China's Financial System in Equilibrium

Jie Luo and Cheng Wang*

April 18, 2017

Abstract

This paper presents 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. Consistent with data, in equilibrium smaller firms obtain finance through the private lending market, larger firms use bank loans, and the largest by way of corporate bonds. The model predicts, and the data supports, that removing the controls on bank lending rates or tightening the supply of external finance reduces bank loans but increases bond finance. We argue that this may partially explain the observed decline in banking and the rise of the bond market in China, over the past ten years. The model also suggests that removing all interest rate controls would increase the rate of return on lending, expanding banking but squeezing direct lending.

JEL: G18, G21, O16
Keywords: China, financial system, monitoring, banking reforms

*First version: May 2016. Luo: School of Economics and Management, Tsinghua University; luojie@sem.tsinghua.edu.cn. Wang: School of Economics, Fudan University; wangcheng@fudan.edu.cn. We thank Steve Williamson, Jun Qian, and the audience at the Tsinghua PBC School of Finance, University of Goettingen, University of Hong Kong, SHUFE, SAIF, and the Antai School of Economics and Management of Shanghai Jiaotong University for comments.

1 Introduction

China's financial system consists of a state-owned, tightly regulated, monopolistic banking sector, a less formal and decentralized direct lending market, an equity market, and a growing bond market. In this paper, we motivate and construct an equilibrium model of the financial market to study China's financial system. The paper explains why bank regulations give rise to the coexistence of monopoly banking and decentralized private lending. It explains how financial resources are allocated, through the different sectors of the system and by means of differential instruments, to firms who differ in net worth and ability in obtaining finance. The source and size distribution of external finance is determined endogenously in the model. The model is then used to evaluate the effects of recent banking reforms, in particular the central bank moves in lifting away controls on bank deposit and lending rates.

1.1 China's financial system – an overview

While there is no official data on the size of the informal lending market, Figure 1 shows how large and important each of the other three parts of China's financial system is, relative to total financing (excluding informal lending). Specifically, it depicts the division between bank loans and the two other types of finance as a fraction of total lending, in time series and for the period 2002-2015.[1] Notice that the equity market is small, and stays small in size relative to the two other mechanisms of lending. Notice, more importantly, the decline in banking and the rise of the market for bonds over the same period.

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) estimate it to be 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.

To picture the dominance of the state owned banks in China's banking system, Figure 2 measures the degree of bank concentration in China, showing the time series of total loans held by the largest five banks, all state-owned, as a fraction of total bank loans in China, relative to the U.S.. Observe that bank concentration has been decreasing but is still much

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

higher in China than in the U.S..[2]

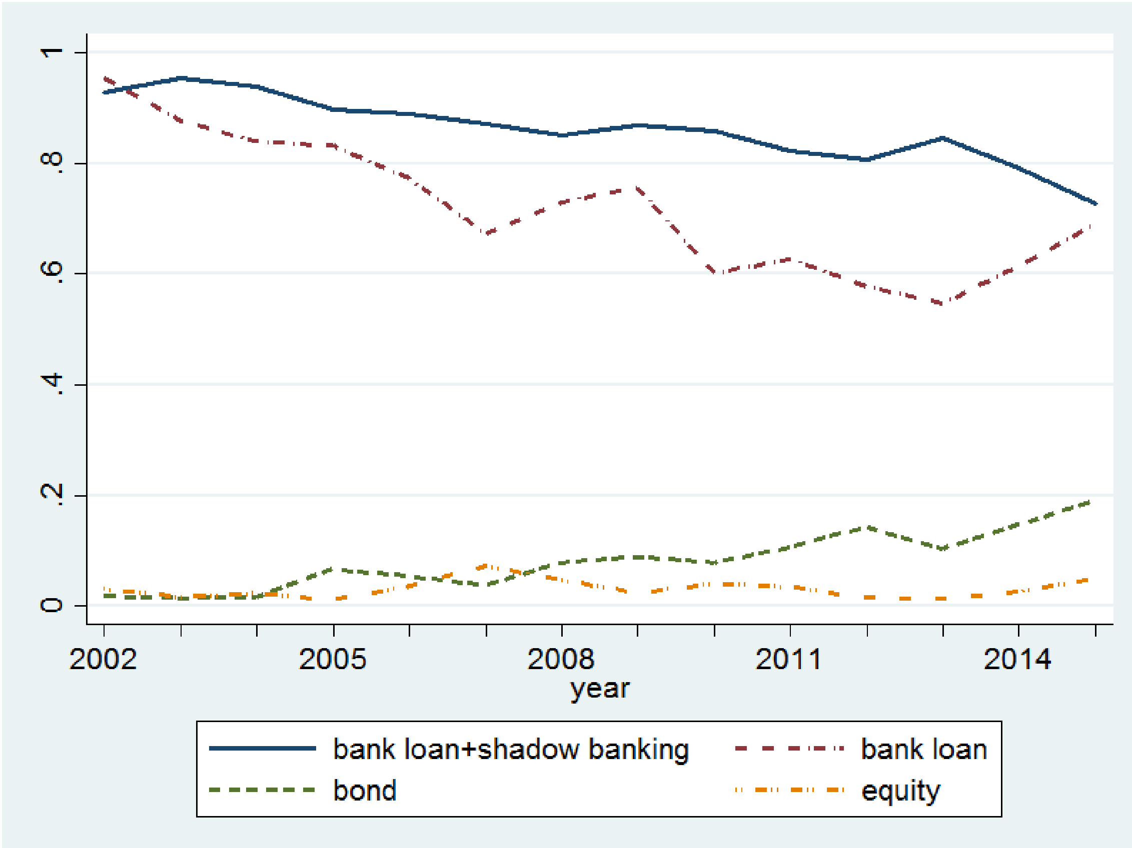


Figure 1: Composition of aggregate financing in China

Source: CEIC.[3]

Note: The fraction of bank loans equals (loans in local currency + loans in foreign currency)/aggregate financing. The fraction of shadow banking equals (trust loans + entrusted loans + banker's acceptance bills)/aggregate financing. The fraction of bond equals corporate bond financing/aggregate financing. The fraction of equity equals non-financial enterprise equity financing/aggregate financing.

The majority of banks in China are commercial banks. According to Bankscope, in 2015 there were 154 commercial banks in China, accounting for 67.7% of total bank assets and 75.9% of bank loans; whereas in the U.S. there were 5064 commercial banks accounting for 28.3% of total bank assets and 33.6% of bank loans.[4] Figure 3 shows the distributions of

[2] Chang et al. (2015) estimate that the share of large national banks in total bank loans was on average 67.4% between 2010 and 2014 (with a share of 51.2% for the Big Four).

[3] 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.

[4] In the U.S. there are about 700 bank holding companies that account for 35% of total bank assets and 32.6% of total bank loans.

commercial banks in the quantity of loans made, in China and the U.S.. Obviously, banks are on average larger and more concentrated in China than in the U.S..

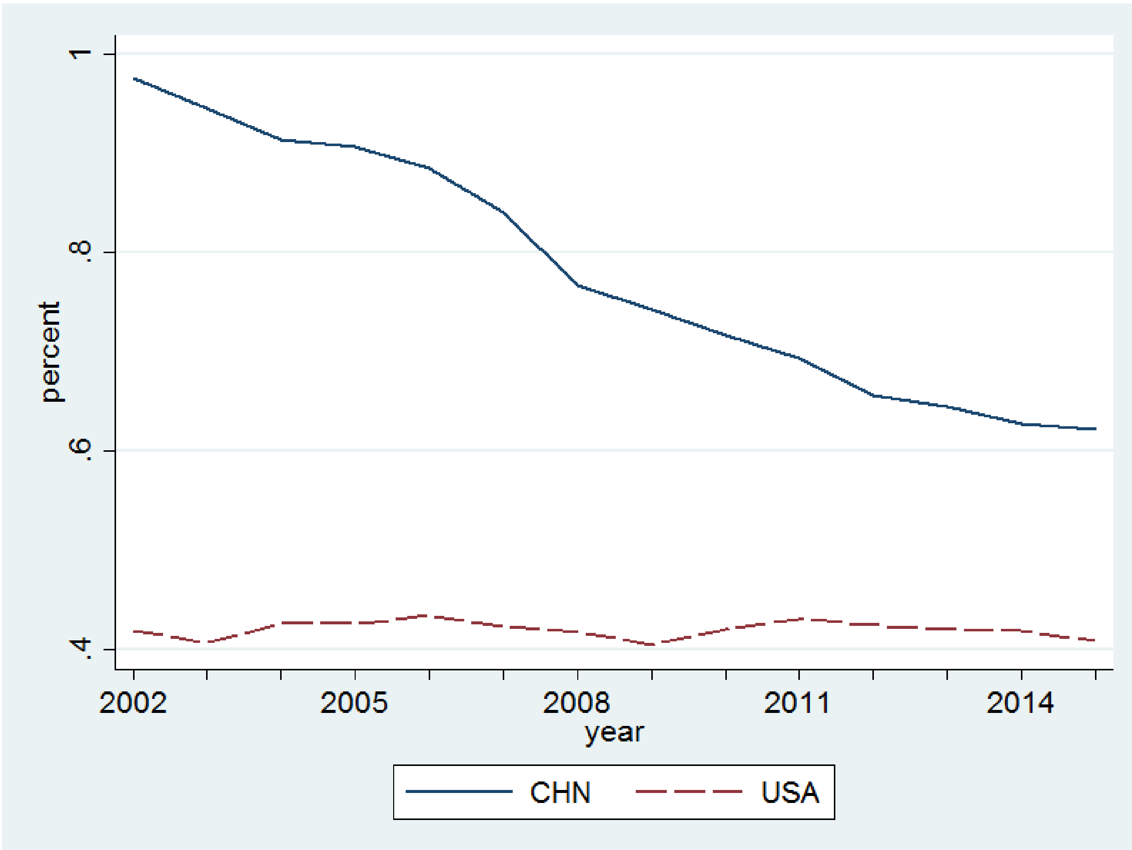


Figure 2: 5-bank loans concentration in commercial banks in China and U.S.

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 and Bank of Communications, and in the U.S. are Wells Fargo Bank, Bank of America, JP Morgan Chase Bank, Citibank and US Bank National Association. The 5-bank concentration within bank holding companies in the U.S. is similar to that within Commercial banks.

Banks in China are largely state owned and subject to state controls, although the last ten years has seen policy moves in lifting up the controls, especial on the deposit and lending rates. 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. The PBC removed the lending rate floors in July 2013, and then, by 2015, its controls on deposit rates.[5] Figure 4 depicts the time series of the policy rates on

[5]Bank regulations exist also on the quantity of loans. In fact, in many cases the PBC conducts its monetary

one year loans and on one year saving deposits.[6] Notice the greater variability in both the policy lending and deposit rates after 2004.

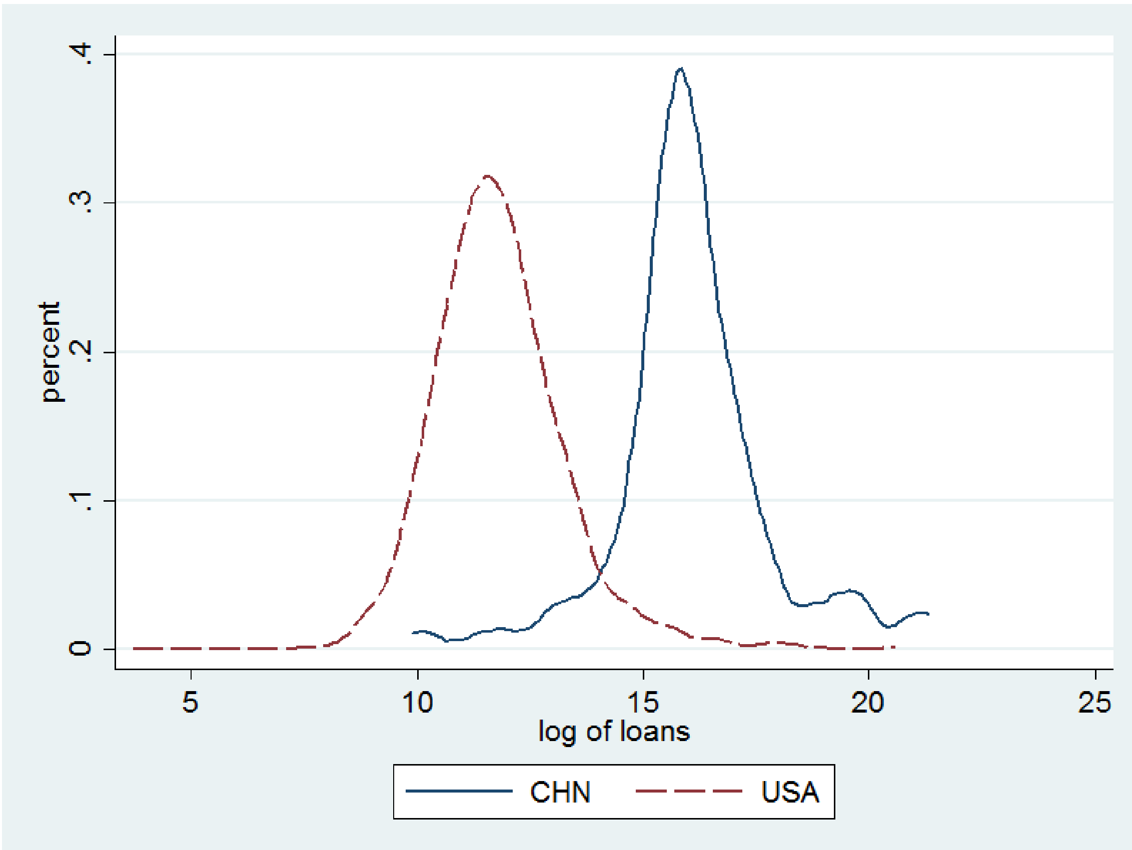


Figure 3: Commercial banks distribution, China and U.S., 2015

Source: Bankscope, self-calculations.

In the private lending market, there is much larger variability in the nominal lending rates, ranging from nearly zero from relatives to more than 30% from money lenders. He et al. (2015) document that interest rates in the private credit markets are much more opaque and higher. They also show that the average lending rates in the private credit market are 2 ∼ 3 times more than the bank lending rates.[7]

policy by way of imposing specific constraints on the quantities of loans commercial banks are allowed to make. We leave this equally important aspect of the Chinese banking system for possible future research.

[6]The policy rates are the benchmarks from which the actual rates are allowed to deviate up to a given maximum percent.

[7]See Figure 6 in their paper.

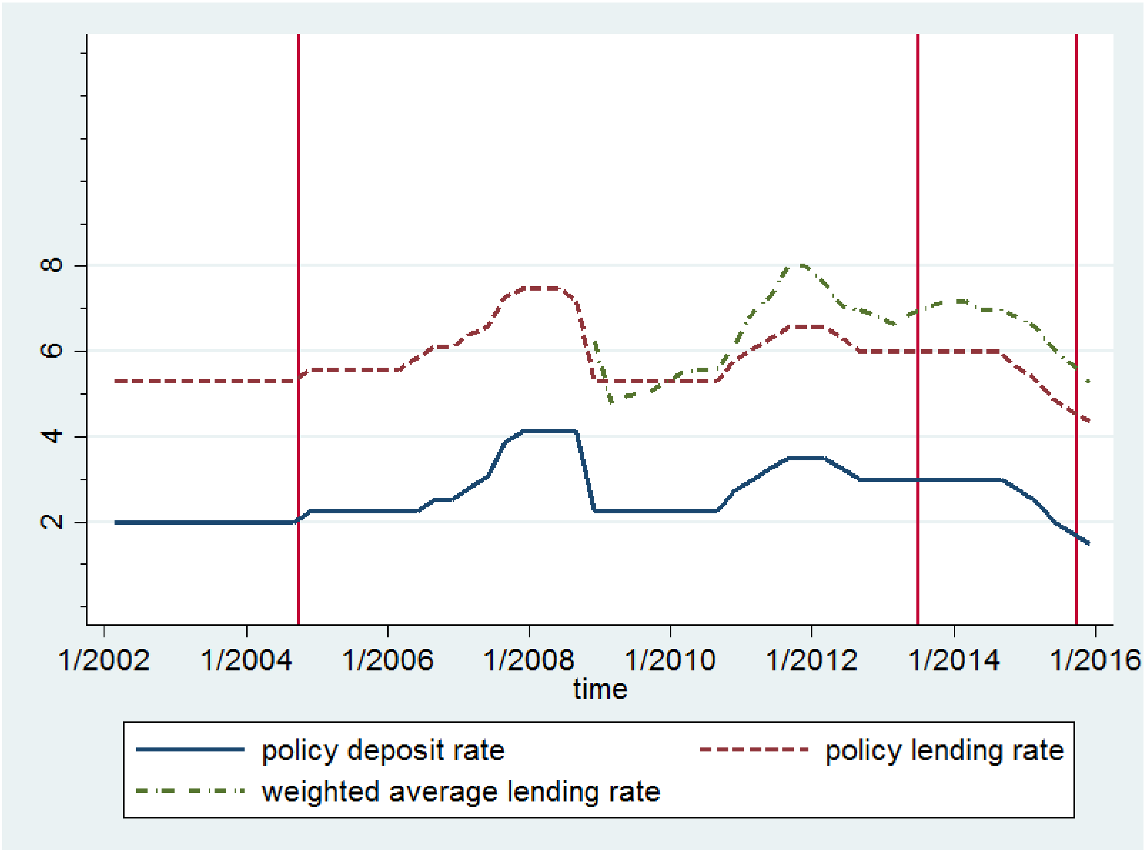


Figure 4: Monthly lending and deposit rates in China

Source: CEIC.
Note: The weighted average lending rate is available only from year 2009.

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 obtaining bank loans, and there are many policy discussions on how to encourage banks to expand loans to smaller businesses. Table 1, which reports a summary of Word Bank's enterprise surveys for China 2012, shows that the percent of firms using bank loans for investment financing is on average much lower in China, relative to other countries in the world. Specifically, for the small firms in the survey, it is 3.8% in China, 16.8% in East Asia and Pacific, and 21.5% across all countries. Allen, Qian and Qian (2005) find that during a small private firm's growth period, the most important financing channel is private credit agencies (PCAs), instead of banks. Dollar and Wei (2007) report that private firms, which have smaller sizes on average, rely less on bank loans but more on families and friends for finance.[8] Ayyagari

[8] Allen, Qian and Qian (2005) argues that the growth of SOEs and foreign companies in China relies heavily on the banking, while the growth of private economy has to rely on alternative financing such as retained earnings, informal financing and in-kind finance (trade credit). Also, Kroeber (2016) mentions that

et al. (2010) also find that in China bank financing is more prevalent with larger firms.[9]

Table 1: Percent of firms using banks to finance investments

	China	East Asia & Pacific	All Countries
Small (5-19)	3.8	16.8	21.5
Medium (20-99)	20.4	23	27.1
Large (100+)	23.3	22.7	30.7

Source: World Bank's Enterprise Surveys data for China 2012.
Note: Only manufacturing firms are included. Small, medium, and large firms are defined by the number of employees.

To look more deeply into the relationship between firm size and bank loans, we rank the firms in the World Bank's Surveys data for China 2012 by size and divide them into 5 groups.[10] Table 2a shows that the fraction of firms that use bank loans as the only source of external finance is increasing in firm size. For the publicly listed firms in China, which are much larger than those in the World Bank's surveys, the fraction of them using bank loans as the only source of external finance initially increases but then decreases, as firm size increases (see Table 2b). To obtain a more comprehensive view, we merge the publicly listed firms and those in World Bank's Enterprise Survey, rank and divide them into 10 groups by size. A clear inverted-U relationship between firm size and the fraction of firms using bank loans as their only source of external finance emerges, as shown in Figure 5.

One might suggest that bank loans are, for some reason, too expensive to smaller firms. This is not the case, as Table 3 shows. Specifically, the third and fourth rows suggest that among those who need a loan but choose not to apply for one, for the small firms the most important reason is that the application procedures were complex; while for larger firms, it is the unfavorable interest rates. The fourth row of the table also indicates that, relative to larger firms, a larger fraction of small firms would like to obtain a bank loan at the ongoing interest rate, but could not. In addition, the seventh row of the table shows that 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.

P2P in China, fills a demand for credit from consumers and going part way to solving the problem of getting financing to small firms.

[9]Using data from the World Bank Investment Climate Survey 2003, they find that in financing capital expenditures, the very large firms use more bank financing (30%) than micro and small firms (15%).

[10]Following the World Bank, firm size is measured as total employment.

Table 2: Number of firms in China, by firm's size and sources of finance

(a) Within manufacture firms in 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 manufacture 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.

Table 3: Percent of reasons why firms did not apply for any line of credit

	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.

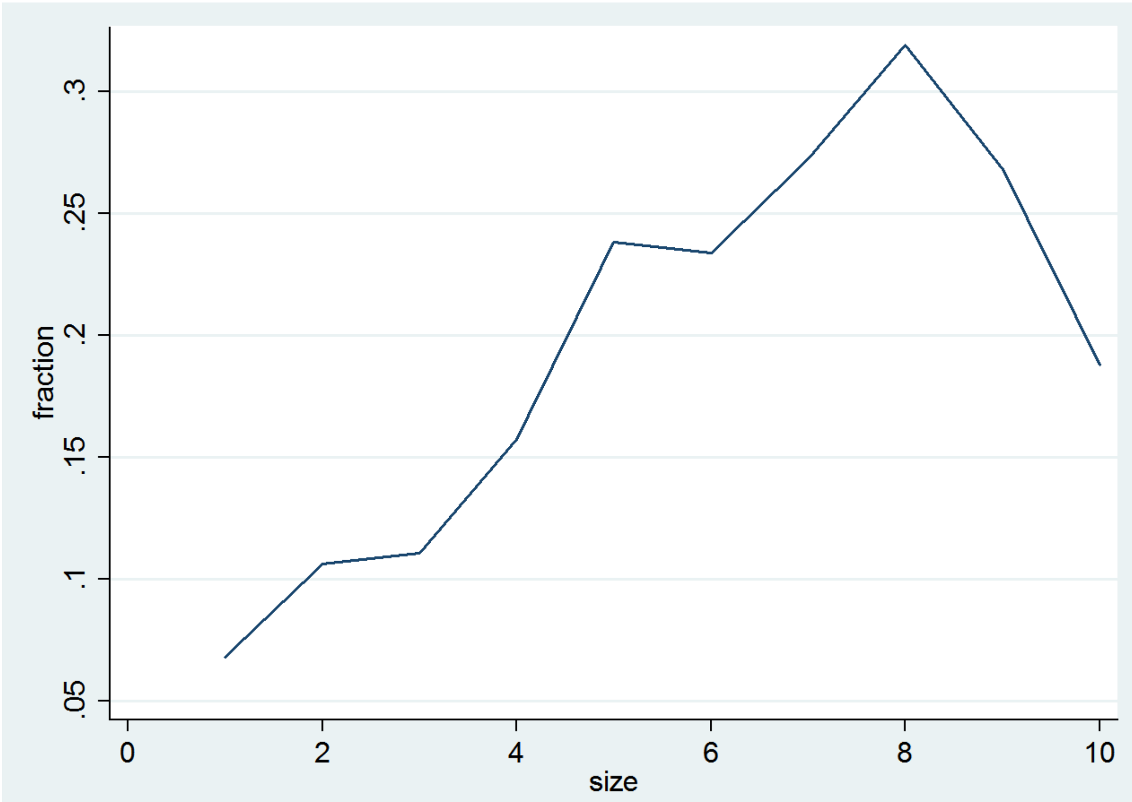


Figure 5: Fraction of firms in China with only bank finance, 2011

Source: World Bank's Enterprise Surveys data for China 2012 and CSMAR.[11]
Note: The X-axis represents the firms' group number, where larger value implies larger size of firms.

China's bond market, where the majority of contracts traded are government and corporate bonds, has grown over the last ten years, from virtually nonexistent to the third biggest in the world, just behind the U.S. and Japan. From Figure 6, although corporate bonds still account for a smaller part of the whole bond market, they have grown fast in relative size over the recent years. Another important feature of China's bond market, as shown in Figure 7, is that the firms who use bonds as a means of external finance are much larger in size than those use bank loans who, in turn, are larger than those who use neither bonds nor bank loans. [12]

[11] 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.

[12] That firms who use bonds for external finance are larger than those who use bank loans is not just observed among Chinese firms.

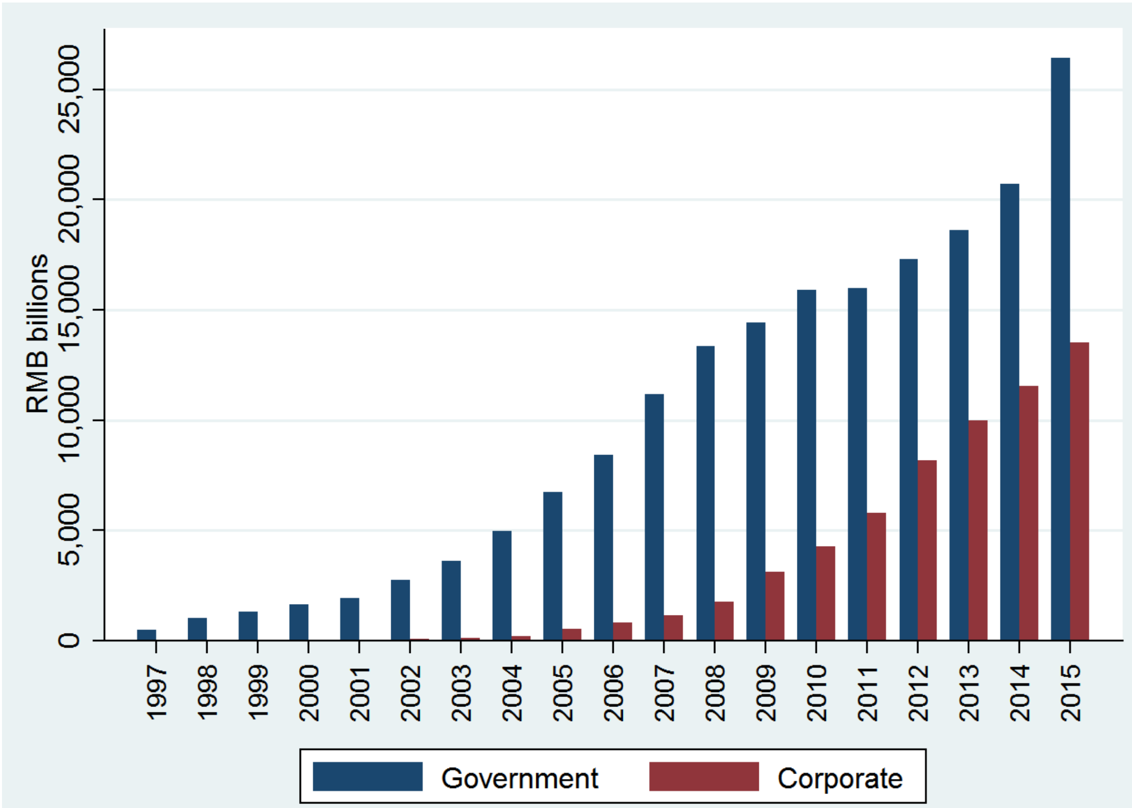


Figure 6: Size of local currency bonds in China

Source: AsianBondsOnline.
Note: Government bonds include obligations of the central government, local governments, and the central bank. Corporate bonds comprise both public and private companies.

1.2 What this paper does

It is not difficult to explain why state owned banks dominate China's financial system.[13] More interesting questions are why the private lending market even exists, and is rising in size relative to the largely state owned banking sector; and why the observed source distribution of finance is such that larger firms are associated with bonds and bank loans, while smaller enterprises obtain finance from the private lending market. In what directions would the composition of the Chinese financial system move when regulations on banking are further loosened? These questions are important, not just for interpreting existing data, but also because of immediate policy concerns. To answer these questions, however, one must first understand how China's financial system works – what's inside it that generates the features and characteristics one observes. This motivates our work.

[13] See, for example, Allen and Qian (2014).

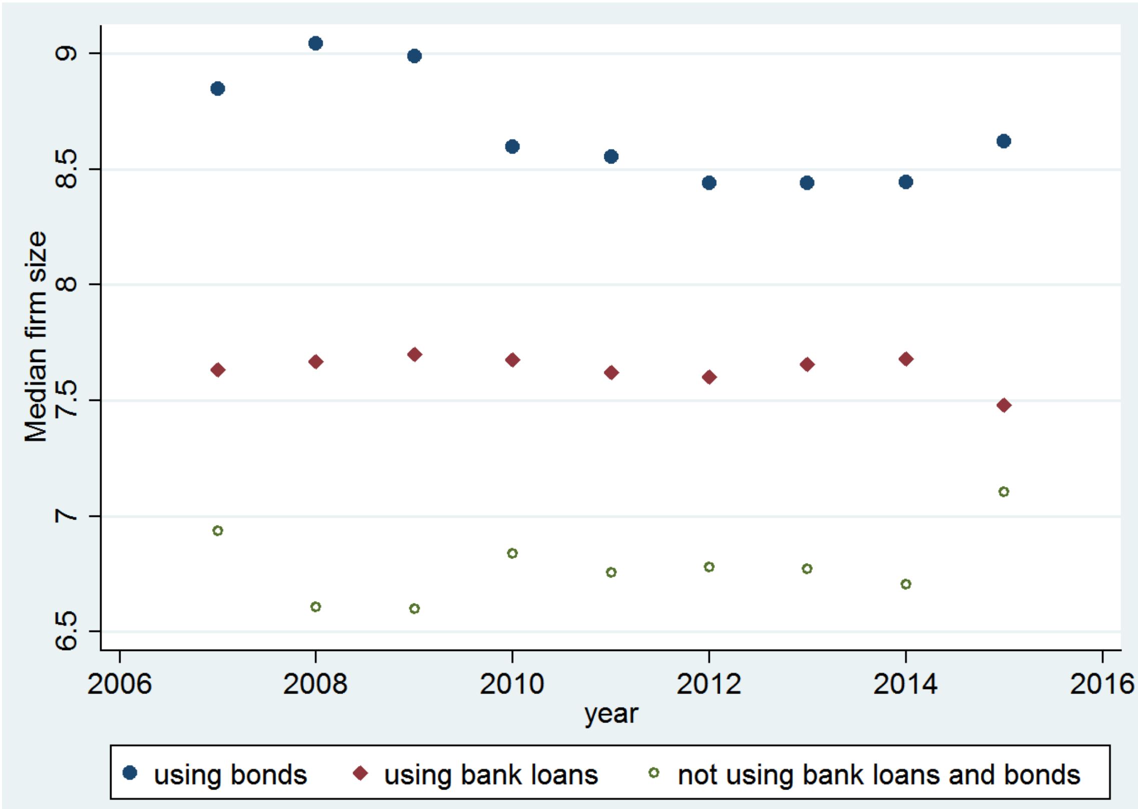


Figure 7: Using versus not using bonds: the median size of listed firms in China, 2007-2015

Source: CSMAR.
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.

In this paper, we first develop a benchmark model to characterize the coexistence of a tightly regulated, monopolistic banking system, and a decentralized direct lending sector where corporate bonds and privately monitored loans are traded. Individual investors are free to lend indirectly through the bank, or directly through the bond market or the market for private lending, while firms are are free to pick any instrument for external finance. The sizes of the submarkets are determined endogenously, and how large each of them is relative to the rest depends on the values of the policy variables, the rate of return paid on bank deposits for example, and the parameters that define the environment, including especially the total supply of external finance. In equilibrium firms with larger net worth obtain finance from the bank, while those with smaller net worth borrow from individual lenders in the private lending market. We then modify the model in ways with which regulations on bank interest rates are lifted, as occurred in the past twenty years, to evaluate the effects of the observed major policy moves. In particular, we use the model to make predictions on what would happen if the bank is set free to compete with private lenders.

We take a standard approach to model lending and financial intermediation (banking), following the ideas of Diamond (1984) and Williamson (1986). Specifically, 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. In the model, private lending coexists with the more efficient bank lending because the low (regulated) deposit rate induces investors to participate in private lending for higher returns; or because a tight supply of external finance dictates a sufficiently high interest rate on private lending to compete credit away from banking.

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 more efficient as delegated monitor. Meanwhile, financing a smaller project requires a fewer times of repetition in monitoring the firm's financial report in the state of bad output.[14] Larger firms also find bonds a favorable means of finance. Their larger net worth allows them to raise a sufficient amount of capital without utilizing costly monitoring.

In the model, a higher deposit rate moves the market towards more bank loans and less private lending and bond finance. We also show that loosening the supply of loanable funds – the quantity of which affected by the supply of money in the economy – shifts the equilibrium composition of the market away from bonds and private lending and towards bank loans; and tightening the supply of loanable funds squeezes out bank lending while expanding monitored private lending and bond finance.

We use the model to evaluate the effects of the recent reforms of banking regulations, specifically those related to the lifting of the deposit and lending rate controls. The model

[14]An empirical literature relates bank loans with state ownership. Allen and Qian (2014) show that the majority of the bank credit goes to state-owned firms in China. Song et al. (2011) show that state-owned firms finance more than 30 percent of their investments through bank loans, compared to less than 10 percent for private firms. Dollar and Wei (2007) report that private firms rely significantly less on bank loans and more on retained earnings and family and friends to finance investments. Now given that most larger firms are state owned, one could speculate that it is the state ownership that gives rise to the observation that most bank loans go to the larger firms – a theory that would need further theoretical construction. There is no state ownership in our model. Instead of relying on state ownership for explaining the data, we argue that the standard theory of banking is sufficient for explaining why banks prefer larger firms.

suggests that removing the controls on the loan rate, which took place in 2004, results in a decline in banking, while at the same time increasing bond finance but reducing private lending. This is consistent with and offers a potential theoretical explanation for the observed decline in banking and the rise of the bond market in China, as shown in Figure 1. The model also suggests that removing all interest rate controls would result in a higher interest rate, crowding out private lending.

Most of the model's predictions are testable, a subset of which are taken to the data to show that they are largely consistent with empirical evidence.

1.3 The literature

Our study builds on the models of costly state verification that are based on Townsend (1979) and Gale and Hellwig (1985). We also build directly on the idea of Diamond (1984) to view financial intermediaries or banks as delegated monitors. Existing theories of financial contracting and intermediation that follow the same ideas include, among others, Boyd and Prescott (1986), Williamson (1986, 1987), Greenwood and Jovanovic (1990), and Greenwood, Sanchez and Wang (2010, 2013). In modeling delegated versus non-delegated monitoring, we offer a novel specification which divides the total cost of monitoring between a fixed component that depends only on the size of the investment, and a variable component which depends also on the measure of lenders providing external finance.

Our work is related also to the larger literature on banking and financial markets. Take Holmström and Tirole (1997) for example, due to moral hazard, only a fraction of external capital can be financed directly by individual investors, the rest must be financed with the participation of monitors (banks). Two elements of our model, however, make it differ from most studies in the literature. First, three asset markets (for monitored bank loans, monitored private contracts, and non-monitored bonds respectively) endogenously coexist in our model. Second, the assumptions of monopoly banking and interest rate regulations give our model a "Chinese look".

There is a literature that studies the coexistence of formal and informal finance in credit markets, including Hoff and Stiglitz (1998) and Stein (2002). Most papers in this literature share the notion that informal lenders hold information advantages over banks in small-business lending, which relies heavily on "soft" information that cannot be directly verified by agents who do not have connections with it. This assumption is in contrast with ours, which holds that the bank has an absolute advantage over private lenders in monitoring the firm's output, but this advantage is comparatively small with smaller firms.

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 strong growth of China's private sector economy. Elliott, Kroeber and Qiao (2015) show that despite its rapid growth, shadow banking remains less important than formal banking as a source of credit in China (as Figure 1 suggests). Besides, they estimate that about two thirds of shadow banking in China results from regulatory arbitrage of the banks. Wang et al. (2015) build an equilibrium model in which commercial banks use shadow banking to evade the restrictions on deposit rates and loan quantities. They argue that shadow banking is able to correct policy distortions and improve social surplus. 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 (2016) study a specific component of shadow banking in China – the wealth management product (WMP) of commercial banks. In a Diamond-Dybvig type banking model with small and big banks, and regulations on deposit rates and the loan-to-deposit ratio, they show a tight loan-to-deposit ratio induces small banks to use WMP for poaching deposits from big banks.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 studies the optimal contracts for financial lending. Section 4 defines and studies the model's general equilibrium. Section 5 studies the effects of the interest rate reforms that were implemented by the PBC over the last ten years. Section 6 takes the model to the data to test some of its major predictions. Section 7 concludes the paper. Proofs 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 in the model, M units of them consumers (investors) and μ units firms (entrepreneurs). Firms are risk neutral and maximize expected profits in period 1. Consumers have the following utility function: $u(c) = c$ where $c\,(\geq 0)$ is consumption in period 1.

Each consumer is endowed with 1 unit of the good in period 0. Firms differ in their capital endowment, k, which is uniformly distributed over the interval $[0, \bar{k}]$ across individual entrepreneurs, with $\bar{k} > 0$. Each entrepreneur is also endowed with an investment project with which any $X(\geq 0)$ units of capital invested in period 0 would return $\tilde{\theta}X$ 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 consumers and offers loans to entrepreneurs. 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 $R_L \in (R_D, \theta_2)$.[15]

Each consumer is free to lend indirectly through the bank, at the fixed interest rate R_D, or directly to individual entrepreneurs through a private lending market. Likewise, each entrepreneur can either borrow from the bank, or directly from individual consumers in the private lending market. For convenience, assume entrepreneurs cannot obtain finance simultaneously from both the bank and a set of individual lenders, and consumers cannot participate in both markets either.

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

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

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 $\gamma_0 + \gamma < \theta_1$.

Observe that equation (1) covers both the case of delegated monitoring, with $\Delta = 0$, and that of non-delegated monitoring, with $\Delta > 0$. Observe also 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.

Given equation (1) then, the bank is always more efficient than individual consumers in lending, as long monitoring is involved.

[15] 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.

3 Optimal Lending

Let r^* denote the market rate of (net expected) return on lending for individual consumers – an endogenous variable whose value will be determined in the equilibrium of the model. Obviously then, $r^* \in [R_D, E(\theta))$. More specifically, if both direct and bank lending are active at the same time, it must hold that $r^* = R_D$. If there is active direct lending but not bank lending, 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 consumers are lenders. Entrepreneurs are free to participate in either side of the market. However, given $r^* < E(\theta)$, it is never optimal for any entrepreneur to lend any fraction of his net worth to the market, directly or indirectly. In the following analysis, therefore, we will take as given that all entrepreneurs are a borrower.

3.1 Direct Lending

Consider first the market where individual consumers/investors lend directly to firms, not through the bank. Consider an individual entrepreneur in this market, with net worth k. To obtain finance, he offers a contract to potential lenders. Assuming deterministic monitoring, the contract takes the form of

$$\sigma_D(k) = \{X(k), S(k), r_1(k), r_2(k)\},$$

where $X(k)$ is the size of the project ($L(k) = X(k) - k$ the size of external finance); $r_i(k)$ is the repayment per unit of the loan in output state $\theta_i,\ i = 1, 2$; and $S(k)$ is the set of reported output states in which the lender monitors the borrower's report – his monitoring policy.

It is straightforward to show that the optimal contract has $S(k) = \emptyset$ or $S(k) = \{\theta_1\}$.[16] In what follows, these two cases are considered separately before the optimal contract is derived.

3.1.1 Non-monitored Direct Lending

Consider first the case where finance is obtained through a contract that prescribes no monitoring, or $S(k) = \emptyset$. In this case, to induce truth telling the entrepreneur's payment to the lender must be constant across the states of output, that is, $r_1(k) = r_2(k) = r_{\mathrm{N}}(k)$, and the entrepreneur's value is given by

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

[16] See the appendix (Section 7.1) for the proof.

subject to

$$r_{\mathrm{N}} L \leq \theta_1 (L + k), \tag{2}$$

$$r_{\mathrm{N}} \geq r^*. \tag{3}$$

Equation (2) is limited liability: total repayment of the loan cannot exceed total output. Equation (3) is individual rationality: the lender must get a rate of return on lending not lower than what the market offers.

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

$$L_{\mathrm{N}}(k) = \frac{\theta_1 k}{r^* - \theta_1}, \quad X_{\mathrm{N}}(k) = \frac{r^* k}{r^* - \theta_1}. \tag{4}$$

With no monitoring, the optimal way to raise finance is to issue a risk-free bond that pays the market interest rate r^*. Notice that at the optimum, constraint (2) binds. That is, in the low output state the repayment of loan is just equal to total output and the entrepreneur's compensation is zero. This allows the firm to raise the maximum amount of finance that the limited liability constraint permits. With the optimal contract, the firm's expected 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. That is, conditional on no-monitoring, a larger entrepreneur net worth supports more finance, a larger project, and higher firm value.[17]

3.1.2 Monitored Direct Lending

Alternatively, the firm could raise finance with a contract that involves investor monitoring: $S(k) = \{\theta_1\}$, in which case the problem of optimal contracting is

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

subject to

$$0 \leq r_1 L \leq \theta_1 (L + k) - \widetilde{C}(L, k), \tag{5}$$

$$0 \leq r_2 L \leq \theta_2 (L + k), \tag{6}$$

[17]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.

$$\theta_1(L+k) - r_1 L - \widetilde{C}(L,k) \geq \theta_1(L+k) - r_2 L, \tag{7}$$

$$\pi_1 r_1 + \pi_2 r_2 \geq r^*, \tag{8}$$

where

$$\widetilde{C}(L,k) = \begin{cases} C(L, L+k) = \gamma_0(L+k) + \gamma L(L+k), & \text{if } L > 0 \\ 0, & \text{if } L = 0 \end{cases}. \tag{9}$$

In the above, equations (5) and (6) are limited liability – what the entrepreneur pays to the lender cannot exceed his total output. Equation (7) is incentive compatibility. Note that given $S(k) = \{\theta_1\}$, the contract must only ensure that the entrepreneur has no incentives to report θ_2 when the true output is θ_1. Equation (8) is a participation constraint. Last, equation (9) says that the cost of monitoring is $C(L, L+k)$ if lending takes place, zero if not.

Monitoring has two effects on the firm's value, one direct, the other indirect, both increasing in the firm's net worth k. The direct effect is that monitoring is costly, and, all else equal, the cost is increasing in k. This reduces the firm's value. The indirect effect is that monitoring, by entering the incentive constraint, affects the firm's ability in repaying its debt and thus its value. To understand this, remember that with no monitoring, truth-telling imposes $r_1 = r_2$. With monitoring, the truth-telling constraint (7) requires instead

$$r_2 - r_1 \geq \widetilde{C}(L,k)/L \geq 0. \tag{10}$$

That is, under monitoring, truth telling imposes a gap between r_1 and r_2, and the size of this gap is increasing in the cost of monitoring, $\widetilde{C}(L,k)$. To focus on the effect of k, fix L. With a smaller k (smaller $\widetilde{C}(L,k)$), a less tight incentive constraint (10) gives the investor larger flexibility in collecting loan repayments, increasing potentially the size of lending and thus the value of the firm. On the other hand, lending is more tightly constrained for a larger k. In particular, when k or the cost of monitoring is sufficiently large, (10) is likely to be binding, or simply infeasible for the contract to implement (remember r_1 must be non-negative and r_2 must not exceed θ_2). This, again, affects adversely the value of the firm. To summarize, the model suggests that monitoring goes better with a smaller rather than a larger k.

Last, remember, as discussed earlier, because there is no coordination and information exchange among individual lenders, each of them incurs on the firm a monitoring cost of $\gamma(L+k)$ to verify the report of θ_1. This repetition in monitoring then implies that, in the state of low output, the total cost of monitoring incurred increases more than linearly in the size of the project, amplifying the effects we have just discussed. This is another aspect of the model which suggests that monitored direct lending is more efficient with firms smaller in k.

3.1.3 Optimal Direct Lending

The entrepreneur'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) of the assumption then assumes that the deposit rate is sufficiently high.[18] Remember $R_D < E(\theta)$.[19]

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:*

$$L_{\mathrm{M}}(k) = \frac{E(\theta) - \pi_1\gamma k - \pi_1\gamma_0 - r^*}{2\pi_1\gamma}, \tag{11}$$

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

$$r_1(k) = \frac{(\theta_1 - \gamma_0)X_{\mathrm{M}} - L_{\mathrm{M}}(k)\gamma X_{\mathrm{M}}(k)}{L_{\mathrm{M}}}, \tag{13}$$

$$r_2(k) = \frac{r^* - \pi_1 r_1(k)}{\pi_2}, \tag{14}$$

and the value of the entrepreneur is

$$V_{\mathrm{M}}(k) = \frac{[E(\theta) + \pi_1\gamma k - \pi_1\gamma_0 - r^*]^2}{4\pi_1\gamma} + kr^*, \tag{15}$$

and $\tilde{k}$ solves

$$V_{\mathrm{M}}(\tilde{k}) = V_{\mathrm{N}}(\tilde{k}).$$

The determination of $\tilde{k}$ is illustrated in Figure 8.[20] With the optimal contract, we have

$$X(k) = \begin{cases} X_{\mathrm{M}}(k), & \forall k < \tilde{k} \\ X_{\mathrm{N}}(k), & \forall k \geq \tilde{k} \end{cases} \tag{16}$$

[18] 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 $k \in [0, \tilde{k}']$.

[19] Note it holds that $\pi_2\theta_2 - \pi_1\theta_1 + \pi_1\gamma_0 < E(\theta)$.

[20] More specifically $\tilde{k}$ must solve $(E(\theta) + \pi_1\gamma\tilde{k} - \pi_1\gamma_0 - r^*)^2/(4\pi_1\gamma) + \tilde{k}r^* = \pi_2(\theta_2 - \theta_1)(\tilde{k}r^*)/(r^* - \theta_1)$, which has a unique solution for $\tilde{k} \in (0, \bar{k})$.

and

$$V(k)=\begin{cases}V_{\mathrm{M}}(k), & \forall k<\tilde{k}\\ V_{\mathrm{N}}(k). & \forall k\geq\tilde{k}\end{cases}. \tag{17}$$

Proposition 2 says that in the market of direct finance, larger firms issue bonds for external finance, while smaller firms use mechanisms that involve monitoring. This is consistent with the findings in Didier and Schmukler (2013) that in China, firms with equity issues (with average employment 2527) are much smaller than that with bond issues (with average employment 4188).[21] In general, equity holders play more active roles in monitoring the management of their investment than the public who hold the firm's commercial paper.

From Proposition 2 and Lemma 1, at the optimum a larger k supports a larger X and larger firm value, monitoring involved or not. This, of course, is anticipated, given our earlier analysis. Specifically, conditional on no monitoring, a larger k increases the entrepreneur's ability in delivering a required debt repayment (in the state of low output), using the firm's net worth as a collateral to support lending. This same effect exists also in the case of monitoring.

The fact that a larger k make finance with monitoring less efficient relative to that with no monitoring is also anticipated, given our earlier discussion.[22]

Remember from Lemma 1 that if the optimal contract prescribes no monitoring (i.e., $k\geq\tilde{k}$), the interest rate is constant and equal to r^* across the output states. In Corollary 7 in the appendix, we show that the optimal direct lending contract has 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, there is spread in interest rate between the two output states, and the spread shrinks as the entrepreneur's net worth grows. Also, for any fixed $k\in[0,\tilde{k}]$, the optimal contract has that $r_1(k)$ is larger when r^* is larger. This holds for a larger r^* reduces the optimal size of the investment, which, in turn, increases the efficiency in monitoring and allows for higher lender returns in the low output state.

Corollary 3. *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 $k\in(\tilde{k},\bar{k}]$.*

[21] See Table 1 in their paper.

[22] In addition, note that conditional on monitoring, a larger k increases the cost of monitoring per unit of external finance, which is given by

$$\frac{C(L,X)}{L}=\frac{\gamma_0 X+\gamma LX}{L}=\gamma_0\frac{X}{L}+\gamma X,$$

where L, the optimal amount of external finance raised, is decreasing in k.

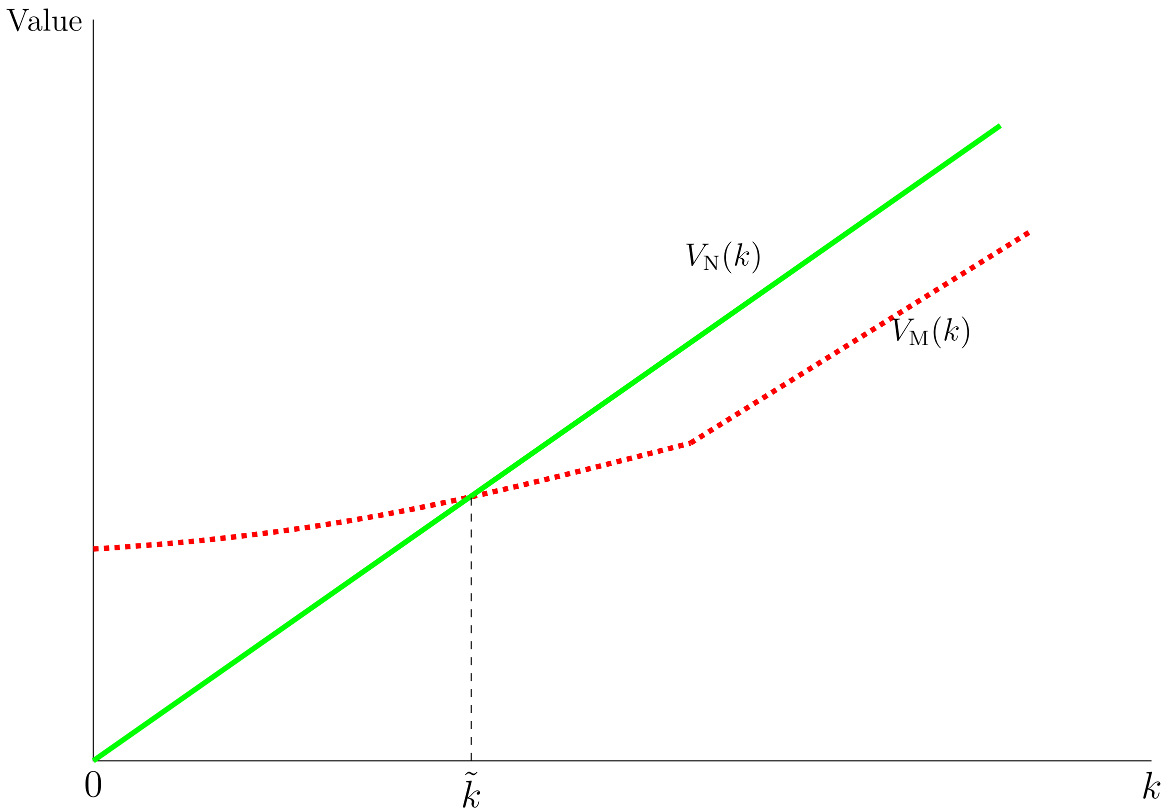


Figure 8: Lender's value functions in direct lending

In other words, on average smaller (in k) firms are more valuable per unit of equity, and they also borrow more relative to equity.[23] This, again, results from the relative inefficiency in monitoring a larger firm. A larger k allows the firm to finance a larger project (larger $X(k)$) which, in turn, implies more duplication in the cost of monitoring. More specifically, conditional on monitoring, the firm's value is

$$\begin{aligned} V_{\mathrm{M}}(k) &= \pi_1\left[\theta_1 X - r_1 L - \gamma_0 X - L\gamma X\right] + \pi_2\left[\theta_2 X - r_2 L\right] \\ &= E(\theta)X - \left[r^* L + \pi_1\gamma_0 X + \pi_1\gamma X L\right]. \end{aligned} \tag{18}$$

The second part of the RHS of the above equation is the total cost of external finance which, in turn, consists of two parts. The first part, $r^* L$, is the reservation return for the lenders. The second part, $\pi_1\gamma_0 X + \pi_1\gamma X L$, is the expected cost of monitoring. Suppose the size of the project (X) increases. Then not only the total cost of monitoring would increase, the marginal cost of that ($\pi_1\gamma X$) would also increase. This gives rise to Corollary 3.

[23] Kato and Long (2006) show empirically that smaller firms in China enjoy higher profitability than larger firms, consistent with the prediction of our model.

Obviously, monitoring allows the contract to support more external finance and the firm to fund a larger investment. In the appendix (Corollary 8), we show that with the optimal direct lending contract, $X_{\mathrm{M}}(k) > X_{\mathrm{N}}(k)$, for all $k \in [0, \tilde{k}]$. This explains the jump in the optimal size of the funded project as a function of k, $X(k)$, at $\tilde{k}$ (see Figure 17).

3.2 Intermediated/Bank Finance

Let $D(\geq 0)$ denote the bank's total deposits from consumers/investors. This is also 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. Notice that we need only study the case of $r^* = R_D$, for otherwise (i.e., $r^* > R_D$) no one lends through the bank and $D = 0$.

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 $L(k) \equiv Z(k) - k$, or equivalently the size of the entrepreneur's project $Z(k)$, and a policy for monitoring the firm's report of output.

Let $\mathbf{B}$, a subset of $[0, \bar{k}]$, denote the set of all entrepreneurs whom the bank is willing to offer a loan to. For each $k \in \mathbf{B}$, the loan must ensure that the entrepreneur 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.

Consider the bank's monitoring policy. Fix $k \in \mathbf{B}$. With the optimal contract, monitoring occurs if and only if the lower output θ_1 is reported. 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

$$L(k) \leq \frac{\theta_1 k}{R_L - \theta_1}, \tag{19}$$

where the right hand side gives the maximum size of the credit the firm could raise with the bank. Given this, the expected value of the firm, which is $E(\theta)(k + L(k)) - R_L L(k)$, is strictly less than $V(k)$.[24] In other words, if the bank never monitors the entrepreneur's report, it would

[24]Specifically,

$$E(\theta)(k + L(k)) - R_L L(k) \leq (E(\theta) - \theta_1)\frac{R_L}{R_L - \theta_1}k < (E(\theta) - \theta_1)\frac{R_D}{R_D - \theta_1}k = V_{\mathrm{N}}(k) \leq V(k),$$

not be able to induce the firm to participate – it could not offer a loan that is sufficiently large to make the entrepreneur better off with a bank loan than with direct lending.

Given the above, the bank's problem becomes

$$\max_{\mathbf{B},\{L(k)\}_{k\in\mathbf{B}}} \mu \int_{\mathbf{B}} \Big\{ \pi_1 \left(\theta_1 - \gamma_0\right) \left(k + L(k)\right) + (\pi_2 R_L - 1) L(k) \Big\} dG(k) + D - R_D D \tag{20}$$

subject to

$$\mathbf{B} \subseteq [0, \bar{k}], \tag{21}$$

$$L(k) \geq 0, \quad \forall k \in \mathbf{B}, \tag{22}$$

$$\mu \int_{\mathbf{B}} L(k) dG(k) \leq D, \tag{23}$$

$$V_b(k, L(k)) \equiv \pi_2 \left\{\theta_2(k + L(k)) - R_L L(k)\right\} \geq V(k), \quad \forall k \in \mathbf{B}, \tag{24}$$

where equation (23) is a resource constraint: total loans made cannot exceed the total supply of bank credit; (24) is a participation constraint: the firms in $\mathbf{B}$ are better off obtaining finance from the bank than from individual lenders directly.

Rewrite (24) as

$$L(k) \geq L_0(k), \ \forall k \in \mathbf{B}. \tag{25}$$

where

$$L_0(k) \equiv \frac{V(k) - \pi_2 \theta_2 k}{\pi_2(\theta_2 - R_L)}, \ \forall k \in [0, \bar{k}], \tag{26}$$

$$Z_0(k) \equiv k + L_0(k) = \frac{V(k) - \pi_2 R_L k}{\pi_2(\theta_2 - R_L)}, \ \forall k \in [0, \bar{k}]. \tag{27}$$

Clearly, $L_0(k)$, derived from the entrepreneur's participation constraint, is the entrepreneur's reservation loan size – the minimum size of the loan with which it is willing to borrow from the bank, and $Z_0(k)$ is the corresponding size of the project. Given the nature of the loan contract (that the entrepreneur is paid 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 $L_0(k)$) can induce the firm to participate.

where the first inequality is from (19), the second holds because $R_L > R_D$.

From (26), 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 entrepreneur's outside value $V(k)$, requiring a lager loan for inducing him to participate. Overall, however, it is shown that $L_0(k)$ and $Z_0(k)$ are increasing in k.[25]

Notice that $V_b(k, L_0(k)) = V(k)$. That is, at the minimum loan the firm is willing to take from the bank, the firm is indifferent between raising finance from the bank and borrowing directly from individual lenders.

Let

$$D_1 \equiv \mu \int_0^{\bar{k}} L_0(k) dG(k), \tag{28}$$

$$D_0 \equiv \mu \int_{\tilde{k}}^{\bar{k}} L_0(k) dG(k). \tag{29}$$

In words, D_1 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 $k \in [\tilde{k}, \bar{k}]$. Remember firms with $k \geq \tilde{k}$ would be able to issue bonds to obtain direct finance, if a bank loan is not available.

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

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

Proposition 4. *The following holds under Assumption 2. (i) Suppose $0 \leq D < D_0$. Then the bank's optimal plan has*

$$L_{\text{B}}(k) = L_0(k), \ \forall k \in \mathbf{B},$$

where $\mathbf{B}$ is any subset of $[\tilde{k}, \bar{k}]$ that solves

$$\mu \int_{\mathbf{B}} L_0(k) dG(k) = D. \tag{30}$$

(ii) Suppose $D_0 \leq D < D_1$. Then it is optimal for the bank to set $\mathbf{B} = [\hat{k}, \bar{k}]$, with

$$L_{\text{B}}(k) = L_0(k), \ \forall k \in [\hat{k}, \bar{k}],$$

[25] From (17) we have $V'(k)$ is weakly increasing in k for $k \in [0, \bar{k}]$. Then, from Assumption 1, $V'(k) \geq V'(0) = r^* + \frac{1}{2}[E(\theta) - r^* - \pi_1\gamma_0] > \pi_2\theta_2 > \pi_2 R_L$. Thus $Z_0'(k) > 0$ and $L_0'(k) > 0$. See appendix (Section 7.8) for more on this.

where $\hat{k}$ *solves*

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

(iii) Suppose $D \geq D_1$*. Then the optimal plan for the bank is to set* $\mathbf{B} = [0, \bar{k}]$*, and with* $\{L_{\mathrm{B}}(k), k \in \mathbf{B}\}$ *be any function that satisfies (22) and (23).*

To understand the proposition, consider the bank's return on lending to firm k in an amount of L, with $L \geq L_0(k)$:

$$\begin{aligned} R_b(k, L) &\equiv \frac{\pi_1(\theta_1 - \gamma_0)(k + L) + \pi_2 R_L L}{L} - R_D \\ &= \pi_1(\theta_1 - \gamma_0)\frac{k}{L} + \pi_1(\theta_1 - \gamma_0) + \pi_2 R_L - R_D, \end{aligned} \tag{31}$$

where the term $\pi_1(\theta_1 - \gamma_0)\frac{k}{L}$, 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 on lending. A larger L, on the other hand, dilutes the gains from utilizing the firm's net worth as collateral for enforcing more repayments in the state of low output, reducing the bank's returns per unit of lending.

In (i) and (ii) where $D < D_1$, there is not enough funds to finance all firms, any external 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.[26] In these cases, what the bank seeks, essentially, is to maximize the number of loans made, by making each loan as small as possible.[27]

Equation (31) also indicates the bank should in general prefer larger to smaller firms. More specifically, given (26) and Corollary 3,

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

In other words, for firms with $k \in [0, \tilde{k}]$, the bank strictly prefers the larger, and between firms with $k \in [\tilde{k}, \bar{k}]$, it is indifferent.

[26]For this, see the appendix (Step 3 in Section 7.7) for related calculations.

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

More specifically, when $0 \leq D < D_0$, the supply of bank credit is so tight that only a subset of firms with $k \geq \tilde{k}$ could get a bank 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 $\mathbf{B} = [\hat{k}_1, \hat{k}_2]$, where $0 \leq \hat{k}_1 < \hat{k}_2 \leq \bar{k}$ (see Figure 9).[28]

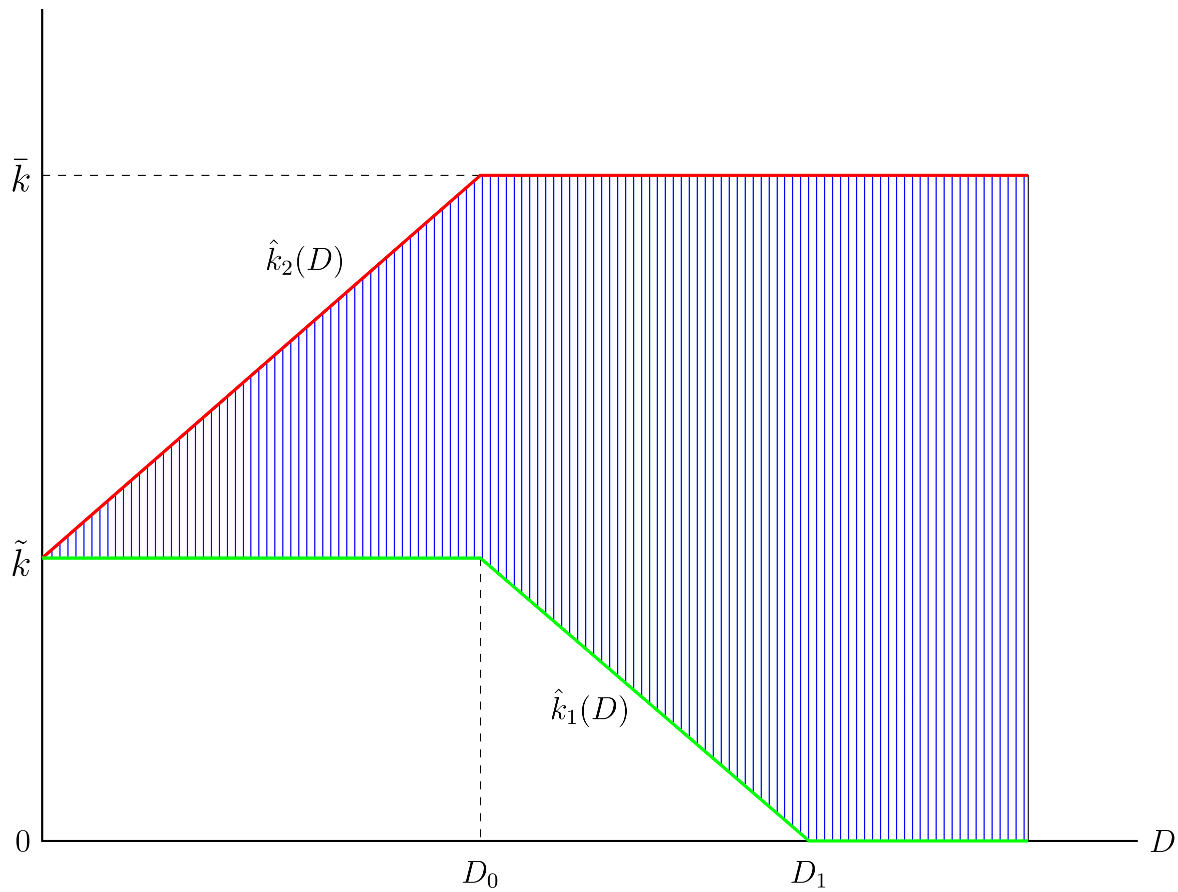


Figure 9: The optimal bank lending set $\mathbf{B}$ conditional on deposit D

In the case $D_0 < D < D_1$, 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

[28]Note, however, that this *rationing* does not imply that those obtaining bank loans are better off than those who do not. In fact, the firms are indifferent in value 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).

giving each of them a loan with their reservation size $L_0(k)$.

Last, in the case $D \geq D_1$, 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 in this case to (i) meet the minimum demand for credit from each firm, and then (ii) lend the rest of the funds to an arbitrary set of firms, on top of their $L_0(k)$. Here (ii) 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 in D and $\hat{k}_2(D)$ is increasing in D, as Figure 9 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 finance.

3.3 Direct vs. Bank Lending

Being more efficient in monitoring, what outcomes, in particular in the size of the external finance it supports, would the bank achieve relative to direct lending? We show that if R_D is sufficiently low, then bank lending always support 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.[29]

The intuition for these results would touch the difference between the two lending mechanisms. On the one hand, while R_L is fixed for bank loans, parties in direct lending are free to adjust the terms of their contract to reflect market conditions. This gives direct lending an upper hand over bank loans. On the other hand, being more efficient in monitoring gives bank loans an advantage over direct lending.[30] 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 entrepreneur 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 (who lends either indirectly through the bank, or directly to a firm), reducing the value of the firm, $V(k)$. This, given the

[29]See Lemma 9 in the appendix.

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

fixed loan rate R_L, puts less pressure on the bank in offering a larger loan for inducing the firm to participate, reducing the size of the loan. A higher R_D also reduces the size of direct finance $(X(k)-k)$. However, since the parities in direct lending are free to adjust the interest rates in the lending contract, the reduction in size of direct finance would be less than that in the bank loan.[31] Overall, therefore, an increase in R_D would result in smaller bank loans relative to monitored private loans.

4 Equilibrium

Definition 1. *A rational expectations equilibrium of the model consists of a market rate of return on lending for consumers r^*, a quantity of deposits D^*, a set $\mathbf{B} \subseteq [0, \bar{k}]$ of entrepreneurs 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 only the market for direct lending opens, with $D^* = 0$ and $\mathbf{B} = \emptyset$.*
4. *The demand for loans equals the supply of loans in the direct lending market:*

$$\mu \int_{[0,\bar{k}]\setminus\mathbf{B}} [X(k) - k]\, dG(k) = M - D^*. \tag{34}$$

[31] To see this more precisely, remember, for any fixed k, in order to induce the firm to participate, $Z_0(k)$ must satisfy

$$\pi_2 \{\theta_2 Z_0(k) - R_L [Z_0(k) - k]\} = V(k). \tag{32}$$

A higher R_D decreases $V(k)$ which, given that R_L is fixed, forces the bank to decrease $Z_0(k)$ in order to decrease the entrepreneur's value on the left hand side of the equation to make it hold. On the other hand, for direct lending, from equations (16) and (17), $X(k)$ must satisfy

$$\pi_2 [\theta_2 X(k) - r_2(k)(X(k) - k)] = V(k). \tag{33}$$

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 $X(k)$.

The above defined equilibrium of the model is formulated more explicitly in a system of equations in the appendix (Section 7.14). We now characterize the outcomes of this equilibrium. To save space, we assume in this rest of the paper $R_D < \bar{R}_D$. [32]

The bank's 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 credit market.[33] 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:

$$\begin{aligned} Q(D, r^*) &= \mu \int_0^{\hat{k}_1(D,r^*)} L_{\mathrm{M}}(k, r^*) dG(k) + \mu \int_{\hat{k}_1(D,r^*)}^{\hat{k}_2(D,r^*)} L_{\mathrm{B}}(k, r^*) dG(k) \\ &+ \mu \int_{\hat{k}_2(D,r^*)}^{\bar{k}} L_{\mathrm{N}}(k, r^*) dG(k). \end{aligned} \tag{35}$$

This is the sum of the demand for direct finance with monitoring, bank loans, and bond finance. Note that the second part of the sum, the demand for bank loans, is equal to D, as the bank's resource constraint binds.

Figure 10 depicts $Q(D, r^*)$ on the D dimension and conditional on $r^* \geq R_D$.[34] Consider first the case of $r^* > R_D$. In this case, there is no bank lending in equilibrium and 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 (4) and (11), are both decreasing in the interest rate 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 $Q(0, r^*)$ with $r^* = R_D$ so that the demand for external finance achieves its maximum conditional on $D = 0$.

What happens in the direct lending market in this case is depicted in Figure 11, where a value of M below $\underline{Q}$ induces an equilibrium interest rate r^* to clear the market. Observe

[32] An earlier version of the paper, available by request, includes also an analysis for the case of $R_D \geq \bar{R}_D$. Similar outcomes arise between the two cases but the data looks more consistent with the one we choose to present, as to be shown later in the paper.

[33] 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.

[34] 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.

that for M sufficiently small, $M \leq \underline{M}$ specifically, the equilibrium interest rate 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 (4) and (11)). That is, a sufficiently high interest rate, which results from a sufficiently small supply of external finance M, would render monitoring being completely crowded out and the risk free bond being the only financial instrument used in equilibrium.[35]

Consider next the case of $r^* = R_D$. In this case, D could take any value from $(0, M]$. In the appendix, Lemma 10, we show that $Q(D, R_D)$ is strictly increasing in D at all $D \in (0, M]$, as depicted in Figure 10, where $Q_0 \equiv Q(D_0, R_D)$ and $Q_1 \equiv Q(D_1, R_D)$. If $D > D_1$, all firms raise credit through the bank, with $Q(D, R_D) = D$. If $0 < D < D_1$, 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.[36]

With these, four cases emerge from Figure 10, in 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, the equilibrium of the model being depicted in Figure 11.

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.

[35] 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

$$L_{\mathrm{M}}(k, r^*) = \max\left\{0, \ \frac{E(\theta) - \pi_1 \gamma k - \pi_1 \gamma_0 - r^*}{2\pi_1 \gamma}\right\},$$

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.

[36] Note that this is conditional on $R_D < \bar{R}_D$ and so bank loans are able to support larger finance relative to direct lending for all k, as depicted in Figure 17 (a), and so the slope of Q in D is positive at all D. Obviously, if $R_D \geq \bar{R}_D$ and Figure 17 (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 7.10.

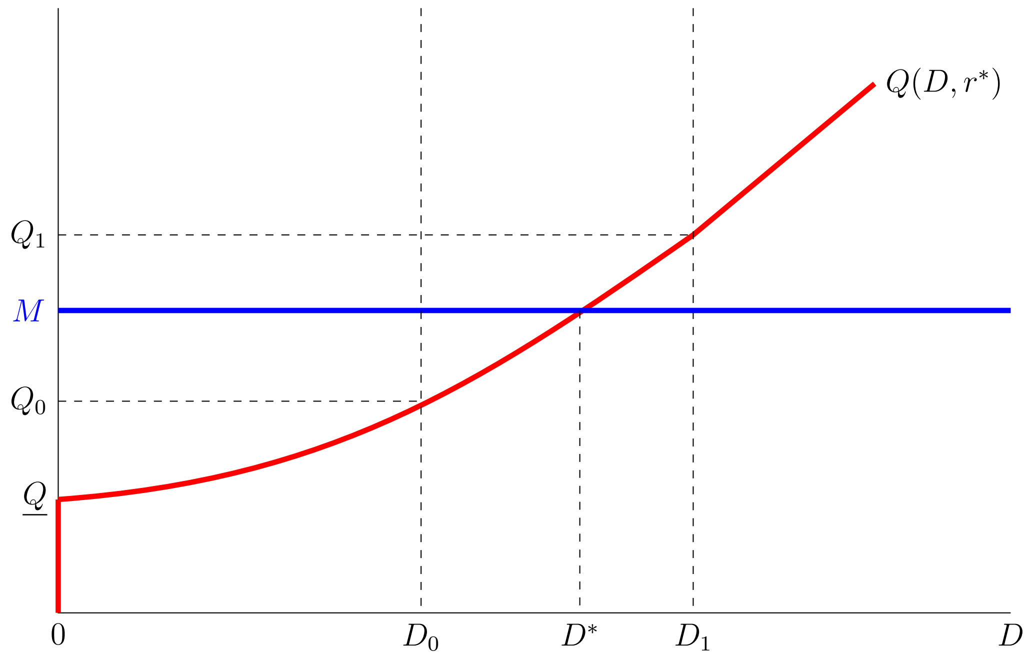


Figure 10: Equilibrium when $0 < M < \underline{Q}$

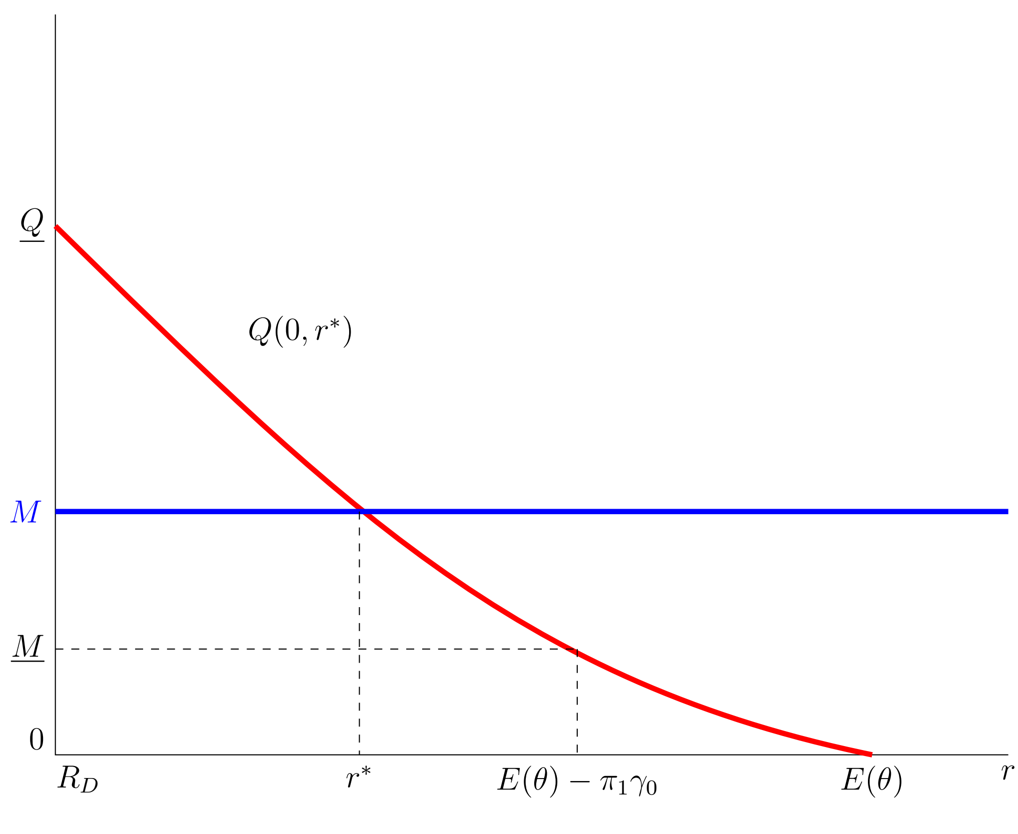


Figure 11: Equilibrium with $D^* = 0$

Case 3: $Q_0 < M < Q_1$. Bank loans and monitored direct finance coexist in the unique equilibrium of the model.

Case 4: $M \geq Q_1$. In equilibrium $D^* \geq D_1$ and, from Proposition 4, all lending takes place indirectly through the bank.

In Cases 2 and 3, where direct lending and bank loans coexist, a larger M implies a higher equilibrium D^*, which, from Figure 9, implies an expanded set of firms obtaining bank loans but a reduced set of firms participating in direct lending. In other words, an increase in the total supply of finance induces a crowding out of direct finance by bank loans: as M increases, D^* is larger while $M - D^*$ is smaller.

So an increase in M reduces the size of direct lending in both absolute and relative measures. Let us think more and look for an interpretation for the mechanisms behind this. Imagine the economy is in an initial equilibrium. Imagine M is increased by a small positive amount Δ. Any positive fraction of this Δ could not have flowed into the market of direct lending, for then the interest rate on direct lending would fall and investors would flow back into bank deposits for the higher deposit rate. In other words, the newly arrived funds must become an addition to the bank's deposits, which now totals $D' \equiv D^* + \Delta$. With 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 demand for credit in the market for direct lending, lowering the interest rate for investors in direct lending, driving them away from direct lending and into bank deposits, until the interest rate on direct lending is restored at R_D. The above described process increases the bank's deposits for the second time, from D' to $D''(> D')$. And this continues, until the bank's deposits settles at its new equilibrium level, which is strictly greater than that of the initial equilibrium.

Observe also that as bank loans crowd out direct lending following the increase in M, the composition of direct lending also changes, for smaller shares of bond finance but larger shares of monitored private lending, from Figure 9.

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 equilibrium outcomes. Figure 12 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, Q_1, and $\underline{Q}$ are all functions of R_D, we write them explicitly as $Q_0(R_D)$, $Q_1(R_D)$ and $\underline{Q}(R_D)$, respectively. These are all decreasing functions and are located relative to each other as the figure depicts.

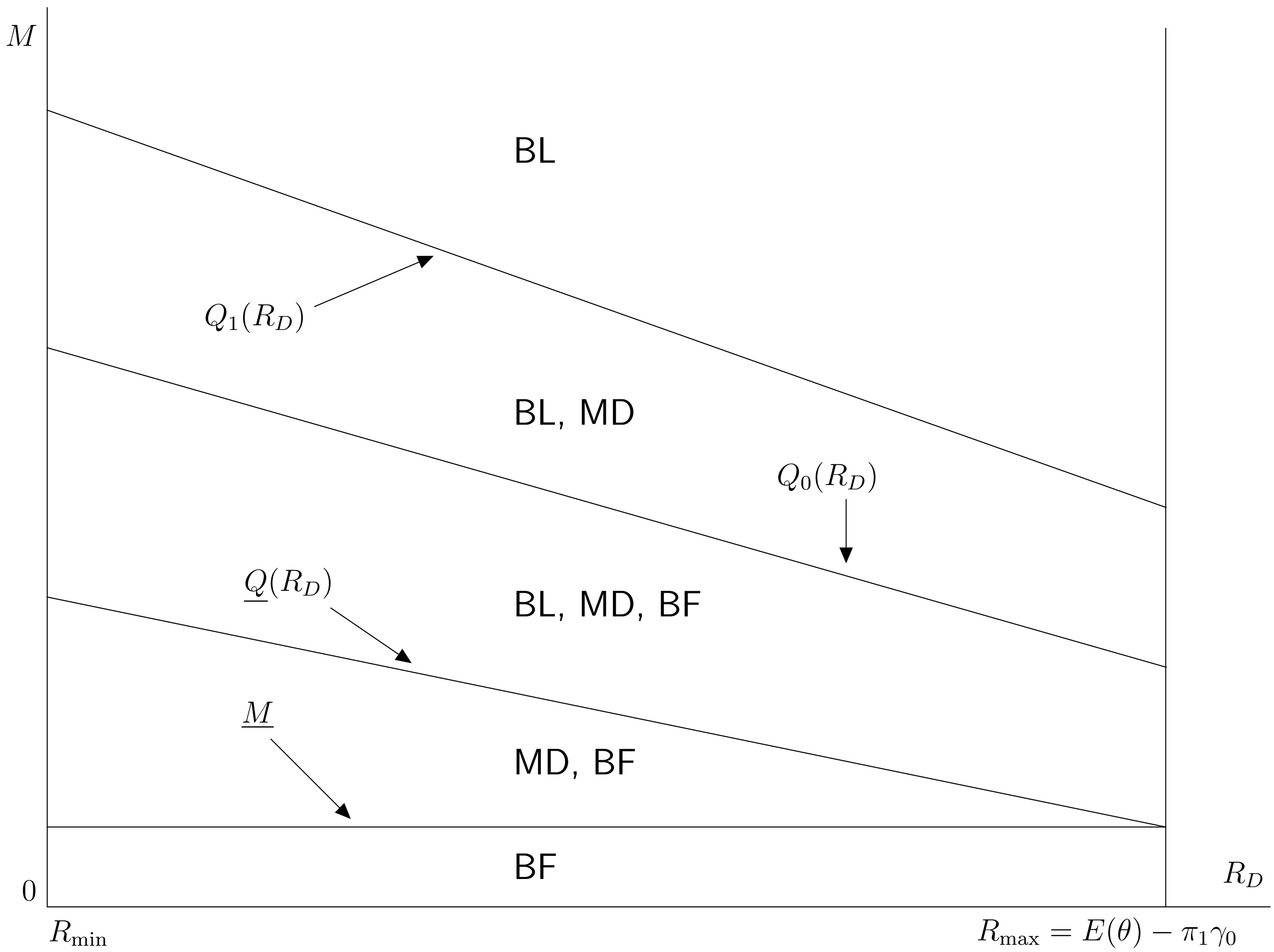


Figure 12: Equilibria with respect to R_D and M

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 pair of 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 in equilibrium bank loans and monitored direct finance coexist.

Figure 12 shows 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 high M crowds out completely the markets for 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 to offer the fixed deposit rate, better ability in competing for deposits from the consumers which, in turn, gives rise to a larger D and more bank loans in equilibrium, at the expense of direct finance.

The figure also shows that, fixing M, a higher R_D moves the market towards (weakly) more (monitored) bank loans and less direct lending. On the one hand, a higher R_D gives the bank stronger 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).

5 Banking Reforms

In this section, we use the model to evaluate, analytically, the effects of the reforms that the central bank of China has implemented, in a sequence of major moves since 2004, in lifting the interest rate controls on commercial bank loans and on deposits.

Given the linearity in the payoff and production functions, and the efficiency of delegated relative to individual monitoring, removing the control on the bank lending rate would result in unbounded investments financed with bank loans. To avoid this, we modify the production function $f(\cdot)$ to make it weakly concave, assuming

$$f(X) = \begin{cases} \tilde{\theta}X, & \text{if } X \leq \bar{X} \\ \tilde{\theta}\bar{X}, & \text{if } X > \bar{X} \end{cases},$$

where $\bar{X}$ is the size of the project beyond which any additional investment would not be productive. Assume $\bar{X}$ is positive and sufficiently large. In particular, we assume $\bar{X} > Z_0(\bar{k})$, so that the outcomes in the prior section continues to hold.[37]

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

To study the effects of the reforms, we suppose $\underline{Q}(R_D) < M < Q_0(R_D)$ so that, consistent with data, all three markets coexist prior to the reforms.

5.1 Removing the lending rate ceiling

In October 2004, the central bank removed its lending rate ceiling on commercial bank loans so that banks are free to charge borrowers any rate above the floor rate, which continues to exist after the reform. To model this, let $\underline{R}_L$ ($\leq R_L$) be the positive floor lending rate. With this, the bank's problem becomes

$$\max_{\mathbf{B},\{Z(k),R_L(k)\}_{k\in\mathbf{B}}} \mu \int_{\mathbf{B}} \left\{ \pi_1(\theta_1 - \gamma_0) Z(k) + \pi_2 R_L(k) [Z(k) - k] \right\} dG(k) \\ + D - \mu \int_{\mathbf{B}} [Z(k) - k] \, dG(k) - R_D D \tag{36}$$

subject to (21), (23) and

$$k \leq Z(k) \leq \bar{X}, \quad \forall k \in \mathbf{B}, \tag{37}$$

$$R_L(k) \geq \underline{R}_L, \quad \forall k \in \mathbf{B}, \tag{38}$$

$$\pi_2 \{\theta_2 Z(k) - R_L(k) [Z(k) - k]\} \geq V(k), \quad \forall k \in \mathbf{B}. \tag{39}$$

As in the benchmark environment, the participation constraint (39) dictates a relationship between the lending rate charged, $R_L(k)$, and the size of the loan, $Z(k) - k$, which, given (37), gives

$$R_L(k) \leq \theta_2 - \frac{V(k) - \pi_2 \theta_2 k}{\pi_2(\bar{X} - k)} \equiv \bar{R}_L(k), \forall k \in \mathbf{B}, \tag{40}$$

$\bar{R}_L(k)$ being the maximum possible lending rate the bank is able to charge on firm k, subject to (37) and (39). It is easy to show that $\bar{R}_L(k)$ is decreasing in k. With a larger k, the entrepreneur'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 entrepreneur at the maximum possible 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.

With Assumption 3, the participation constraint (39) is binding and the bank's rate of return on lending to firm k is

$$\begin{aligned} R_b(k) &= \frac{\pi_1(\theta_1 - \gamma_0)(L(k) + k) + \pi_2 R_L(k)L(k)}{L(k)} - R_D \\ &= E(\theta) - \pi_1\gamma_0 - R_D + \frac{(E(\theta) - \pi_1\gamma_0)k - V(k)}{L(k)}, \end{aligned} \tag{41}$$

where since $(E(\theta) - \pi_1\gamma_0)k - V(k) < 0$ (which holds for all $k \in [0, \bar{k}]$ from (17)), $R_b(k)$ is larger when $L(k)$ is larger. Notice that this is in contrast with what happens in the benchmark model. With a freely adjustable lending rate, the bank is able to collect more repayments per unit of loan in the high output state θ_2. This gives the bank incentives for larger loans. A larger loan also dilutes the net cost of lending to firm k, resulting in a higher average rate of return to the bank.

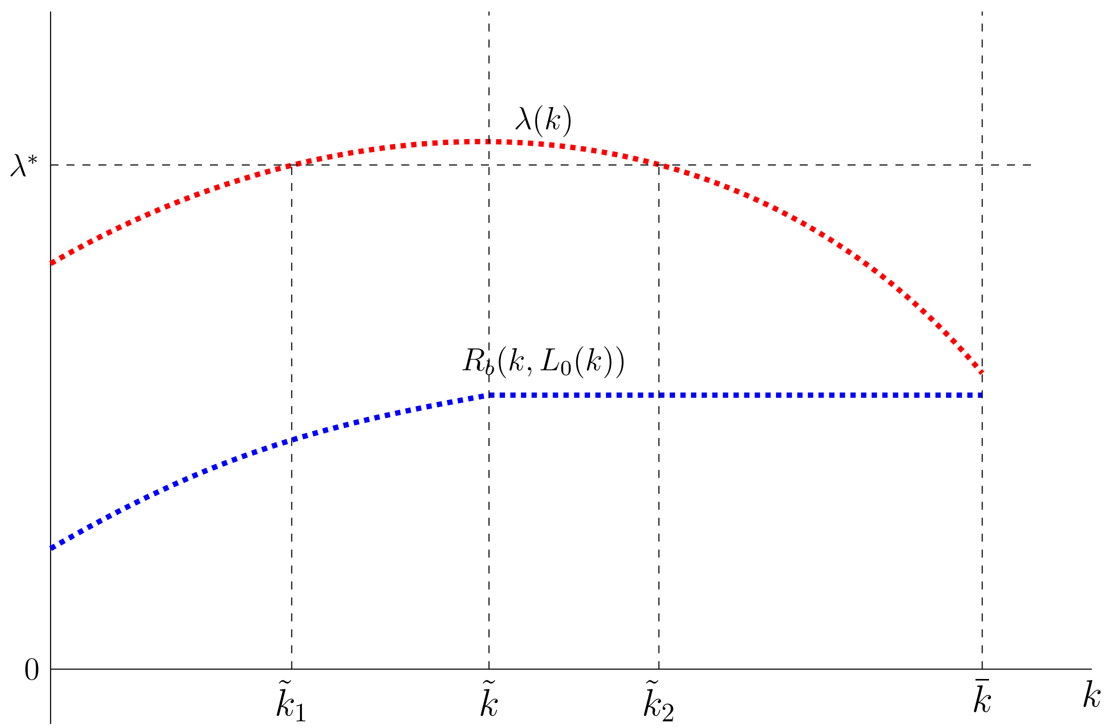


Figure 13: The scenario where $\mathbf{B} = [\tilde{k}_1, \tilde{k}_2]$

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

Thus 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 $R(k) = \bar{R}_L(k)$, defined in (40), to maximize the repayments per unit of loan in the high output state θ_2. Remember, with fixed R_L, the bank wants the loans to be of the minimum possible size. There, by keeping the loans small, the bank lends to more firms, maximizing its benefits from using their net worth as collateral in loan contracting. Here, with a flexible R_L, the bank wants and is able to make larger loans, maximizing its surplus in the state of the high output.

Moreover, for any $k \in [0, \bar{k}]$,

$$\bar{R}_L(k) = \theta_2 - \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\bar{X} - k)} > \theta_2 - \frac{V(k) - \pi_2\theta_2 k}{\pi_2(Z_0(k) - k)} = R_L \geq \underline{R}_L, \tag{42}$$

and so constraint (38) does not bind. With these, the bank's problem is reduced to choosing $\mathbf{B}$ to maximize its total profits subject to the resource constraint (23), and the solution has

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

where

$$\lambda(k) = \frac{(E(\theta) - \pi_1\gamma_0)\bar{X} - V(k)}{\bar{X} - k} - R_D \tag{43}$$

is the bank's expected net rate of return on the loan to firm k, and λ^* is determined by

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

To maximize total profits, the bank would pick the firms with the largest $\lambda(k)$s subject to the total funds available, as depicted in Figure 13,

A larger k has two effects on $\lambda(k)$. First, a larger k implies a larger $V(k)$ and this reduces the returns on lending to firm k. Second, a larger k implies a smaller bank loan $(\bar{X} - k)$, resulting in a higher average net return of lending, which increases $\lambda(k)$. In the appendix (Section 7.11) 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 13.

5.1.1 The distribution of finance

In Figure 13, $\mathbf{B} = [\tilde{k}_1, \tilde{k}_2]$. That is, in equilibrium firms with $k \in [\tilde{k}_1, \tilde{k}_2]$ would be financed with a bank loan, others obtaining finance directly from individual lenders. Moreover, given $0 < \tilde{k}_1 < \tilde{k} < \tilde{k}_2 < \bar{k}$, it follows from Proposition 2 that firms with $k \in [0, \tilde{k}_1)$ would seek monitored private finance, and those with $k \in (\tilde{k}_2, \bar{k}]$ would obtain credit by way of issuing bonds. So the 2004 banking reform should not have changed the general patten of the source distribution of finance across firms, which is that small firms seek monitored private lending, medium sized firms are financed with bank loans, and large firms issue bonds.

As is obvious from Figure 14, 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.

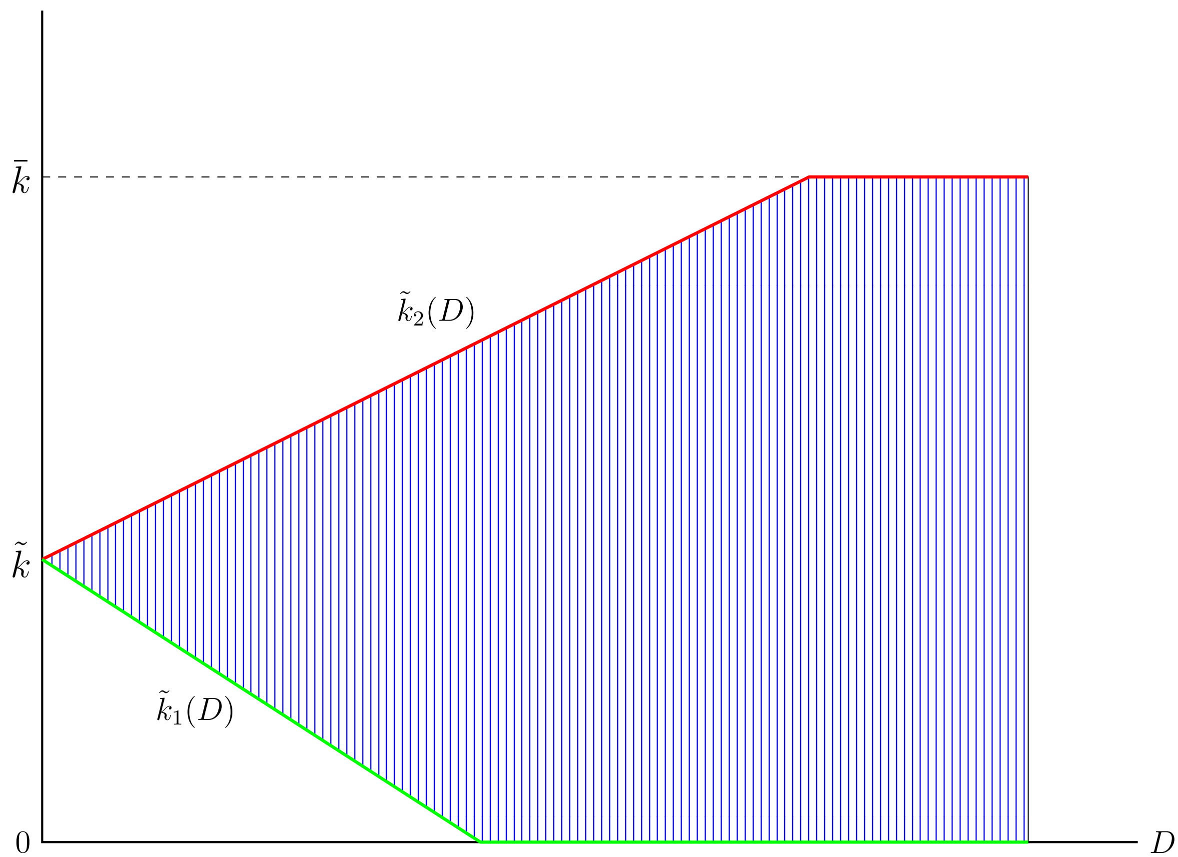


Figure 14: The division of total finance as a function of D.

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

$$\widetilde{Q}(D) = \mu \int_0^{\tilde{k}_1} L_{\mathrm{M}}(k) dG(k) + \mu \int_{\tilde{k}_1}^{\tilde{k}_2} [\bar{X} - k] dG(k) + \mu \int_{\tilde{k}_2}^{\bar{k}} L_{\mathrm{N}}(k) dG(k). \tag{44}$$

As is for $Q(D)$ in (35) in the benchmark case, it is easy to verify that $\widetilde{Q}(D)$ is increasing in

D.[38] The equilibrium bank deposits, denoted $\widetilde{D}^*$, then solves

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

as depicted in Figure 15.

Obviously, a larger M results in a larger $\widetilde{D}^*$ and, from Figure 13, 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 more loanable funds are available in the economy, bank loans would crowd out both monitored private lending and bond finance, as in the benchmark model.

In the appendix (Section 7.12) we show $\widetilde{Q}(D) > Q(D)$ for all $D \in (0, D_0)$ (remember at these Ds all three markets are active in the benchmark model). What happens is that, for fixed 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.

Suppose R_D and M satisfy $\underline{Q}(R_D) < M < Q_0(R_D)$ so that all three markets are active in the benchmark model. Observe then from Figure 15 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 5. *(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 removing the lending rate ceiling, an increase in M increases equilibrium bank deposits and loans, but squeezes bond finance and monitored private lending, as in the case of fixed bank lending rate.*

Part (ii) of the proposition confirms that removing the lending rate ceiling would not alter the direction in which a variation in M induces a change in the size of banking. Look now at the mechanism behind (i) of the proposition. After removing the lending rate ceiling, the bank would want each of the loans in its portfolio to be larger and be charged a higher rate (the $\bar{R}_L(k)$). For the given D then, the bank must take some firms out of its portfolio $\mathbf{B}$. These firms, leaving the bank to join the market for direct lending, would then increase the demand for finance in that market, pushing up the interest rate on direct lending. This, however, would induce depositors to leave the bank and join direct lending, cutting D and lowering

[38] 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.

the interest rate on direct lending. The story continues. With the decreased D, the bank must again adjust its portfolio of lending to make $\mathbf{B}$ even smaller, moving more firms into direct lending, pushing up again the interest rate on direct lending, inducing more consumers to leave the bank and join direct lending. And this goes on until the market settles at a new and lower equilibrium D, the $\widetilde{D}^*$ in Figure 15.

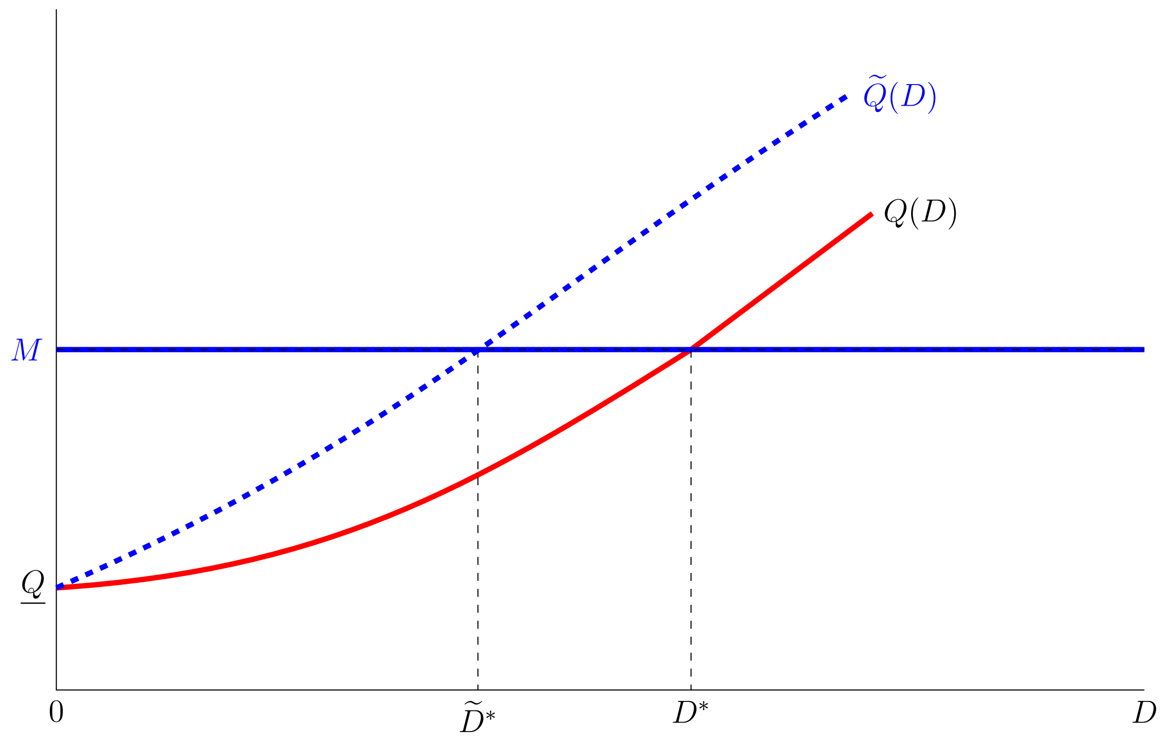


Figure 15: Equilibrium after removing the lending rate ceiling

A remark is order. Removing the lending rate ceiling is supposed to make the bank more able to compete in the market for finance. The outcome, however, seems to go in the opposite direction, weakening, instead of strengthening, the bank's standing in the financial system. What happens is that the deposit rate R_D, held fixed by the regulator, essentially forces the bank to choose larger profits on individual contracts at the expense of total amount of credit extended. Suppose the bank is free to choose the values of both R_L and R_D. Then, R_D would go up to at least partially balance the effects described in the above paragraph.

5.2 Removing all controls on lending rates

In July 2013, the central bank also scraped the floor on bank lending rates. The effects of this reform depends, of course, on whether the floor, $\underline{R}_L$, binds before being removed. By equation (42), if the floor is lower than the lending rate in the benchmark model (before the reform), removing the floor has no effects on the equilibrium outcomes of the model. If the floor is large enough, then removing it increases the equilibrium measure of firms receiving a bank loan, expanding the set $\mathbf{B}$ to include some of the larger firms which were not given a bank loan before the reform.

5.3 Removing deposit rate controls

Following the lending rate reforms, in October 2015 the central bank removed also its control on deposit rates. With this, all the restrictions on interest rates have been lifted, and the bank is free to choose the deposit rate R_D, the lending rates $\{R_L(k)\}$, as well as its loan portfolio $\mathbf{B}$, and the size of each loan, $\{Z(k)\}_{k\in\mathbf{B}}$, to maximize expected profits.

We define an equilibrium of the model as a measure of consumers who choose to lend through the bank $D^* \in [0, M]$ and an interest rate on direct lending r^* (not the market rate of lending for consumers in the benchmark model), which the agents in the economy take as given and produce outcomes consistent with them. We continue to focus on equilibria where direct lending and bank loans coexist.

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

$$\max_{D,R_D,\mathbf{B},\{R_L(k),Z(k)\}_{k\in\mathbf{B}}} \mu \int_{\mathbf{B}} \Big\{ \pi_1 (\theta_1 - \gamma_0) Z(k) + \pi_2 R_L(k) [Z(k) - k] \Big\} dG(k)$$
$$+ D - \mu \int_{\mathbf{B}} [Z(k) - k] \, dG(k) - R_D D \tag{45}$$

subject to (21), (23), (37), (39) and

$$D = \begin{cases} M, & \text{if } R_D > r^*, \\ D^*, & \text{if } R_D = r^*, \\ 0, & \text{if } R_D < r^*. \end{cases} \tag{46}$$

Notice that what equation (46) describes, namely D as a function of R_D, is not continuous and has a non-convex image. The solution to the above problem has:

(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 (40)).

(iii) $\mathbf{B} = \{k : \lambda(k) \geq \lambda^*\}$, where $\lambda(k), k \in [0, \bar{k}]$, is given in (43), and λ^* solves

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

Following from (iii), and as depicted in Figure 13, we have $\mathbf{B} = [\tilde{k}_1(D), \tilde{k}_2(D)]$. Thus, as in the case of fixed R_D but flexible $R_L(k)$, and by the same logic, here in equilibrium the bank would include in its loan portfolio medium-sized firms whose net worth is neither too large nor too small. The largest firms would raise finance from the bond market, and the smallest firms with private lending.

Now in order for r^* and D^* to constitute an equilibrium, the solution to the bank's problem must have $D = D^*$ and the market for direct lending clears:

$$\mu \int_0^{\tilde{k}_1(D^*)} L_{\mathrm{M}}(k)dG(k) + \mu \int_{\tilde{k}_2(D^*)}^{\bar{k}} L_{\mathrm{N}}(k)dG(k) = M - D^*. \tag{47}$$

Proposition 6. *Removing the control on R_D results in a higher equilibrium interest rate for 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, each individual firm in the private lending market is raising a smaller amount of finance ($X(k) - k$ smaller), and operating a smaller project.*

When the bank is free to set the interest rate on deposits, increased competition for funds between the bank and the firms in the direct lending market bids up the returns for consumers. The bank, with a new instrument for raising deposits, is also able to attract more deposits, expanding banking at the expense of direct lending.

To conclude, note that with all the interest rate controls on banking removed, one would think the bank is able to replicate, or do strictly better than, any contract the market for direct finance could offer. In particular, the bank, being the more efficient delegated monitor, should be able to crowd out monitored direct finance 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 (46)), is hard to rule out.

6 Empirical Support

Does the model make sense empirically? In this section, we take two major predictions of the model to the data, seeking both for empirical support for our analysis, and for explanations for the observed decline in banking and the rise of the bond market in China over the last 15 years, as Figure 1 shows.

Prediction 1. *Increasing the economy's total supply of external finance M shifts the equilibrium composition of aggregate finance away from bonds and towards bank loans, and tightening M squeezes bank lending but expands the market of bond finance.*

Prediction 2. *All else equal, removing the bank lending rate ceiling moves the market towards less bank loans and more private lending and bond finance.*

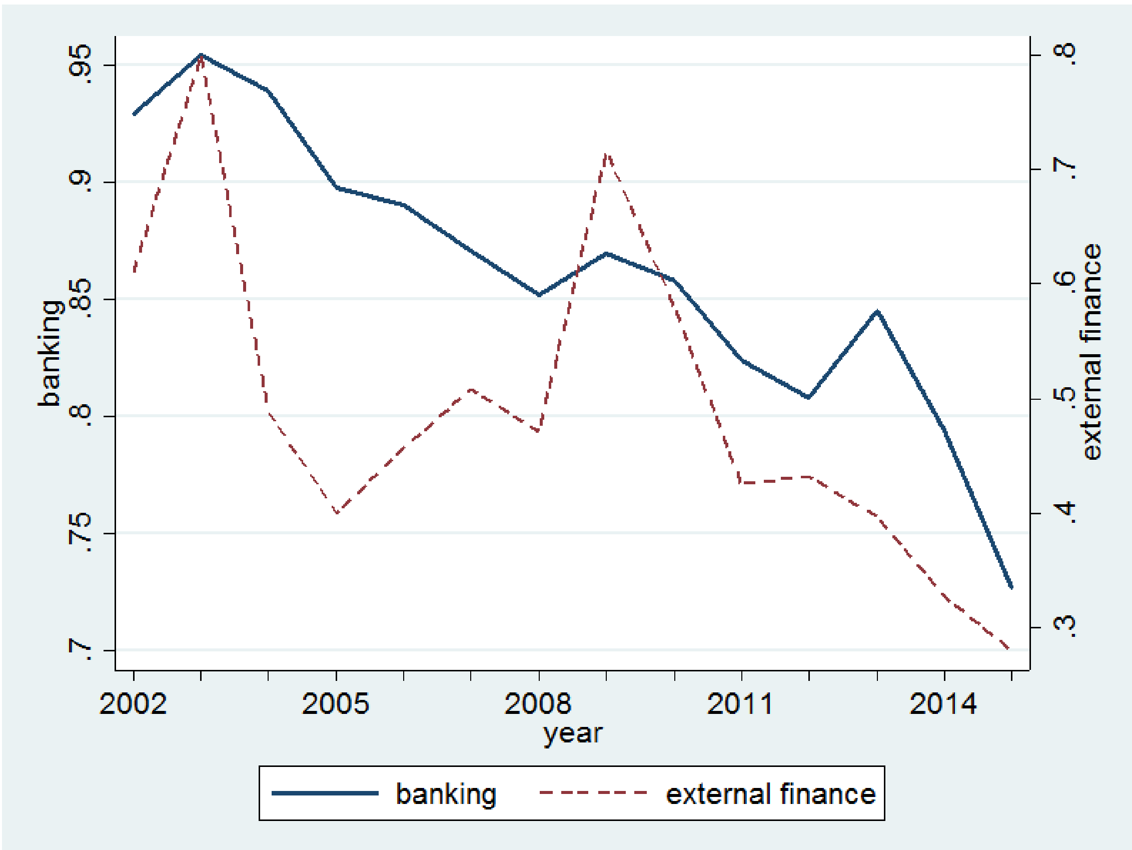


Figure 16: Banking and aggregate external finance

Source: CEIC.

Prediction 1 follows from Figure 12 and Prediction 2 from the discussions in Section 5.

In the model, holding the distribution function $G(k)$ of firms fixed, the ratio of the measure of investors over that of firms, M/μ, measures the tightness of the credit market. Clearly, if the tightness of credit stays constant, then the equilibrium size composition of the system remains constant (i.e., the equilibrium sizes of bank loans, bond finance and private lending remain constant relative to each other). Given this, in the regressions that link empirically the supply of external finance to the variability in the size composition of the financial system, we measure the supply of external finance not directly as the M in the model, but as the ratio of total external finance to total investment (internally plus externally financed), or

$$\frac{M}{\mu \int_0^{\bar{k}} kdG(k) + M}$$

which, obviously, is increasing in M/μ.

The data is from CEIC, covering 2002-2015, over which the bank lending rate ceiling was removed in 2004, but the deposit rate control stayed in place throughout the sample period. Part of the data is displayed in Figure 16, where "banking" measures the fraction of bank loans in aggregate financing – to represent the D in the model, and "external finance" measures aggregate financing as a fraction of total fixed investment – to represent the M/μ in the model.[39]

Observe that, in the data, the movements in banking and total external finance do seem serially correlated, as Prediction 1 claims. Observe also the steady drop in banking starting from 2004, the year the central bank removed its lending rate ceiling on bank loans – a policy shift which, according to the analysis in Section 5, should reduce banking. It, however, is theoretically ambiguous whether this results from decreases in M, or the removal of the lending rate ceiling.

To test the predictions, consider first the regression

$$Bank_t = \beta_0 + \beta_1 \times M_t + \beta_2 \times RD_t + \beta_3 \times D1_t + \beta_4 \times D2_t + \varepsilon_t,$$

where $Bank_t$ denotes the quantity of bank loans as a fraction of aggregate financing in period t; M_t, as discussed earlier, is aggregate financing as a fraction of total fixed investment in period t; RD_t is the nominal rate of return on one year saving deposits; $D1_t$ is a dummy variable which takes value 1 for the periods in the years 2005 and 2006, and 0 for all other periods; and $D2_t$ is a dummy variable which takes value 1 for each period in the years after and including 2007, and 0 otherwise. A period is a quarter.

The dummy variables are designed to catch the potential downward shifts in the demand for bank loans following the 2004 reform that Prediction 1 suggests. We hypothesize, however,

[39] According to the PBC, aggregate financing to the real economy (AFRE) "refers to the outstanding of financing provided by the financial system to the real economy during the period, where real economy means non-financial enterprises and households." Clearly, AFRE does not include private lending. Our interpretation of the data, therefore, assumes that total private lending and the calculated total finance tend to move in same directions.

that the policy takes time to enforce, and thus the downward shift in bank loans also takes time to unfold, in two stages following its implementation.

Table 4: Bank loans in Predictions 1 and 2

	(1)	(2)	(3)	(4)
External finance	0.297***	0.247***	0.385***	0.283***
	(0.0563)	(0.0504)	(0.0830)	(0.0768)
Deposit rate	-0.00498	0.0504**	-0.000429	0.0505**
	(0.0224)	(0.0235)	(0.0218)	(0.0218)
Year 2005 – 2006		-0.0377		-0.0343
		(0.0336)		(0.0325)
After 2007		-0.118***		-0.116***
		(0.0305)		(0.0296)
Constant	0.716***	0.707***	0.697***	0.726***
	(0.0559)	(0.0502)	(0.0700)	(0.0635)
Seasonal effect			Controlled	Controlled
Observations	46	46	46	46
R-squared	0.400	0.576	0.501	0.666

Standard errors in parentheses, with *** p<0.01, ** p<0.05, * p<0.1

The outcomes from the regression, displayed in Table 4, support the model's predictions. In particular, the negative signs of the estimated coefficients of $D1_t$ and $D2_t$ suggest a negative effect from removing the cap on R_L which, as the regression postulates, is released in two steps. Notice that the absolute value of the estimated β_3 is lower than that of the estimated β_4, indicating, obviously, that the size of the effect diminished in time.

To focus on the market for bond finance, we now use the quantity of bond finance as a fraction of aggregate financing as the dependent variable of the regression:

$$Bonds_t = \beta_0 + \beta_1 \times M_t + \beta_2 \times RD_t + \beta_3 \times D1_t + \beta_4 \times D2_t + \varepsilon_t,$$

where $Bonds_t$ denotes the fraction of bonds in aggregate financing in period t.[40] The outcome of this regression, which consistent with the predictions of the model, is displayed in Table 5.

Table 5: Bonds in Predictions 1 and 2

	(1)	(2)	(3)	(4)
External finance	-0.229***	-0.177***	-0.324***	-0.215***
	(0.0484)	(0.0396)	(0.0741)	(0.0643)
Deposit rate	-0.0119	-0.0683***	-0.0181	-0.0688***
	(0.0193)	(0.0185)	(0.0195)	(0.0182)
Year 2005 – 2006		0.0489*		0.0442
		(0.0264)		(0.0272)
After 2007		0.123***		0.120***
		(0.0240)		(0.0248)
Constant	0.221***	0.226***	0.264***	0.228***
	(0.0481)	(0.0395)	(0.0625)	(0.0532)
Seasonal effect			Controlled	Controlled
Observations	46	46	46	46
R-squared	0.342	0.611	0.411	0.653

Standard errors in parentheses, with *** p<0.01, ** p<0.05, * p<0.1

To close this section, note that the model also predicts that as the credit market tightens (M/μ decreases), the share of private lending in the economy's total external finance should increase, and removing the controls on the bank lending and deposit rates should reduce the

[40]This includes both government and corporate bonds, and we use the sum as a proxy for the corporate bonds in the model, given, from Figure 6, that both of them have been rising after 2002.

weight of private lending. While it is beyond the scope of this paper to offer a formal test for these predictions, they do not appear inconsistent with the so far very limited data/estimates that the literature is able to provide, namely the ratio of private lending to bank loans rose from 4.6% in 2003 to 6.4% in 2012, as cited in the introduction of the paper.

7 Conclusion

We have constructed a model for China's financial system in which a state owned bank competes with lending through a private market that is subject to no restrictions imposed by the state. The model accounts theoretically for the coexistence of bank loans, corporate bonds, and private lending as means of external finance in China's financial system. The model also offers an empirically tested explanation for the observed decline of the market for bank loans and rise of that for corporate bonds. The model also predicts that removing interest rate controls on the bank squeezes out private lending from China's financial system.

The model is "small" and can be extended in potentially many ways for better understanding China's financial system. For example, the environment can be enriched by way of introducing a government which issues a public debt, through or not through the banking sector. The model can then be used to study the effects of public debt in allocating financial resources. Properly calibrated, one should also be able to use the model for evaluating the effects of more efficient banking/monitoring on the distribution of finance among the different means of finance.

Appendix

7.1 Optimality of direct lending has $S(k) = \emptyset$ or $S(k) = \{\theta_1\}$

Fixed any $k \in [0, \bar{k}]$. Suppose $S(k) = \{\theta_2\}$. The problem of optimal contracting becomes

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

subject to

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

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

$$\theta_2 X - r_2(X-k) - \widetilde{C}(X-k,k) \geq \theta_2 X - r_1(X-k), \tag{48}$$

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

Let $\{r_1^*, r_2^*, X^*\}$ be a solution. From constraint (48) we have $r_2^* < r_1^*$. Consider now an alternative plan $\{S'(k), r_1', r_2', X^*\}$ with $S'(k) = \emptyset$ and $r_1' = r_2' = \pi_1 r_1^* + \pi_2 r_2^* \leq r_1^*$. This new plan is feasible and gives the entrepreneur an extra value of $\pi_2 \widetilde{C}(X^* - k, k)$. A contradiction.

Suppose $S(k) = \{\theta_1, \theta_2\}$. Then the problem of optimal contracting is

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

subject to

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

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

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

Let $\{r_1^*, r_2^*, X^*\}$ be a solution. Suppose $r_2^* > r_1^*$. Consider an alternative plan $\{S'(k), r_1', r_2', X^*\}$ with $S'(k) = \{\theta_1\}$ and $r_1' = r_1^* - \varepsilon, r_2' = r_2^* + \frac{\widetilde{C}(X^*-k,k)}{X^*-k}$, where ε is positive and sufficiently small. This new plan is feasible and gives the entrepreneur a higher value, which is a contradiction. Suppose $r_2^* \leq r_1^*$. Then consider an alternative plan $\{S'(k), r_1', r_2', X^*\}$ with $S'(k) = \emptyset$ and $r_1' = r_2' = \pi_1 r_1^* + \pi_2 r_2^* \leq r_1^* \leq \theta_1$, which is feasible and gives the entrepreneur a higher value. Again a contradiction.

7.2 Proof of Lemma 1

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

$$\max_{\{X\}} \left\{ (E(\theta) - r^*)X + r^* k \right\}$$

subject to

$$k \leq X \leq \frac{r^* k}{r^* - \theta_1}, \tag{49}$$

where equation (49) is from (2). Clearly, the optimal X has $X = r^* k/(r^* - \theta_1)$. That is, it is optimal to maximize the size of the lending. Substituting the optimal solution into the entrepreneur's objective delivers the desired results on the entrepreneur's values .

7.3 Proof of Proposition 2

Let $\Phi \equiv \{k \in [0,\bar{k}] \mid V_{\mathrm{M}}(k) > V_{\mathrm{N}}(k)\}$. This is set of entrepreneurs who prefer monitored direct lending to bond finance. To prove the proposition we need only show $\Phi = [0,\tilde{k})$ and for all $k \in \Phi$, equations (11) - (15) hold at the optimum.

Step 1 Fix any $k \in \Phi$ and suppose the optimal contract conditional on $S(k) = \{\theta_1\}$ is $\{r_1, r_2, X\}$.

Notice 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 has $X > k$ and so $\widetilde{C}(X-k, X) = C(X-k, X)$. Notice also that the participation constraint (8) binds, or $\pi_1 r_1 + \pi_2 r_2 = r^*$. For otherwise r_2 can be reduced to make the entrepreneur strictly better off.

The incentive constraint (7) 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 (5) gives

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

or

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

Now consider an alternative plan at k, $\{S'(k), r_1', r_2', X'\}$, with $S'(k) = \emptyset$ and $r_1' = r_2' = \pi_1 r_1 + \pi_2 r_2$, and $X' = X$. This plan is feasible (satisfying all the constraints at k), 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 to $k \in \Phi$.

Given the above, the entrepreneur's problem is 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

$$0 \leq r_1(X-k) \leq \theta_1 X - \gamma_0 X - \gamma X(X-k),$$

$$0 \leq r_2(X-k) \leq \theta_2 X, \tag{50}$$

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

Notice that the objective does not depend on r_1 and r_2 directly. Maximizing the objective subjective only to the constraint $X \geq k$ gives

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

where

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

with the entrepreneur's value being

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

Step 2 We show that there exists a unique $\tilde{k} \leq k'$ such that

$$V_{UC}(k) - V_{\mathrm{N}}(k) \begin{cases} > 0, & k \in [0, \tilde{k}) \\ = 0, & k = \tilde{k} \\ < 0, & k \in (\tilde{k}, \bar{k}] \end{cases}. \tag{51}$$

To see this, notice first that from Lemma 1 and Step 1, $V_{UC}(0) > V_{\mathrm{N}}(0) = 0$ and $V_{UC}(k') < V_{\mathrm{N}}(k')$. Notice second that for all $k \in [0, k')$,

$$\frac{dV_{UC}(k)}{dk} < E(\theta) < \pi_2(\theta_2 - \theta_1)\frac{r^*}{r^* - \theta_1} = \frac{dV_{\mathrm{N}}(k)}{dk}.$$

Step 3 To prove part (ii) of this proposition, we show for all $k \in [0, \tilde{k})$, the contract with

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

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

and

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

is optimal conditional on $S(k) = \{\theta_1\}$, and so $V_{\mathrm{M}}(k) = V_{UC}(k)$ for all $k \in [0, \tilde{k}]$. This, together with equation (51), gives $\Phi = [0, \tilde{k})$, or part (i) of the proposition

From Step 1, the above specified contract attains the "unconstrained" value $V_{UC}(k)$, so to show it is optimal we need only show that it is feasible. Notice first that given

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

constraint (6) is satisfied. 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 constraint (5) is satisfied. The incentive constraint (7) is satisfied because $V_{UC}(k) > V_{\mathrm{N}}(k)$ and the participation constraint (8) is also satisfied. The proposition is proved.

7.4 Proof of Corollary 3

From (16) and (17), for any $k \in [\tilde{k}, \bar{k}]$ we have

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

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

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

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.

7.5 Corollary 7, proof and intuition

Corollary 7. *The optimal direct lending contract 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 (8) binds for all $k \in [0, \tilde{k}]$. The incentive constraint (7) 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,

$$\begin{aligned}\frac{dr_1(k)}{dk} &= \frac{(\theta_1 - \gamma_0)(X_{\mathrm{M}}(k) - 1/2k)}{(X_{\mathrm{M}}(k) - k)^2} - \frac{1}{2}\gamma \\ &= \frac{2\pi_1\gamma(\theta_1 - \gamma_0)(E(\theta) - \pi_1\gamma_0 - r^*)}{(E(\theta) - \pi_1\gamma k - \pi_1\gamma_0 - r^*)^2} - \frac{1}{2}\gamma \\ &\geq \frac{2\pi_1\gamma(\theta_1 - \gamma_0)}{E(\theta) - \pi_1\gamma_0 - r^*} - \frac{1}{2}\gamma \\ &= \frac{\gamma}{E(\theta) - \pi_1\gamma_0 - r^*}\left(2\pi_1\theta_1 - 2\pi_1\gamma_0 + \pi_1\theta_1 + r^* - \pi_2\theta_2 - \pi_1\gamma_0\right) \\ &> 0,\end{aligned}$$

where the last inequality is from Assumption 1. Moreover,

$$\frac{dr_2(k)}{dk} = -\frac{\pi_1}{\pi_2} r_1'(k) < 0.$$

The intuition behind the above proof is as follows. As k increases, $r_1(k)$ increases, as a larger entrepreneur 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 (14), 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 $r_2(k)$.

□

7.6 Corollary 8 and proof

Corollary 8. *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 $X_{\mathrm{M}}(k) > X_{\mathrm{N}}(k)$. □

7.7 Proof for Proposition 4

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

Step 1 We show that the budget constraint (23) 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 $\pi_1 (\theta_1 - \gamma_0) + \pi_2 R_L - 1 > 0$. 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

$$\max_{\mathbf{B}; L(k), k \in \mathbf{B}} C_1 \int_{\mathbf{B}} k dG(k) + C_2 \tag{52}$$

subject to

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

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

$$\mu \int_{\mathbf{B}} L(k) dG(k) = D, \tag{53}$$

$$\pi_2 \theta_2 k + \pi_2 (\theta_2 - R_L) L(k) \geq V(k), \ \forall k \in \mathbf{B}, \tag{54}$$

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 (54) 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 (24), $Z(k) - k > L_0(k), k \in H$. 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

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

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

Step 4 Consider the case where $D \geq D_1$. Suppose $\mathbf{B} \subset [0, \bar{k}]$. From **Step 3**, the participation constraint (54) 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 (53) we have

$$D = \mu \int_{\mathbf{B}} \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\theta_2 - R_L)} dG(k) < \mu \int_0^{\bar{k}} \frac{V(k) - \pi_2\theta_2 k}{\pi_2(\theta_2 - R_L)} dG(k) = D_1.$$

A contradiction. So when $D \geq D_1$, we have $\mathbf{B} = [0, \bar{k}]$.

Now the total net worth of the entrepreneurs $\mu \int_{\mathbf{B}} kdG(k)$ is constant. From (52) 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 $D < D_1$. From (53) and (54) we have

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

Thus $\mathbf{B} \subset [0, \bar{k}]$, which implies resource constraint (54) 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

$$\max_{\mathbf{B} \subseteq [0, \bar{k}]} \int_{\mathbf{B}} kdG(k) \tag{55}$$

subject to

$$\int_{\mathbf{B}} L(k) dG(k) = \frac{D}{\mu}, \tag{56}$$

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

Let λ be the Lagrange multiplier of the constraint (56). 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 3,

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

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 $D \in [D_0, D_1)$, 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.

7.8 $Z_0(k)$ and $L_0(k)$ are increasing in k

From (17),

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

which is increasing in k at $k \in [0, \tilde{k})$ and constant in k for $k \in [\tilde{k}, \bar{k}]$. Moreover, $V'(k) \geq V'(0) = r^* + \frac{1}{2}[E(\theta) - r^* - \pi_1\gamma_0]$. These, with Assumption 1, then imply $V'(k) > \pi_2\theta_2 > \pi_2 R_L$ for $k \in [0, \bar{k}]$ 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

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

7.9 Lemma 9, proof, and Figure 17

Lemma 9. *Let* $\bar{R}_D \equiv \pi_1\theta_1 + 2\pi_2 R_L - \pi_2\theta_2 - \pi_1\gamma_0$. *Then the optimal direct and bank lending contracts have*

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

(ii) $R_D \geq \bar{R}_D$: $Z_0(k) < X(k)$ *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. From equation (27), $Z_0(k)$ satisfies

$$\pi_2 \left\{\theta_2 Z_0(k) - R_L \left[Z_0(k) - k\right]\right\} = V(k), \quad \forall k \in [0, \bar{k}]. \tag{57}$$

From Proposition 2 and Corollary 7,

$$\pi_2 \left[\theta_2 X(k) - r_2(k)(X(k) - k)\right] = V(k), \quad \forall k \in [0, \bar{k}]. \tag{58}$$

From (57) and (58),

$$Z_0(k) \geq X(k) \Leftrightarrow R_L \geq r_2(k). \tag{59}$$

Notice from Corollary 7 that $r_2(k)$ is decreasing in k and

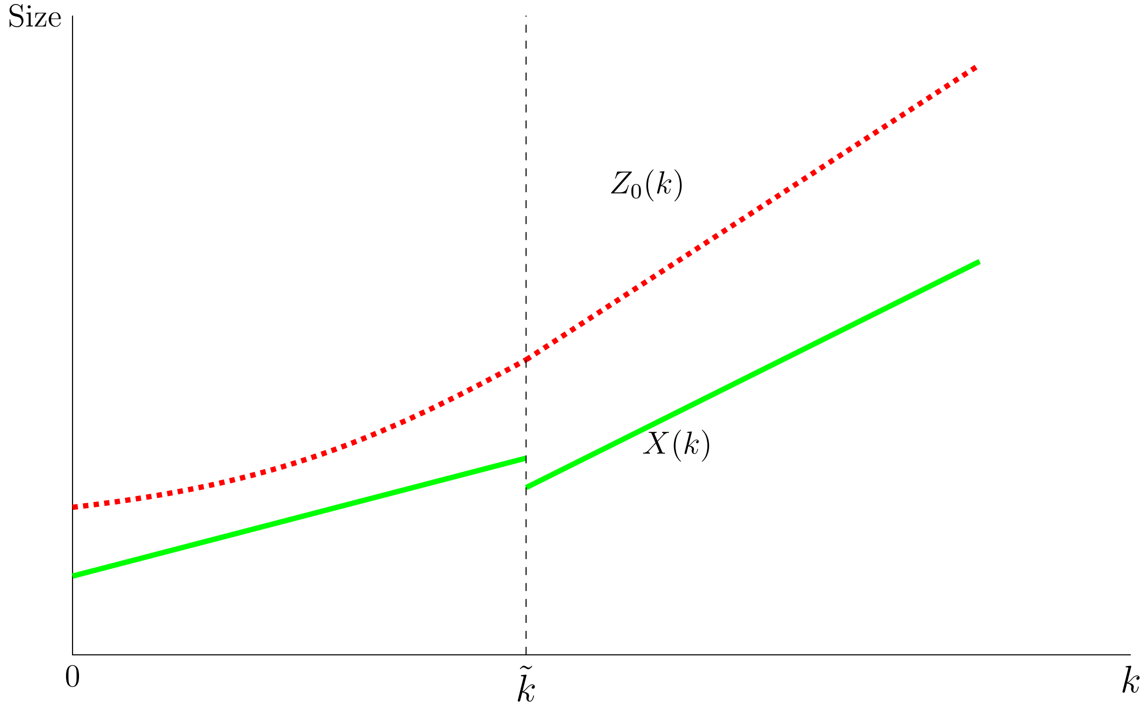
$$r_2(0) = \frac{R_D + \pi_2\theta_2 - \pi_1\theta_1 + \pi_1\gamma_0}{2\pi_2}.$$

Thus if $R_D < \bar{R}_D$, then $R_L > r_2(0) > r_2(k)$ and $Z_0(k) > X(k)$ for all $k \in [0, \bar{k}]$, and so (i) holds. If $R_D \geq \bar{R}_D$, then $R_L < r_2(0)$, and $r_2(\tilde{k}) = R_D < R_L$. So there exists some $k^* \in [0, \tilde{k}]$ so that

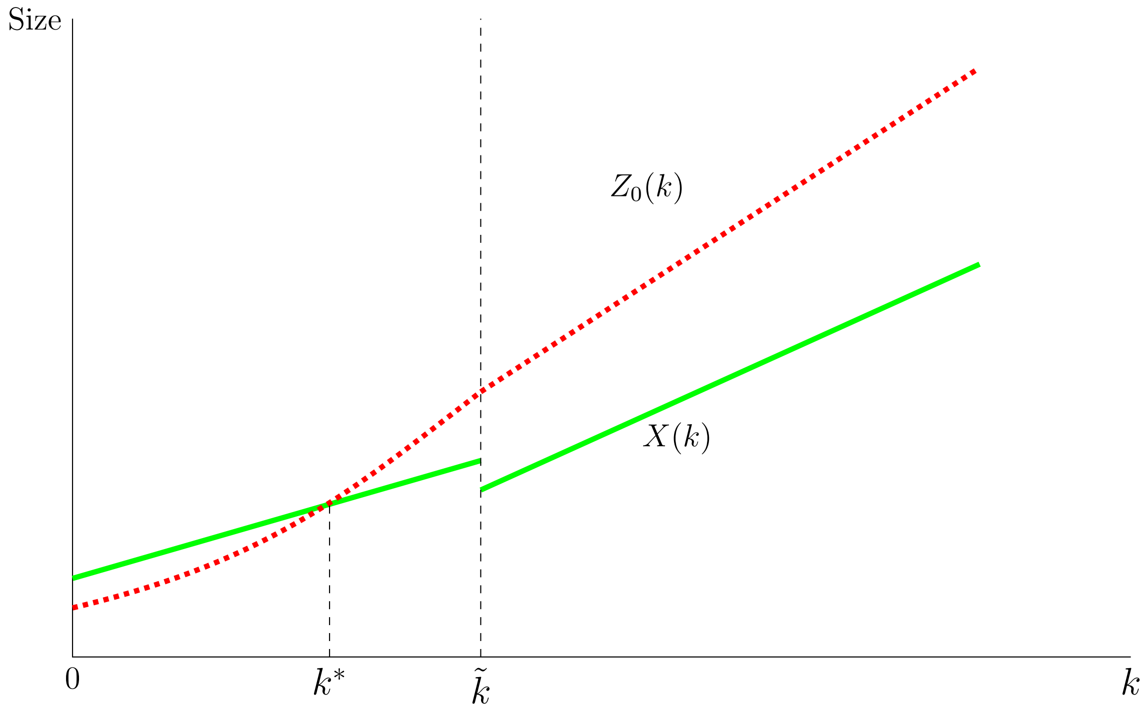
$$R_L \begin{cases} < r_2(k), & if \quad k \in [0, k^*) \\ = r_2(k), & if \quad k = k^* \\ > r_2(k), & if \quad k \in (k^*, \bar{k}] \end{cases}$$

which, given (59), proves part (ii) of the lemma. □

Figure 17 depicts what Lemma 9 states for the two cases, $R_D > \bar{R}_D$ and $R_D \leq \bar{R}_D$, respectively.



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



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

Figure 17: The optimal size of the project: direct and bank lending

7.10 Lemma 10 and proof

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

Lemma 10. *With the optimal contracts, $Q(D)$ is strictly increasing in D at all $D \in (0, M]$.*

Proof. Given $R_D < \bar{R}_D$, it follows from Lemma 9 that $Z(k) \geq Z_0(k) > X(k), \forall k \in [0, \bar{k}]$. And from Proposition 4, $\hat{k}_1(D)$ is weakly decreasing in D and $\hat{k}_2(D)$ weakly increasing in D. Given equation (35), the lemma is proved. $\square$

Suppose $R_D \geq \bar{R}_D$. Then $Q(D)$ is decreasing in D over the interval $[D_1', D_1]$, where $D_1' = \mu \int_{k^*}^{\bar{k}} L_B(k) dG(k)$ and k^* is defined in part (ii) of Lemma 9. This case is depictted in Figure 18 below.

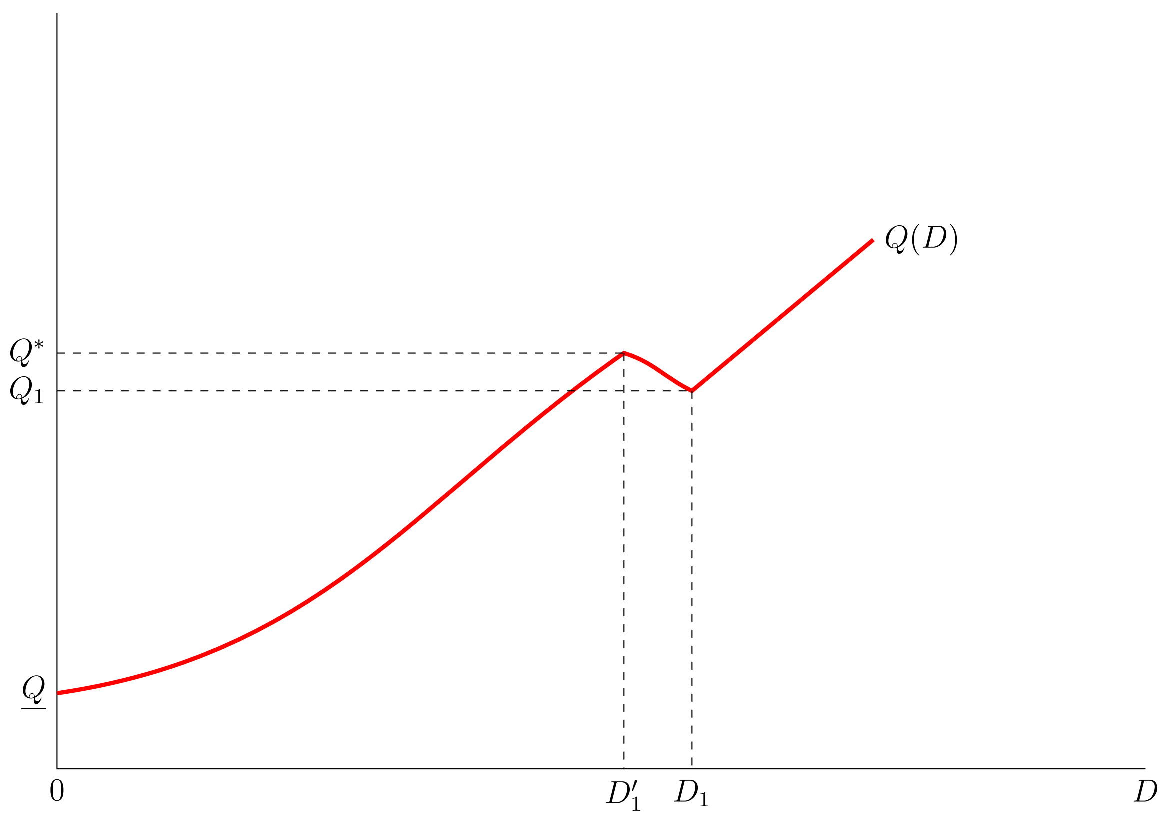


Figure 18: The demand function under $R_D \geq \bar{R}_D$

7.11 The monotonicity of $\lambda(k)$ in Figure 13

From equations (17) and (43) we have

$$\lambda(k) = \begin{cases} \frac{(E(\theta)-\pi_1\gamma_0)\bar{X}-(\bar{X}-k)R_D-[E(\theta)+\pi_1\gamma k-R_D-\pi_1\gamma_0]^2/(4\pi_1\gamma)-kR_D}{\bar{X}-k}, & \forall k < \tilde{k} \\ \frac{(E(\theta)-\pi_1\gamma_0)\bar{X}-(\bar{X}-k)R_D-\pi_2(\theta_2-\theta_1)R_D/(R_D-\theta_1)k}{\bar{X}-k}, & \forall k \geq \tilde{k} \end{cases}.$$

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

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

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

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

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

7.12 $\widetilde{Q}(D) > Q(D)$, $\forall D \in (0, D_0)$

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

$$\int_{\tilde{k}_1}^{\tilde{k}} [\bar{X} - k]dG(k) < \int_{\tilde{k}_2}^{\hat{k}} [Z_0(k) - k]dG(k). \tag{60}$$

Apply the First Mean Value Theorem for Integrals, there exist $k' \in (\tilde{k}_1, \tilde{k})$ and $k'' \in (\tilde{k}_2, \hat{k})$ 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 $\bar{X} - k' > Z_0(k'') - k''$, (60) and the above two equations give

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

And suppose $\bar{X}$ is large enough. Given that $X(k)$ is increasing in k at all $k \in [0, \bar{k}]$ (Lemma 1 and Proposition 2), apply again the First Mean Value Theorem for Integrals to obtain

$$\widetilde{Q}(D) - Q(D) = -\mu \int_{\tilde{k}_1}^{\tilde{k}} [X(k) - k] dG(k) + \mu \int_{\tilde{k}_2}^{\hat{k}} [X(k) - k] dG(k) > 0.$$

7.13 After removing all interest rate regulations

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

From Figure 13, it holds that $\lambda(k) > R_b(k, L_0(k))$ for all $k \in [0, \bar{k}]$, and so $\lambda(k) > \max\{R_b(k, L_0(k')), k' \in [0, \bar{k}]\} > 0$, supposing of course that the loans the bank makes offer positive rates of return.

Assume $\bar{k}$ is large enough that $\lambda(0) > \lambda(\bar{k})$ (note that $\lambda(k)$ is strictly decreasing after $\tilde{k}$ and goes to $-\infty$).

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

$$U(D) = \mu \int_{\tilde{k}_1(D)}^{\tilde{k}_2(D)} \lambda(k)(\bar{X} - k) dG(k). \tag{61}$$

This is the bank's total profits earned, conditional on D. Now note that in equilibrium, as in the main body of the paper, the bank's choice of D is restricted to be from the set $\{0, D^*, M\}$, where D^* is taken as the initial state of D in which the bank sits right before the reform occurs, $D = 0$ as what the bank attains if the bank lowers the deposit rate to be below the regulated R_D before the reform, and $D = M$ as that of the bank if it raises the deposit rate to be above the regulated R_D. We thus have, from (47),

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

Suppose the economy is initially in the 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 $\tilde{k}_2(M) < \bar{k}$.[41] This implies the total demand for credit in the direct lending market is larger than the supply of credit (see equation (47)). This will increase r^* and then 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 (34)).

7.14 Formulating the equilibrium in a system of equations

Following Definition 1, more specifically 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^*) satisfy:

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

(II) $\tilde{k}$ solves

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

[41] Otherwise, suppose $\tilde{k}_2(M) = \bar{k}$. Given $\lambda(0) > \lambda(\bar{k})$, we have $\tilde{k}_1(M) = 0$ and then

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

A contradiction.

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 $D^* \geq D_1$. 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

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

(b) $\mathbf{B} = [\hat{k}, \bar{k}]$ where $\hat{k} \in (0, \bar{k})$ if $D^* \in (D_0, D_1)$. 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:

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

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

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, C., K. Chen, D. Waggoner, and T. Zha (2015) "Trends and Cycles in China's Macroeconomy," *National Bureau of Economic Research Working Paper 21244*.

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," *National Bureau of Economic Research Working Paper 21890*.

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

Didier, T. and S. Schmukler (2013) "The Financing and Growth of Firms in China and India: Evidence From Capital Markets," *Journal of International Money and Finance*, Vol. 39, pp. 111–137.

Dollar, D. and S. Wei (2007) "Das (Wasted) Kapital: Firm Ownership and Investment Efficiency in China," *National Bureau of Economic Research Working Paper 13103*.

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," *Research paper, 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.

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.

——— (2013) "Quantifying the Impact of Financial Development on Economic Development," *Review of Economic Dynamics*, Vol. 16(1), pp. 194–215.

Hachem, K. C. and Z. M. Song (2016) "Liquidity Regulation and Unintended Financial Transformation in China," *National Bureau of Economic Research Working Paper 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.

Hoff, K. and J. E. Stiglitz (1998) "Moneylenders and Bankers: Price-increasing Subsidies in a Monopolistically Competitive Market," *Journal of Development Economics*, Vol. 55(2), pp. 485–518.

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

Kato, T. and C. Long (2006) "CEO Turnover, Firm Performance and Enterprise Reform in China: Evidence from New Micro Data," *Journal of Comparative Economics*, Vol. 34(4), pp. 796–817.

Kroeber, A. (2016) "China's Economy: What Everyone Needs to Know," *Oxford University Press*.

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.

Madestam, A. (2014) "Informal Finance: A Theory of Moneylenders," *Journal of Development Economics*, Vol. 107, pp. 157–174.

Song, Z, K. Storesletten, and F. Zilibotti (2011) "Growing Like China," *American Economic Review*, Vol. 101(1), pp. 196–233.

Stein, J. C. (2002) "Information Production and Capital Allocation: Decentralized versus Hierarchical Firms," *Journal of Finance*, Vol. 57(5), pp. 1891–1921.

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.

——— (1987) "Costly Monitoring, Loan Contracts, and Equilibrium Credit Rationing," *Quarterly Journal of Economics*, Vol. 102, pp. 135–145.