi_1\gamma_0) - \frac{V(k) - (E(\theta) - \pi_1\gamma_0)k}{L(k)} - R_D, \tag{A.21}$$

where since $V(k) - (E(\theta) - \pi_1\gamma_0)k > 0$ (which holds for all $k \in [0, \bar{k}]$ from (18)), $R_b(k)$ is larger when $L(k)$ is larger. This is so because a larger loan dilutes the net cost of lending to the firm – the second term of the right hand side of the above equation – which in turn results in a higher average rate of return to the bank.[43] Observe the contrast between this and what happened in the benchmark model, where a larger loan implies a lower rate of return on lending.

[43] The bank's rate of return on lending is given by

$$R_b(k) = \frac{\pi_1(\theta_1 - \gamma_0)(L(k) + k) + \pi_2 R_L(k)L(k)}{L(k)} - R_D.$$

Use the binding participation constraint (A.19) to solve for $R_L(k)$ and plug it into the above expression to get

$$R_b(k) = \frac{\pi_1(\theta_1 - \gamma_0)(L(k) + k) + \pi_2\theta_2(L(k) + k) - V(k)}{L(k)} - R_D,$$

which then gives (A.21).

Given the above then, for any $k \in \mathbf{B}$, it is optimal to set $L(k) = \bar{X} - k$, or $Z(k) = \bar{X}$, while the optimal lending rate is set at $R_L(k) = \bar{R}_L(k)$, defined in (A.20). Remember, with a fixed R_L, the bank wants the loans to be of the minimum size. There, by keeping the loans small, the bank lends to more firms, maximizing the use of firm net worth as collateral for enforcing credit repayments. Here, with a flexible R_L, the bank is able to make loans larger to minimize its cost of lending per unit of the loan.

Moreover, constraint (A.18) does not bind,[44] and so the bank's problem is reduced to choosing $\mathbf{B}$ to maximize its total profits subject to constraint (21), and the solution has

$$\mathbf{B} = [\hat{k}_1, \hat{k}_2] = \{k : \lambda(k) \geq \lambda^*\},$$

where $\lambda(\cdot)$ is the bank's expected rate of return on the loan to firm k and defined in (32). And λ^* is determined by

$$\mu \int_{\{k:\, \lambda(k) \geq \lambda^*\}} (\bar{X} - k) dG(k) = D.$$

That is, to maximize profits, the bank includes in its portfolio firms with the largest $\lambda(k)$s subject to the total funds available, as depicted in Figure 7.

Lemma 13. *$\lambda(k)$ is increasing in k for $k \in [0, \tilde{k}]$, and decreasing in k for $k \geq \tilde{k}$.*

Proof. From equations (18) and (32) we have $\forall k < \tilde{k}$

$$\lambda(k) = \frac{(E(\theta) - \pi_1\gamma_0)\bar{X} - (\bar{X} - k)R_D - [E(\theta) + \pi_1\gamma k - R_D - \pi_1\gamma_0]^2 / (4\pi_1\gamma) - kR_D}{\bar{X} - k},$$

and $\forall k \geq \tilde{k}$

$$\lambda(k) = \frac{(E(\theta) - \pi_1\gamma_0)\bar{X} - (\bar{X} - k)R_D - \pi_2(\theta_2 - \theta_1)R_D k/(R_D - \theta_1)}{\bar{X} - k}.$$

Then for $k \in (0, \tilde{k})$,

$$\lambda'(k) = \frac{E(\theta) - R_D - \pi_1\gamma_0 - \pi_1\gamma k}{2(\bar{X} - k)^2} \left(\bar{X} - \frac{E(\theta) - R_D - \pi_1\gamma_0 + \pi_1\gamma k}{2\pi_1\gamma} \right) > 0;$$

[44] It is straightforward to show that for any $k \in [0, \bar{k}]$,

$$\bar{R}_L(k) = \theta_2 - \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\bar{X} - k)} > \theta_2 - \frac{V(k) - \pi_2\theta_2 k}{\pi_2(Z_0(k) - k)} = R_L \geq \underline{R}_L. \tag{A.22}$$

and for $k \in (\tilde{k}, \bar{k})$,

$$\lambda'(k) = -\frac{\pi_1\gamma_0 + \theta_1(E(\theta) - R_D)/(R_D - \theta_1)}{(\bar{X} - k)^2}\bar{X} < 0.$$

Clearly, $\lambda(k)$ is increasing in k for $k \in (0, \tilde{k})$, and decreasing in k for $k \geq \tilde{k}$. $\square$

A larger k has two effects on $\lambda(k)$. First, it implies a larger $V(k)$ which reduces the returns on lending to firm k. Second, it implies a smaller bank loan $(\bar{X} - k)$, which results in a higher average return on lending, increasing $\lambda(k)$. Figure 7 depicts the function $\lambda(\cdot)$ as in Lemma 13, where $\mathbf{B} = [\tilde{k}_1, \tilde{k}_2]$. Moreover, given $0 < \tilde{k}_1 < \tilde{k} < \tilde{k}_2 < \bar{k}$, it follows from Proposition 2 that firms with $k \in [0, \tilde{k}_1)$ seek monitored private finance, and those with $k \in (\tilde{k}_2, \bar{k}]$ obtain credit by way of issuing bonds. So removing the lending rate ceiling does not change the model's prediction that small firms use private loans, medium sized firms are financed with bank loans, and large firms issue bonds.

As is obvious from Figure A.7 in the appendix, a larger D, by giving a lower λ^*, results in a lower $\tilde{k}_1$ but a larger $\tilde{k}_2$, implying both less bond finance and less monitored private lending.

To determine the equilibrium D, let $\widetilde{Q}(D)$ be the total demand for finance which, after the removal of the lending rate ceiling, is given by

$$\widetilde{Q}(D) = \mu\int_0^{\tilde{k}_1} L_{\mathrm{M}}(k)dG(k) + \mu\int_{\tilde{k}_1}^{\tilde{k}_2}[\bar{X} - k]dG(k) + \mu\int_{\tilde{k}_2}^{\bar{k}} L_{\mathrm{N}}(k)dG(k). \tag{A.23}$$

As is for $Q(D)$ in (30) in the benchmark case, it is easy to verify that $\widetilde{Q}(D)$ is increasing in D.[45] The equilibrium bank deposits, denoted $\widetilde{D}^*$, then solves

$$\widetilde{Q}(\widetilde{D}^*) = M,$$

as depicted in Figure 8.

Obviously, a larger M results in a larger $\widetilde{D}^*$ and, from Figure 7, a lower λ^* which, in turn, implies a lower $\tilde{k}_1$ and a higher $\tilde{k}_2$. In other words, after removing the lending rate ceiling, any time there is an expansion in the supply of credit in the economy, bank loans would crowd out both monitored private lending and bond finance, as in the benchmark model.

Lemma 14. *$\widetilde{Q}(D) > Q(D)$, $\forall D \in (0, D_0)$.*

[45] We drop r^*, which is assumed to be fixed at R_D in this part of the analysis, as an argument of the functions $\widetilde{Q}$ and Q.

Proof. Fix any $D \in (0, D_0)$. Note that $r^* = R_D$ in both cases. We have

$$Q(D) = \mu \int_0^{\tilde{k}} [X(k) - k] dG(k) + \mu \int_{\tilde{k}}^{\hat{k}} [Z_0(k) - k] dG(k) + \mu \int_{\hat{k}}^{\bar{k}} [X(k) - k] dG(k),$$

and

$$\widetilde{Q}(D) = \mu \int_0^{\tilde{k}_1} [X(k) - k] dG(k) + \mu \int_{\tilde{k}_1}^{\tilde{k}_2} [\bar{X} - k] dG(k) + \mu \int_{\tilde{k}_2}^{\bar{k}} [X(k) - k] dG(k),$$

with $\tilde{k}_1 < \tilde{k} < \tilde{k}_2 < \hat{k}$ and

$$D = \mu \int_{\tilde{k}_1}^{\tilde{k}_2} [\bar{X} - k] dG(k) = \mu \int_{\tilde{k}}^{\hat{k}} [Z_0(k) - k] dG(k),$$

or

$$\int_{\tilde{k}_1}^{\tilde{k}} [\bar{X} - k] dG(k) + \int_{\tilde{k}}^{\tilde{k}_2} [\bar{X} - k] dG(k) = \int_{\tilde{k}}^{\tilde{k}_2} [Z_0(k) - k] dG(k) + \int_{\tilde{k}_2}^{\hat{k}} [Z_0(k) - k] dG(k).$$

Given $\bar{X} > Z_0(k)$ for all $k \in [0, \bar{k}]$, we have

$$\int_{\tilde{k}_1}^{\tilde{k}} [\bar{X} - k] dG(k) < \int_{\tilde{k}_2}^{\hat{k}} [Z_0(k) - k] dG(k). \tag{A.24}$$

Apply the mean value theorem, there exist $k' \in (\tilde{k}_1, \tilde{k})$ and $k'' \in (\tilde{k}_2, \hat{k})$ such that

$$\int_{\tilde{k}_1}^{\tilde{k}} [\bar{X} - k] dG(k) = (\bar{X} - k') \int_{\tilde{k}_1}^{\tilde{k}} dG(k)$$

and

$$\int_{\tilde{k}_2}^{\hat{k}} [Z_0(k) - k] dG(k) = (Z_0(k'') - k'') \int_{\tilde{k}_2}^{\hat{k}} dG(k).$$

Given $k' < k''$ and so $\bar{X} - k' > Z_0(k'') - k''$, (A.24) and the above two equations give

$$\int_{\tilde{k}_2}^{\hat{k}} dG(k) / \int_{\tilde{k}_1}^{\tilde{k}} dG(k) > (\bar{X} - k')/(Z_0(k'') - k'') > 1.$$

Apply again the mean value theorem to obtain

$$\begin{aligned}
\widetilde{Q}(D) - Q(D) &= \mu \int_{\tilde{k}_2}^{\hat{k}} [X(k) - k] dG(k) - \mu \int_{\tilde{k}_1}^{\tilde{k}} [X(k) - k] dG(k) \\
&= (X(k^{**}) - k^{**}) \mu \int_{\tilde{k}_2}^{\hat{k}} dG(k) - (X(k^*) - k^*) \mu \int_{\tilde{k}_1}^{\tilde{k}} dG(k) \\
&> \left[(X(k^{**}) - k^{**})(\bar{X} - k')/(Z_0(k'') - k'') - (X(k^*) - k^*) \right] \mu \int_{\tilde{k}_1}^{\tilde{k}} dG(k) \\
&> 0,
\end{aligned}$$

where the last inequality we suppose $\bar{X}$ is large enough. $\square$

What happens is that, for any given D, removing the lending rate ceiling allows the bank to lend more at a higher interest rate to each individual firm. This reduces the measure of firms obtaining a bank loan, increasing the measure of firms in direct lending and their demand for external finance.

Observe from Figure 8 that $\widetilde{D}^* < D^*$. That is, removing the lending rate ceiling results in decreased equilibrium quantity of bank deposits or loans. In addition, given $\tilde{k}_1(\widetilde{D}^*) < \hat{k}_1(D^*) = \tilde{k}$ and $\tilde{k}_2(\widetilde{D}^*) < \tilde{k}_2(D^*) < \hat{k}_2(D^*)$, the equilibrium share of monitored private finance in total lending would decline, but that of bond finance would increase.

Proposition 4. *(i) Fixing M and R_D, removing the lending rate ceiling results in a decline in banking and private lending, but an increase in bond finance. (ii) After the removal of the lending rate ceiling, an increase in M increases the equilibrium bank deposits and loans, but squeezes bond finance and monitored private lending, as in the case of fixed bank lending rate.*

So removing the lending rate ceiling does not alter the direction in which a change in M affects banking. Consider the story behind (i) of the proposition. After removing the rate ceiling, the bank would want its loans to be larger and charge a higher rate (the $\bar{R}_L(k)$). With the given D, it must then take out from its initial portfolio $\mathbf{B}$ a set of larger firms and replace them with a group of smaller firms. This immediately expands the market for bonds – the larger firms, upon leaving the bank, would get finance by way of issuing bonds – and at the same time reduces private lending. The story continues. The adjustment in the bank's portfolio would result in a net increase in the demand for direct finance, pushing up the interest rate on direct lending. This, in turn, would induce individual investors to substitute bank deposits for direct lending, cutting D and lowering the interest rate on direct lending. With the decreased D, the bank must again adjust its loan portfolio to make $\mathbf{B}$ even smaller, moving more (large) firms into direct lending, pushing up again the interest rate on direct lending, inducing more investors to leave bank deposits and join direct lending. And this goes on, until the market settles at a new and lower equilibrium D, the $\widetilde{D}^*$ in Figure 8, together with an expanded bond market but a smaller market for private lending.

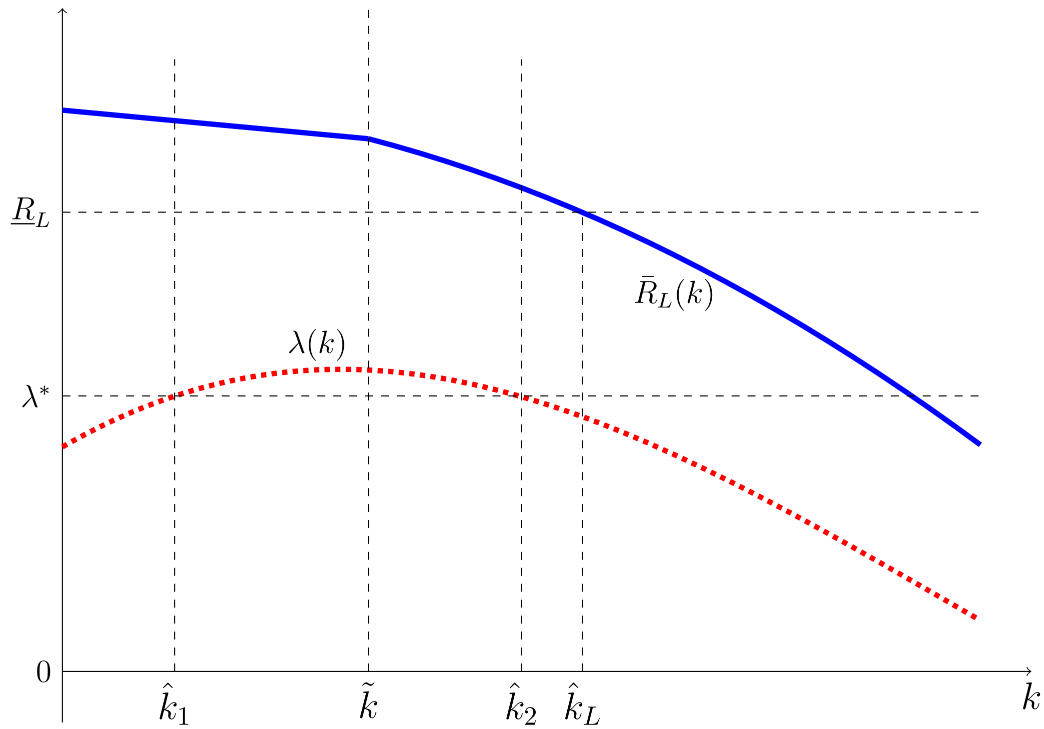
To close this part of the analysis, note that removing the lending rate ceiling is supposed to make the bank more competitive as a credit provider. The outcome of the reform, however, weakens, instead of strengthening, the bank's position in the financial system. The fixed deposit rate R_D plays an important role in the story. It forces the bank to choose larger profits on individual contracts at the expense of the total amount of loans made. Suppose the bank is free to choose optimally both R_L and R_D. Then it may raise R_D to at least partially offset the above effect.

Appendix A.14.2. All lending rate controls removed

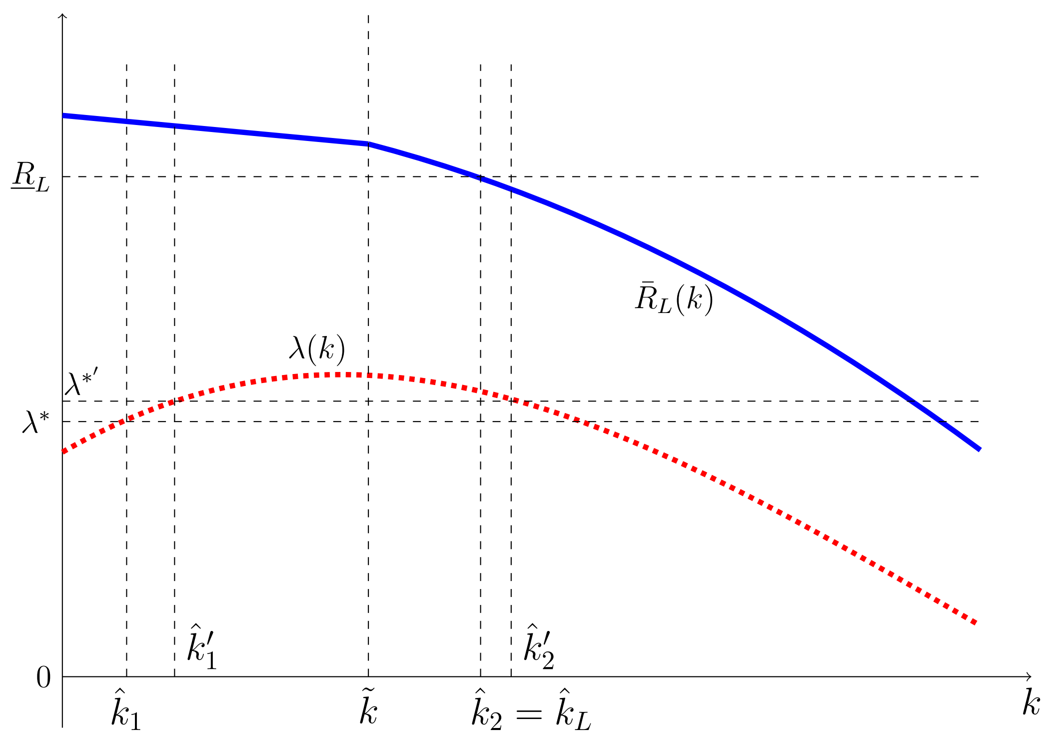
As discussed in the main body of the paper, the effect of the 2013 reform to remove the lending rate floor depends on whether the floor, $\underline{R}_L$, binds before being removed. Let $\hat{k}_L$ solve

$$\bar{R}_L(\hat{k}_L) = \underline{R}_L.$$

From Figure A.16, it is clear that in Case (a), where constraint (A.18) does not bind, the reform has no effects on the outcome of the model. In Case (b), where constraint (A.18) binds, removing the lending rate floor moves the set $\mathbf{B}$ to the right in k to put larger firms in the bank's portfolio. Specifically, after removing the floor, the new λ^* is higher, and both the new $\hat{k}_1$ and new $\hat{k}_2$ are larger ($\lambda^{*\prime} > \lambda^*, \hat{k}_1' > \hat{k}_1$ and $\hat{k}_2' > \hat{k}_2$).



(a) Constraint (A.18) does not bind



(b) Constraint (A.18) binds

Figure A.16: The effects of removing the lending rate floor

Appendix A.14.3. Removing all deposit rate controls

An equilibrium of the model is now defined as a measure of investors who choose to lend through the bank $D^* \in [0, M]$ and an interest rate on direct lending r^* which all agents in the markets take as given.

Taking D^* and r^* as given, the bank solves

$$\max_{R_D, \mathbf{B}, \{R_L(k), Z(k)\}_{k \in \mathbf{B}}} \mu \int_{\mathbf{B}} \Big\{ \pi_1 (\theta_1 - \gamma_0) Z(k) + \pi_2 R_L(k) [Z(k) - k] \Big\} dG(k) \\ + D - \mu \int_{\mathbf{B}} [Z(k) - k] \, dG(k) - R_D D \tag{A.25}$$

subject to (19), (21), (A.17), (A.19) where

$$D \equiv \begin{cases} M, & \text{if } R_D > r^* \\ D^*, & \text{if } R_D = r^* \\ 0, & \text{if } R_D < r^* \end{cases}. \tag{A.26}$$

Note that the bank does not choose D directly, but it chooses D indirectly by way of choosing R_D, with the given D^*, r^*, and M, as in A.26.

The solution to the above problem has:[46]

(i) $R_D = r^*$.

(ii) For all $k \in \mathbf{B}$, $Z(k) = \bar{X}$, and $R_L(k) = \bar{R}_L(k)$ (given in (A.20)).

(iii) $\mathbf{B} = \{k : \lambda(k) \geq \lambda^*\}$, where $\lambda(k), k \in [0, \bar{k}]$, is given in (32), and λ^* solves

$$\mu \int_{\{k:\, \lambda(k) \geq \lambda^*\}} (\bar{X} - k) dG(k) = D.$$

And it follows from (iii) that $\mathbf{B} = [\tilde{k}_1(D, r^*), \tilde{k}_2(D, r^*)]$, as in the case of fixed R_D but flexible $R_L(k)$ (Figure 7). That is, in equilibrium the bank includes in its loan portfolio medium-sized firms with net worth levels that are neither too large nor too small. The largest firms would raise finance from the bond market, and the smallest with monitored private lending.

Let $D(D^*, r^*)$ be the investors' deposits induced by the bank's optimal response (in R_D) to D^* and r^* which it takes as given. For r^* and D^* to constitute an equilibrium, the solution

[46] Notice that what equation (A.26) describes, namely D as a function of R_D, is not continuous and has a non-convex image.

to the bank's problem must have

$$D(D^*, r^*) = D^*,$$

and the market for direct lending must clear:

$$\mu \int_0^{\tilde{k}_1(D^*,r^*)} L_{\mathrm{M}}(k, r^*) dG(k) + \mu \int_{\tilde{k}_2(D^*,r^*)}^{\bar{k}} L_{\mathrm{N}}(k, r^*) dG(k) = M - D^*, \tag{A.27}$$

where $L_{\mathrm{M}}(k, r^*) = X_{\mathrm{M}}(k, r^*) - k$ and $L_{\mathrm{N}}(k, r^*) = X_{\mathrm{N}}(k, r^*) - k$ are given in (5) and (12) .

Proposition 5. *Let direct lending and bank loans coexist before the reform. Removing the control on R_D results in a higher equilibrium interest rate on direct lending and deposits (r^* and R_D higher). It also squeezes the market for direct lending while expanding the market for bank loans (D^* larger). With a higher interest rate, firms in the private lending market each raise a smaller amount of finance ($X(k) - k$ smaller) and operate a smaller project.*

Proof. Denote the deposit rate and equilibrium quantity of deposits in the benchmark model as $\underline{R}_D$ and $\underline{D}$ respectively. Remember we are assuming $\underline{Q}(\underline{R}_D) < M < Q_0(\underline{R}_D)$ so that in equilibrium all there markets are active in the benchmark model.

Assume $\bar{X}$ is large enough so that $\lambda(k) > 0$ for all $k \in [0, \bar{k}]$ (note that $\lambda(k)$ is strictly increasing in $\bar{X}$ and goes to $E(\theta) - \pi_1\gamma_0$ as $\bar{X} \to \infty$), supposing of course that the loans the bank makes offer positive expected rates of return to the bank.

For all $D \in [0, M]$, let

$$U(D) = \mu \int_{\tilde{k}_1(D)}^{\tilde{k}_2(D)} \lambda(k)(\bar{X} - k) dG(k) \tag{A.28}$$

denote the bank's total profits earned, conditional on D. With this, the bank's optimal loan portfolio is given by

$$\mathbf{B} = \begin{cases} [\tilde{k}_1(M), \tilde{k}_2(M)], & \text{if } U(M) > U(D^*) \text{ and } U(M) \geq 0 \\ [\tilde{k}_1(D^*), \tilde{k}_2(D^*)], & \text{if } U(M) \leq U(D^*) \text{ and } U(D^*) \geq 0 \cdot \\ \emptyset, & \text{otherwise} \end{cases}$$

Now suppose the economy is initially in an equilibrium state of the benchmark model with $R_D = r^* = \underline{R}_D$. Given $\lambda(k) > 0$ for all $k \in [0, \bar{k}]$ then, the optimal choice of the bank has $D = M$ and $\mathbf{B} = [\tilde{k}_1(M), \tilde{k}_2(M)]$. But given $\underline{Q}(\underline{R}_D) < M < Q_0(\underline{R}_D)$ we have $\tilde{k}_2(M) < \bar{k}$ or

$\tilde{k}_1(M) > 0$.[47] This implies the total demand for credit in the direct lending market is larger than the supply of credit (see equation (A.27)). This will increase r^* and then raise R_D. Thus in equilibrium $R_D = r^* > \underline{R}_D$. And with a higher equilibrium interest rate, the size of direct lending is smaller for all $k \in [0, \bar{k}]$, which, in turn, results in a larger quantity of equilibrium deposits for the bank (see the market clearing condition (29)). □

Lastly, note that with all the interest rate controls on banking removed, one would think that the bank, being the more efficient monitor, should be able to crowd out monitored direct lending completely. From the above discussion, however, monitored private lending is active in equilibrium if $\tilde{k}_1(D^*) > 0$ which, given the non-convexity of the bank's choices in D (see (A.26)), is hard to rule out.

Appendix A.14.4. The effect of the deposit rate ceiling

To look at the effect of the deposit rate ceiling, let $\bar{R}_D$ be the deposit rate ceiling and add constraint $R_D \leq \bar{R}_D$ to problem (A.25).

Suppose the equilibrium has $r^* > \bar{R}_D$. Then for r^* and $D^* = 0$ to constitute an equilibrium, r^* must be strictly greater than $\bar{R}_D$ and satisfy the market clearing condition for direct lending:

$$\mu \int_0^{\tilde{k}_1(0,r^*)} L_{\mathrm{M}}(k, r^*) dG(k) + \mu \int_{\tilde{k}_2(0,r^*)}^{\bar{k}} L_{\mathrm{N}}(k, r^*) dG(k) = M, \tag{A.29}$$

where $L_{\mathrm{M}}(k, r^*) = X_{\mathrm{M}}(k, r^*) - k$ and $L_{\mathrm{N}}(k, r^*) = X_{\mathrm{N}}(k, r^*) - k$ are as defined earlier in equations (5) and (12), given that the interest rate on direct lending is r^*.

Suppose the equilibrium has $r^* \leq \bar{R}_D$ and the constraint $R_D \leq \bar{R}_D$ binds for the bank. Then for r^* and D^* to constitute an equilibrium, r^* must equal $\bar{R}_D$ and D^* must satisfy $D(\bar{R}_D, D^*) = D^*$ and the market clearing condition for the direct lending market

$$\mu \int_0^{\tilde{k}_1(D^*,\bar{R}_D)} L_{\mathrm{M}}(k, \bar{R}_D) dG(k) + \mu \int_{\tilde{k}_2(D^*,\bar{R}_D)}^{\bar{k}} L_{\mathrm{N}}(k, \bar{R}_D) dG(k) = M - D^*. \tag{A.30}$$

Suppose the equilibrium has $r^* \leq \bar{R}_D$ and the constraint $R_D \leq \bar{R}_D$ does not bind for the

[47] Otherwise, suppose $\tilde{k}_1(M) = 0$ and $\tilde{k}_2(M) = \bar{k}$. We have

$$M \geq \mu \int_0^{\bar{k}} (\bar{X} - k) dG(k) > Q_0(\underline{R}_D).$$

A contradiction.

bank. Then for r^* and D^* to constitute an equilibrium, r^* must be no greater than $\bar{R}_D$ and (r^*, D^*) must satisfy $D(r^*, D^*) = D^*$ and the market clearing condition for the direct lending market

$$\mu \int_0^{\tilde{k}_1(D^*,r^*)} L_{\mathrm{M}}(k, r^*) dG(k) + \mu \int_{\tilde{k}_2(D^*,r^*)}^{\bar{k}} L_{\mathrm{N}}(k, r^*) dG(k) = M - D^*. \tag{A.31}$$

Now suppose M is large enough that

$$M - D^* > \mu \int_0^{\tilde{k}_1(D^*,\bar{R}_D)} L_{\mathrm{M}}(k, \bar{R}_D) dG(k) + \mu \int_{\tilde{k}_2(D^*,\bar{R}_D)}^{\bar{k}} L_{\mathrm{N}}(k, \bar{R}_D) dG(k).$$

From the analysis in Section 3.1, the functions $L_{\mathrm{M}}(k, r^*)$ and $L_{\mathrm{N}}(k, r^*)$ are decreasing in r^*. So to clear the market for direct lending, we must have $D^* > 0$ and $r^* < \bar{R}_D$. In other words, only $r^* < \bar{R}_D$ and $D^* > 0$ could constitute an equilibrium, and the equilibrium deposit rate $R_D = r^*$ would be strictly less than $\bar{R}_D$.

Appendix A.15. Optimal lending under constraint (36)

Under the borrowing constraint (36), the problem of non-monitored direct lending becomes

$$V_{\mathrm{N}}(k) \equiv \max_{\{r_{\mathrm{N}}; L \geq 0\}} \left\{ \pi_1 \theta_1 (L + k) + \pi_2 \theta_2 (L + k) - r_{\mathrm{N}} L \right\} \tag{A.32}$$

subject to (3), (4) and

$$L \leq \alpha k. \tag{A.33}$$

As in Lemma 1, the optimal solution has

$$L'_{\mathrm{N}}(k) = \min \left\{ \alpha, \frac{\theta_1}{r^* - \theta_1} \right\} k,$$

$$X'_{\mathrm{N}}(k) = \min\{E(\theta)(1 + \alpha) - r_{\mathrm{N}} \alpha, \frac{r^*}{r^* - \theta_1}\} k.$$

On the other hand, the problem of monitored direct lending becomes

$$V_{\mathrm{M}}(k) \equiv \max_{\{r_1, r_2, L \geq 0\}} \left\{ \pi_1 \left[\theta_1 (L + k) - r_1 L - \widetilde{C}(L, k) \right] + \pi_2 \left[\theta_2 (L + k) - r_2 L \right] \right\} \tag{A.34}$$

subject to (6), (7), (8), (9) and

$$L \leq \alpha k. \tag{A.35}$$

Now suppose $\alpha > \theta_1/(r^*-\theta_1)$.[48] Then as in Proposition 2, there exists a cut-off of k, $\tilde{k}' \in (0, \tilde{k})$, such that the optimal direct lending contract has

$$X'(k) = \begin{cases} X'_{\mathrm{M}}(k) = \min\left\{(1+\alpha)k, X_{\mathrm{M}}(k)\right\}, & \forall k < \tilde{k}' \\ X'_{\mathrm{N}}(k) = \frac{r^*}{r^*-\theta_1}k, & \forall k \geq \tilde{k}' \end{cases}$$

and

$$V'(k) = \begin{cases} V'_{\mathrm{M}}(k) = E(\theta)X'_{\mathrm{M}}(k) - r^*(X'_{\mathrm{M}}(k) - k) - \pi_1 C(X'_{\mathrm{M}}(k) - k, k), & \forall k < \tilde{k}' \\ V'_{\mathrm{N}}(k) = \pi_2(\theta_2 - \theta_1)\frac{r^*}{r^*-\theta_1}k, & \forall k \geq \tilde{k}' \end{cases}, \quad \text{(A.36)}$$

where $\tilde{k}$ and $X_{\mathrm{M}}(k)$ are given in Proposition 2, $X'_{\mathrm{M}}(\cdot)$ and $X'_{\mathrm{N}}(\cdot)$ are the solutions to problems (A.32) and (A.34) respectively, $V'_{\mathrm{M}}(\cdot)$ and $V'_{\mathrm{N}}(\cdot)$ are the associated value functions, and $\tilde{k}'$ solves $V'_{\mathrm{M}}(\tilde{k}') = V'_{\mathrm{N}}(\tilde{k}')$.

The bank's problem becomes

$$\max_{\mathbf{B}, \{R_1(k), R_L(k), L(k)\}_{k\in\mathbf{B}}} \mu \int_{\mathbf{B}} \Big\{(\pi_1 R_1(k) + \pi_2 R_L(k) - 1)L(k)\Big\} dG(k) + D - R_D D$$

subject to (19), (21), (A.17), (A.19) and

$$L(k) \leq \alpha k, \ \forall k \in \mathbf{B}. \quad \text{(A.37)}$$

From the analysis in Appendix A.14.1, the solution to the bank's problem has

$$L'_{\mathrm{B}} = \min\left\{\alpha k, \bar{X} - k\right\}.$$

Now the bank's rate of return on lending to firm k becomes

$$\lambda(k,\alpha) = E(\theta) - \pi_1\gamma_0 - \frac{V(k) - (E(\theta) - \pi_1\gamma_0)k}{\min\{\alpha k, \bar{X} - k\}} - R_D.$$

Together with (A.36) and Corollary 7 in Appendix A.5, we have

$$\lambda(k,\alpha) \begin{cases} \text{is increasing in } k, & \forall k < \tilde{k} \\ \text{is constant in } k, & \forall k \in [\tilde{k}, \bar{X}/(1+\alpha)] \\ \text{is decreasing in } k, & \forall k > \bar{X}/(1+\alpha) \end{cases},$$

as depicted in Figure A.8.

[48] Otherwise only bond finance exists in equilibrium.

Appendix A.16. Equilibrium with government backed lending

The market clearing condition for credit is now modified to read

$$\begin{aligned} M - G &= \mu \int_0^{k_1} L_M(k)dG(k) + \mu \int_{k_1}^{k_2} L_B(k)dG(k) + \mu \int_{k_2}^{\bar{k}} L_N(k)dG(k) \\ &= \mu \int_0^{k_1} (X(k) - k)dG(k) + \mu \int_{k_1}^{k_2} (Z(k) - k)dG(k) + \mu \int_{k_2}^{\bar{k}} (X(k) - k)dG(k). \end{aligned}$$

Note that the value of G would affect the bank's equilibrium lending portfolio (k_1^* and k_2^*), but has no effect on the size of the firm's project, (i.e., $X(\cdot)$ or $Z(\cdot)$). And, remember, $\overline{Q}$ and $\underline{Q}$ are given as

$$\underline{Q} = \mu \int_0^{\bar{k}} (X(k) - k)dG(k),$$

and

$$\overline{Q} = \mu \int_0^{\bar{k}} (Z(k) - k)dG(k),$$

and none of them depend on G.