China’s Financial System in Equilibrium

Jie Luo and Cheng Wang*

June 25, 2018

Abstract

We present a macro view of China’s financial system, where a state-owned monopolistic banking sector coexists, endogenously, with markets for corporate bonds and private loans. The source and size distributions of external finance are determined jointly in the model’s equilibrium. Consistent with data, in equilibrium smaller firms obtain finance from the private lending market, larger firms through bank loans, and the largest by way of corporate bonds. The model suggests that removing the controls on bank lending rates or tightening the supply of credit reduces bank loans but increases bond finance. It also suggests that removing all interest rate controls would increase the rate of return on lending, expanding banking but squeezing direct lending. The model is calibrated to China’s financial market data to show that the observed rise of bonds over bank loans results partially from tightened credit, but mainly from reduced variability in the productivity of Chinese firms. The model also suggests that the private loans market has varied in size between 0 to 10 percent of the economy’s total credit.

**JEL:** G18, G21, O16

**Keywords:** China, financial system, monitoring, banking reforms

---

*First version: May 2016. Luo: National Institute for Fiscal Studies, 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, Xuewen Liu and the audience at Fudan University, Tsinghua PBC School of Finance, University of Goettingen, University of Hong Kong, SAIF, Shanghai Jiaotong University, 2017 Asian Meeting of the Econometric Society, 2017 China Meeting of the Econometric Society, 2017 North American Summer Meeting of the Econometric Society, and the 2017 China Financial Research Conference for discussions.*
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.

Figure 1: Composition of aggregate financing in China

Source: CEIC.

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.
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. 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.

Banks in China are largely state owned. 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 higher in China than in the U.S..

Banks are subject to state controls, although the last 15 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

---

1 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.

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

3 There is also a decline of banking in other emerging economies after 2010. See Figure 23 in the appendix. The same figure also shows that banking as a fraction to total credit to the private non-financial sector has been stable over the same time. Observe the scale difference, between China and other emerging economies, in both the duration and the scale of the decline. In addition, Chang, Fernández and Gulan (2017) argue that, for most emerging economies, the decline in banking could be explained by the drop of global interest rates. This, however, could not explain China. Interest rates in China have been quite stable over the 15 years, as to be shown shortly, in Figure 3.

4 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).
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. Figure 3 depicts the time series of the policy rates on one year loans and on one year saving deposits. Notice the greater variability in both the policy lending and deposit rates after 2004.

![Figure 2: 5-bank loans concentration in commercial banks in China and U.S.](image)

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.

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.

---

5 Bank regulations exist also on loan quantities. In fact, in many cases the PBC conducts its monetary 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.
They also show that the average lending rates in the private credit market are 2 ∼ 3 times more than the bank lending rates.\textsuperscript{7}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Monthly lending and deposit rates in China}
\end{figure}

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 World 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. 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

\textsuperscript{7}See Figure 6 in their paper.
on bank loans but more on families and friends for finance. Ayyagari et al. (2010) also find that in China bank financing is more prevalent with larger firms.

Table 1: Percent of firms using banks to finance investments

<table>
<thead>
<tr>
<th>Firm Size</th>
<th>China</th>
<th>East Asia &amp; Pacific</th>
<th>All Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small (5-19)</td>
<td>3.8</td>
<td>16.8</td>
<td>21.5</td>
</tr>
<tr>
<td>Medium (20-99)</td>
<td>20.4</td>
<td>23</td>
<td>27.1</td>
</tr>
<tr>
<td>Large (100+)</td>
<td>23.3</td>
<td>22.7</td>
<td>30.7</td>
</tr>
</tbody>
</table>

Note: Only manufacturing firms are included. Small, medium, and large firms are defined by the number of employees.

Firms in the World Bank’s surveys are on average much smaller than publicly listed firms (see Table 4 in the appendix). To obtain a more comprehensive view, we merge the publicly listed firms and those in the 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 the only source of external finance emerges, as shown in Figure 4.

Table 2: Percent of reasons why firms did not apply for any line of credit

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


Allen, Qian and Qian (2005) argue 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 P2P in China, fills a demand for credit from investors and going part way to solving the problem of getting financing to small firms.

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%).

Following the World Bank, firm size is measured as total employment.
One might suggest that bank loans are, for some reason, too expensive to smaller firms. This is not the case, as Table 2 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.

![Figure 4: Fraction of firms in China with only bank finance, 2011](image)

Source: World Bank’s Enterprise Surveys data for China 2012 and CSMAR.\textsuperscript{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 substantially over the last twenty years, from virtually nonexistent to the third biggest in the world, just behind the U.S. and Japan. From Figure 5, although corporate bonds still account for a smaller part of the whole market, they have grown fast in relative

\textsuperscript{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.
size. Another important feature of China’s bond market, as shown in Figure 6, 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.\(^{12}\)

![Figure 5: Size of local currency bonds in China](source)

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 private lending. 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

\(^{12}\)That firms who use bonds for external finance are larger than those who use bank loans is not just observed among Chinese firms.

\(^{13}\)See, for example, Allen and Qian (2014).
financial system works – what’s inside it that generates the features and characteristics one observes. This motivates our work.

![Graph showing median firm size](image)

**Figure 6:** 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 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. We show that in equilibrium firms with larger net worth obtain finance from the bank or by bond issuance, while those with smaller net worth borrow from individual investors in the private lending market.
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. 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. The model predicts 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 then 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 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

---

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.
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.

In the last part of the analysis, we take the model quantitatively, calibrating it to China’s financial market data, and using it to quantify the explanations for the observed recent decline in banking relative to bonds. The model suggests that the reduced variability in the firm’s productivity accounts for most of the rise of the bonds/bank-loans ratio. This seemingly surprising result is consistent with the logic of the model. Bonds, the instrument that has the least ability in monitoring the borrower and enforcing incentives, expanded in an economy that experienced decreasing variability in productivity and reduced demand for repayment incentives.

The calibrated model also allows us to back out from it the measure of private loans as a fraction of total credit. It shows that over the time period 2002-2015 the private loans market has varied in size between 0 to 10 percent of the economy’s total credit. The model also suggests a sharp reduction in private lending around 2009 and then a steady increase of private loans afterwards.

1.3 The literature

This 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 to the larger literature on banking and financial markets. In Holmström and Tirole (1997) for example, with 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). In De Fiore and Uhlig (2011), firms exposed to high risk of default abstain from production, firms with relatively low risk choose to raise external finance through bonds, only firms with intermediate degrees of risk choose to sign a contract with banks. Two elements of our model make us 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 unique “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. In contrast to these theories, it holds in our model 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). They also 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. They show a tight loan-to-deposit ratio induces small banks to use WMP for poaching deposits from big banks.

In the following, 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 calibrates the model to the data. Section 7 concludes the paper. Proofs of theoretical results are in the appendix.

2 Model

There are two time periods: \( t = 0, 1 \). In period 0 a financial market opens where lending and borrowing take place, and in period 1 production and consumption take place. There is a
single good in the model that can be used as capital or consumption.

There is a continuum of agents, among them \( M \) units are investors and \( \mu \) units firms. Firms maximize expected profits in period 1. Investors maximize expected consumption in period 1. Profits and consumption must be non-negative.

Each lender 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 firms, with \( \bar{k} > 0 \). Each firm is also endowed with an investment project with which any \( X(\geq 0) \) units of capital invested in period 0 returns \( \tilde{\theta}X \) units of output in period 1, where \( \tilde{\theta} \) is a random variable that takes value \( \theta_1 \) with probability \( \pi_1 \), and \( \theta_2 \) with probability \( \pi_2 \), with \( \theta_2 > \theta_1 > 0 \) and \( \pi_1 = 1 - \pi_2 \in (0, 1) \).

A bank in the model takes deposits from investors and offers loans to firms. This bank is “state owned” and subject to regulations. Let \( R_D \) denote the gross rate of return on deposits and \( R_L \) the gross interest rate charged on loans. The values of \( R_D \) and \( R_L \) are fixed by the state and are such that \( 0 < R_D < R_L \). Naturally, assume \( R_D \in (\theta_1, E(\theta)) \) and \( R_L \in (R_D, \theta_2) \).\(^{15}\)

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

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

\[
C(\Delta, X) = \gamma_0 X^\eta + \gamma \Delta X^\eta,
\]

where \( X \) is the size of the project, \( \Delta \) the measure of lenders who provide the external finance, and \( \gamma_0 \) and \( \gamma \) are positive constants. Let \( \gamma_0 < \theta_1 \) and \( \eta > 0 \). In the following Section 3 to 5, we let \( \eta = 1 \) for obtaining an analytical solution to the model. In the calibration of the model in Section 6, we let \( \eta \) be a free parameter for matching quantitatively the chosen targets.

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

\(^{15}\)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.
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. Last, observe that given equation (1), the bank as a lender is always more efficient than individual investors, as long monitoring is involved in the lending.

To close the model, assume monitoring, delegated or non-delegated, is deterministic. That is, any lending contract could only specify to monitor in a given state of the world with probability one or zero. Assume also that any lending relationship in this environment must meet a limited liability constraint: in no state of the world the firm be required to make a credit repayment that exceeds the output it produces.

3 Optimal Lending

Let \( r^* \) denote the market rate of (net expected) return on lending for individual investors. This is an endogenous variable whose value will be determined in the equilibrium of the model, with \( r^* \in [R_D, E(\theta)] \). 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 investors are lenders. Firms are free to participate in either side of the market. However, given \( r^* < E(\theta) \), it is never optimal for any firm to lend any fraction of his net worth to the market, directly or indirectly. In the following analysis, therefore, we take as given that all firms are a borrower.

3.1 Direct Lending

Consider first the market where individual investors lend directly to firms, not through the bank. Consider an individual firm in this market, with net worth \( k \). To obtain finance, it offers a contract to potential lenders. Given deterministic monitoring, the contract takes the general 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 $S(k) = \{\theta_1\}$. In the following, we consider the two cases separately and in turn, before deriving the optimal contract.

### 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 firm’s loan repayment must be constant across the states of output, or $r_1(k) = r_2(k) = r_N(k)$, and its value is given by

$$V_N(k) \equiv \max_{r_N; L \geq 0} \left\{ \pi_1 \theta_1(L + k) + \pi_2 \theta_2(L + k) - r_N L \right\}$$

subject to

$$r_N L \leq \theta_1(L + k), \quad (2)$$

$$r_N \geq r^*, \quad (3)$$

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

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

$$L_N(k) = \frac{\theta_1 k}{r^* - \theta_1}, \quad X_N(k) = \frac{r^* k}{r^* - \theta_1}. \quad (4)$$

That is, 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 firm’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

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

Notice that $L_N(k)$, $X_N(k)$ and $V_N(k)$ are all linear and increasing in $k$. That is, conditional on no-monitoring, a larger firm net worth supports more finance, a larger project, and higher firm value.\(^{17}\)

---

\(^{16}\)See the appendix (Section 8.2) for the proof.  
\(^{17}\)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 $L_N(k)$ 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.
3.1.2 Monitored Direct Lending

Alternatively, the firm could raise finance with a contract that involves monitoring: \( S(k) = \{ \theta_1 \} \). In this case the problem of optimal contracting is given by

\[
V_M(k) \equiv \max_{\{r_1, r_2, L \geq 0\}} \left\{ \pi_1 \left[ \theta_1(L + k) - r_1 L - \tilde{C}(L, k) \right] + \pi_2 \left[ \theta_2(L + k) - r_2 L \right] \right\}
\]

subject to

\[
0 \leq r_1 L \leq \theta_1(L + k) - \tilde{C}(L, k), \tag{5}
\]

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

\[
\theta_1(L + k) - r_1 L - \tilde{C}(L, k) \geq \theta_1(L + k) - r_2 L, \tag{7}
\]

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

where

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

In the above, (5) and (6) are non-negativity and limited liability, and (7) is incentive compatibility. Note that given \( S(k) = \{ \theta_1 \} \), the contract must only ensure that the firm has no incentives to report \( \theta_2 \) when output is \( \theta_1 \). Equation (8) is a participation constraint. Lastly, (9) says that the cost of monitoring is \( C(L, L + k) \) if lending takes place, zero if not.

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

\[
r_2 - r_1 \geq \tilde{C}(L, k)/\gamma \geq 0. \tag{10}
\]

or only a gap between \( r_1 \) and \( r_2 \). The size of this gap, however, is increasing in the cost of monitoring \( \tilde{C}(L, k) \) which, in turn, is increasing in \( k \) for any fixed \( L \). With a smaller \( k \) (smaller \( \tilde{C}(L, k) \)), a less tight incentive constraint (10) gives lenders 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 \( k \) is
sufficiently large to make $\tilde{C}(L,k)$ sufficiently large, (10) is likely to bind, or simply infeasible for the contract to implement (remember $r_1$ must be non-negative and $r_2$ must not exceed $\theta_2$). This reduces the firm’s value. To summarize, in monitored direct lending, monitoring goes better with a smaller rather than a larger $k$.

Lastly, in monitored direct lending, where each lender imposes on the firm a monitoring cost of $\gamma(L+k)$ to verify the report of $\theta_1$, the repetition in monitoring implies that the total cost of monitoring incurred increases more than linearly in the size of the project, and this strengthens the effects discussed above.

3.1.3 Optimal Direct Lending

The firm’s optimal finance is now determined, under

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

Part (i) ensures that the mean output of the project is sufficiently high so that once it is financed, on average the firm has enough to cover the reservation return of the lender plus the fixed cost in monitoring which is assumed to occur in the state of low output. Part (ii) assumes that the deposit rate is sufficiently high so that the non-negativity constraint $r_1 \geq 0$ in (5) does not bind.\(^{18}\)

**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:

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

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

$$r_1(k) = \frac{(\theta_1 - \gamma_0)X_M(k) - L_M(k)\gamma X_M(k)}{L_M(k)}, \tag{13}$$

$$r_2(k) = \frac{r^* - \pi_1r_1(k)}{\pi_2}. \tag{14}$$

\(^{18}\)Suppose (ii) is violated. Then the constraint $0 \leq r_1L$ binds for all $k < \tilde{k}'$, where $\tilde{k}' = (\pi_2\theta_2 - \pi_1\theta_1 + \pi_1\gamma_0 - r^*)/(\pi_1\gamma)$. It then follows that $r_1(k) = 0$ and $X(k) = k + (\theta_1 - \gamma_0)/\gamma$, for all $k \in [0, \tilde{k}']$. This changes the analysis slightly but would not change the qualitative conclusions derived. Note also that the assumption $R_D > R_{\text{min}}$ will not be maintained later in the paper when the model is calibrated to China’s financial market data.
and the value of the firm is

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

and \( \tilde{k} \) solves

\[ V_M(\tilde{k}) = V_N(\tilde{k}). \]

The determination of \( \tilde{k} \) is illustrated in Figure 7.\(^{19}\) With the optimal contract then,

\[
X(k) = \begin{cases} 
X_M(k), & \forall k < \tilde{k} \\
X_N(k), & \forall k \geq \tilde{k}
\end{cases}
\] (16)

and

\[
V(k) = \begin{cases} 
V_M(k), & \forall k < \tilde{k} \\
V_N(k), & \forall k \geq \tilde{k}
\end{cases}
\] (17)

That a larger \( k \) makes finance with monitoring less efficient relative to that with no monitoring was anticipated from discussions in the prior sections.\(^{20}\) Proposition 2 also suggests that conditional on monitoring, as conditional on no monitoring (Lemma 1), at the optimum a larger \( k \) supports a larger \( X \) and larger firm value. Remember, conditional on no monitoring, a larger \( k \), used as collateral, increases the firm’s ability to repay its debt in the state of low output. This same effect exists also in monitored private lending.\(^ {21}\)

Notice, interestingly, that the optimal size of the loan \((L(k))\) is increasing in \( k \) in the case of no monitoring, but decreasing in \( k \) in the case of monitoring. Conditional on no monitoring, a larger \( k \) supports larger repayments in the state of low output and hence a larger loan. In monitored direct lending, however, a larger \( k \) also increases the cost of monitoring and at an increasing rate, resulting in a smaller loan.\(^ {22}\)

\(^{19}\)More specifically \( \tilde{k} \) must solve \((E(\theta) + \pi_1 \gamma \tilde{k} - \pi_1 \gamma_0 - r^*)^2/(4\pi_1 \gamma) + \tilde{k}r^* = \pi_2(\theta_2 - \theta_1)/(r^* - \theta_1)\), which has a unique solution for \( \tilde{k} \in (0, \bar{k}) \).

\(^{20}\)In addition, note that conditional on monitoring, a larger \( k \) increases the cost of monitoring per unit of external finance, which is given by

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

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

\(^{21}\)But not between monitoring and no monitoring, as shown in Corollary 8 and Figure 25 in the appendix.

\(^{22}\)Because of the convex monitoring cost, the marginal net return on lending, which is \( E(\theta) - r^* - \pi_1(\gamma_0 + \gamma k + 2\gamma L) \), is decreasing in both \( k \) and \( L \).
In general, monitoring allows the contract to support more external finance and hence a larger project. In the appendix (Corollary 8), we show that the optimal size of the project conditional on monitoring, $X_M(k)$, strictly exceeds that conditional on no monitoring, $X_N(k)$, for all $k$. Observe also from Figure 25, again in the appendix, the jump in the optimal size of the project as a function of $k$, $X(k)$, at $\tilde{k}$ which divides monitoring and no-monitoring.

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 the appendix (Corollary 7), we show that with monitored direct lending, that is for all $k \in [0, \tilde{k})$, $r_1(k) < r^* < r_2(k)$ and $r'_1(k) > 0$, $r'_2(k) < 0$. That is, if the optimal contract prescribes monitoring, then there is an interest rate spread between the two output states, and the spread is smaller if the firm is larger in net worth. In addition, as is straightforward to show, for any fixed $k \in [0, \tilde{k}]$, $r_1(k)$ is larger if $r^*$ is larger. This is so because a larger $r^*$ reduces the optimal size of the investment and increases the efficiency in monitoring which, in turn, permits higher lender returns in the low output state.

Figure 7: Lender’s value functions in direct lending
3.2 Intermediated/Bank Finance

Let $D(\geq 0)$ denote the bank’s total deposits from investors. This is also the total supply of bank loans, an endogenous variable of the model whose value would depend especially on $r^*$, the market interest rate for all lenders. Note that we need only consider 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 $\theta_1$ given $\theta_1 < R_L$, it must submit all of its output to the bank. Given $R_L$, as part of the lending contract the bank then chooses the size of the loan $L(k)$, or equivalently the size of the firm’s project $Z(k)(\equiv L(k) + k)$, and a policy for monitoring the firm’s output.

Let $B$, a subset of $[0, \bar{k}]$, denote the set of all firms whom the bank is willing to offer a loan to. For each $k \in B$, the loan must ensure that the firm gets a value no less than $V(k)$ – the value the direct lending market could guarantee and thus the bank must take as the firm’s reservation value.

Consider the bank’s monitoring policy. Fix $k \in B$. With the optimal contract, monitoring occurs if and only if the lower output $\theta_1$ is reported. To see this, first it is straightforward to show that monitoring a report of $\theta_2$ 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

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

(18)

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 $V(k)$.\textsuperscript{23} In other words, if the bank never monitors the firm’s report, it would not be able to induce the firm to participate – it could not offer a loan that is sufficiently large to make the firm better off with a bank loan than with direct lending.

\textsuperscript{23}Specifically,

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

where the first inequality is from (18), the second holds because $R_L > R_D$. 

19
Given the above, the bank’s problem becomes

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

subject to

$$B \subseteq [0, \bar{k}], \quad (20)$$

$$L(k) \geq 0, \quad \forall k \in B, \quad (21)$$

$$\mu \int_B L(k) dG(k) \leq D, \quad (22)$$

$$V_b(k, L(k)) \equiv \pi_2 \{\theta_2 (k + L(k)) - R_L L(k)\} \geq V(k), \quad \forall k \in B, \quad (23)$$

where equation (22) is a resource constraint: total loans made cannot exceed the total supply of bank credit; and (23) is a participation constraint: the firms in $B$ are better off obtaining finance from the bank than from individual lenders directly.

Now rewrite (23) as

$$L(k) \geq L_0(k), \quad \forall k \in B, \quad (24)$$

where

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

and let

$$Z_0(k) \equiv k + L_0(k) = \frac{V(k) - \pi_2 R_L k}{\pi_2 (\theta_2 - R_L)}, \quad \forall k \in [0, \bar{k}]. \quad (26)$$

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

From (25), 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 $\theta_2$ after repaying the bank, reducing $L_0(k)$. Second, a larger $k$ increases the firm’s outside value $V(k)$, requiring a larger loan for participation.
Overall, however, it is shown that $L_0(k)$ and $Z_0(k)$ are both increasing in $k$.\footnote{See Section 8.9 in the appendix for proof.} In addition, notice $V_b(k, L_0(k)) = V(k)$. That is, at $L_0(k)$, the firm is indifferent between raising finance from the bank and borrowing directly from individual lenders.

We now move on, letting

\begin{align}
D_1 &\equiv \mu \int_0^\bar{k} L_0(k)dG(k), \\
D_0 &\equiv \mu \int_{\hat{k}}^{\bar{k}} L_0(k)dG(k).
\end{align}

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 $k \in [\hat{k}, \bar{k}]$ – those who would issue bonds for finance if a bank loan were not available.

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

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

The bank’s optimal policy is now given.

**Proposition 3.** The following holds under Assumption 2. (i) Suppose $0 \leq D < D_0$. Then the bank’s optimal plan has

$$L_B(k) = L_0(k), \ \forall k \in B,$$

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

$$\mu \int_B L_0(k)dG(k) = D.$$ \hfill (29)

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

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

where $\hat{k}$ solves

$$\mu \int_{\hat{k}}^{\bar{k}} L_0(k)dG(k) = D.$$ \hfill (28)

(iii) Suppose $D \geq D_1$. Then the optimal plan for the bank is to set $B = [0, \bar{k}]$, and with

$$\{L_B(k), k \in B\}$$

be any function that satisfies (21) and (22).
To understand these results, consider the bank’s rate of return on lending to a firm $k$ with a loan of size $L$, with $L \geq L_0(k)$:

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

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

(30)

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

In (i) and (ii) where $D < D_1$, the bank could not offer a credit to all firms, any capital above $L_0(k)$ could then be reallocated to a firm not yet receiving bank credit, and this gives extra returns to the bank. In these cases, what the bank seeks, essentially, is to maximize the number of loans made, by making each loan as small as possible. Under $D < D_1$, equation (30) also indicates the bank should in general prefer larger to smaller firms. More specifically, given (25) and Corollary 6 in the appendix,

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

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

---

25 For this, see the appendix (Step 3 in Section 8.8) for related calculations.

26 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 $\theta_2$ 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.
In case (i) with $0 \leq D < D_0$, 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 $B = [\hat{k}_1, \hat{k}_2]$, where $0 \leq \hat{k}_1 < \hat{k}_2 \leq \bar{k}$ (Figure 8).\(^{27}\)

In case (ii) with $D_0 < D < D_1$, 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 giving each of them a loan with their reservation size $L_0(k)$.

\(^{27}\)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).
Last, in the case of $D \geq D_1$, 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 to (a) meet the minimum demand for credit from each firm, and then (b) lend the rest of the funds to an arbitrary set of firms, on top of their $L_0(k)$.

Obviously, $\hat{k}_1(D)$ is decreasing in $D$ and $\hat{k}_2(D)$ is increasing in $D$, as Figure 8 illustrates. To conclude this section then, we claim that as $D$ increases, the use of bank loans relative to total finance increases monotonically, while the use of bond finance and monitored private lending decrease monotonically as a fraction of total external finance.

### 3.3 Direct vs. Bank Lending

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

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

A larger $R_D$ increases the value of the individual investor but reduces that of the firm, $V(k)$. This, given the fixed loan rate $R_L$, lowers the pressure on the bank in inducing the firm to participate, reducing the size of the loan offered. A higher $R_D$ also reduces the size of direct finance (i.e., $X(k) - k$). However, since interest rates are free to adjust in lending

---

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

29 See Lemma 9 in the appendix.

30 Lending with the risk free bond could be viewed as an outcome under infinite monitoring costs.
contract, the reduction in the size of direct finance would be less than that in the bank loan.\textsuperscript{31} Overall, 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 investors \( r^* \), a quantity of deposits \( D^* \), a set \( B \subseteq [0, \bar{k}] \) of firms whom the bank offers a loan to and the corresponding loan contracts \( \{(Z(k), R_L) : k \in B\} \), 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 \( B \) and the loan contracts \( \{(Z(k), R_L) : k \in B\} \) solve the bank’s optimization problem, as described in Section 3.
   (b) Entrepreneurs with net worth \( k \in B \) choose optimally to accept the loan the bank offers, those with \( k \notin B \) obtain finance from the direct lending.

3. Suppose \( r^* > R_D \). Then only the market for direct lending opens, with \( D^* = 0 \) and \( B = \emptyset \).

4. The demand for loans equals the supply of loans in the direct lending market:

\[ \mu \int_{[0,\bar{k}] \setminus B} [X(k) - k] \, dG(k) = M - D^*. \tag{31} \]

The above defined equilibrium of the model is formulated more explicitly in a system of

\textsuperscript{31}To see this more precisely, remember, for any fixed \( k \), in order to induce the firm to participate, \( Z_0(k) \) must satisfy

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

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 firm’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

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

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 \( X(k) \).
equations in the appendix (Section 8.15). We now characterize the outcomes of this equilibrium. To save space, we assume in the 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 $r^* \in [R_D, \theta(E))$, let $Q(D, r^*)$ denote the economy’s total demand for external finance:

$$Q(D, r^*) = \mu \int_{0}^{k_1(D, r^*)} L_M(k, r^*)dG(k) + \mu \int_{k_1(D, r^*)}^{k_2(D, r^*)} L_B(k, r^*)dG(k) + \mu \int_{k_2(D, r^*)}^{\bar{k}} L_N(k, r^*)dG(k),$$

which is the sum of the demand for monitored direct finance, for bank loans, and for bond finance. Note that the second part of the sum is equal to $D$, as the bank’s resource constraint binds.

Figure 9 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

$$Q(0, r^*) = \mu \int_{0}^{\hat{k}(r^*)} L_M(k, r^*)dG(k) + \mu \int_{\hat{k}(r^*)}^{\bar{k}} L_N(k, r^*)dG(k),$$

where $L_N(k, r^*)$ and $L_M(k, r^*)$, 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 $\bar{Q}$, where $\bar{Q}$ is the value of $Q(0, r^*)$ at $r^* = R_D$, with which the demand for external finance achieves its maximum conditional on $D = 0$.

What happens in the direct lending market in the case of $D = 0$ is depicted in Figure 10, where a value of $M$ below $\bar{Q}$ induces an equilibrium interest rate $r^*$ to clear the market.

\(^{32}\)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.
Observe that for $M$ sufficiently small, $M \leq \frac{M}{\bar{k}}$ specifically, the equilibrium interest rate $r^*$ would be so high that $L_M(k, r^*) = 0$ for all $k \in (0, \bar{k})$, while $L_N(k, r^*)$ remains positive for all $k \in [\bar{k}, \tilde{k}]$ (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.\footnote{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\footnote{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 25 (a), and so the slope of $Q$ in $D$ is positive at all $D$. Obviously, if $R_D \geq \bar{R}_D$ and Figure 25 (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 8.11.}

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 $D$, as depicted in Figure 9, 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.\footnote{With these, four cases emerge from Figure 9, in how the economy’s total supply of external finance, $M$, is divided, in equilibrium, among the three different instruments for finance.

**Case 1:** $M \leq Q$. All lending takes place directly between individual firms and investors, the equilibrium of the model being depicted in Figure 10.

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

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

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

and

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

where $L_N(k, r^*)$ is positive for all $r^* < E(\theta)$, whereas $L_M(k, r^*)$ is zero for all $r^* > \bar{v}^* \equiv E(\theta) - \pi_1 \gamma_0$. Notice that $\bar{v}^*$ is decreasing in $\pi_1 \gamma_0$. That is, a larger expected cost of monitoring makes monitoring more vulnerable in the market for monitoring.
**Case 4:** \( M \geq Q_1 \). In equilibrium \( D^* \geq D_1 \) and, from Proposition 3, 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 8, 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 credit induces a crowding out of direct finance by bank loans: as \( M \) increases, \( D^* \) is larger while \( M - D^* \) is smaller.

![Figure 9: Equilibrium when \( 0 < M < Q \)](image)

So an increase in \( M \) reduces the size of direct lending in both absolute and relative measures. Why is this? Imagine the economy were in an initial equilibrium. Imagine \( M \) is increased by a small positive amount \( \Delta \). Any fraction of this \( \Delta \) could not have flowed into the market for direct lending, for then the interest rate on direct lending would fall and investors would flow into bank deposits, which now offer a higher interest rate. In other words, the new funds must become an addition to the bank’s deposits, which now total \( D' \equiv D^* + \Delta \). With \( D' \), however, the bank would re-optimize, to expand its \( B \) to \( B' \), with \( B \subset B' \). This, in turn, would take firms away from direct lending, reducing the demand for direct lending, lowering the interest rate, and driving investors away from direct lending and into bank deposits, until the interest rate on direct lending is restored at \( R_D \). 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 settle at a new equilibrium level, which is strictly greater than that of the initial equilibrium. To summarize, the increase in $M$ results in an increase in banking (by more than the size $\Delta$), but a decrease in direct lending (a smaller $M - D^*$).

![Figure 10: Equilibrium with $D^* = 0$](image)

Observe also that as bank loans crowd out direct lending following the increase in $M$, the composition of direct lending also changes, towards smaller shares of bond finance but larger shares of monitored private lending, from Figure 8.

### 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 outcome. Figure 11 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 $Q$ are all functions of $R_D$, we write them explicitly as $Q_0(R_D)$, $Q_1(R_D)$ and $Q(R_D)$, respectively. These are all decreasing functions and are located relative to each other as the figure depicts.
Figure 11: 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 11 shows, again, that for fixed \( R_D \), increasing the supply of external finance \( M \) shifts the equilibrium composition of lending away from direct finance and towards bank loans; and tightening the supply of external finance squeezes bank lending but expands the market for direct finance. In particular, a sufficiently high \( M \) crowds out completely bond finance and monitored private lending to result in an equilibrium where bank loans is the only means of external finance; and a sufficiently small \( M \) gives rise to an equilibrium where bonds are the only source of external finance. The intuition, discussed earlier, is that a larger \( M \) puts downward pressure on the interest rate on direct lending, giving the bank, who is constrained to offer the fixed deposit rate, better ability in competing for deposits from the investors which, in turn, gives rise to a larger \( D \) and more bank 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

We now use the model to evaluate, analytically, the effects of the reforms that the central bank of China has enforced, 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 $Q(R_D) < M < Q_0(R_D)$ 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 the lending rate ceiling on commercial bank loans. This allows banks to set the lending rate on any individual loan anywhere above the floor rate, which continued to exist after the reform. Let $R_L$ denote the positive floor lending rate. With this reform, the bank’s optimization becomes

$$\max_{B, \{Z(k), R_L(k)\}_{k \in B}} \mu \int_B \left\{ \pi_1 (\theta_1 - \gamma_0) Z(k) + \pi_2 R_L(k) [Z(k) - k] \right\} dG(k)$$

$$+ D - \mu \int_B [Z(k) - k] dG(k) - R_D D$$

subject to (20), (22) and

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

$$R_L(k) \geq R_L \quad \forall k \in B,$$

$$\pi_2 \{\theta_2 Z(k) - R_L(k) [Z(k) - k]\} \geq V(k) \quad \forall k \in B.$$  (36)

As in the benchmark environment, the participation constraint (36) dictates a relationship between the lending rate charged, $R_L(k)$, and the size of the loan, $Z(k) - k$: a larger loan allows for a higher lending rate ($R_L(k)$) that the bank can charge on it. This, given (34), implies

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

where $\bar{R}_L(k)$ is the maximum lending rate the bank is able to charge on firm $k$, subject to (34) and (36). Naturally, $\bar{R}_L(k)$ is decreasing in $k$. With a larger $k$, the firm’s reservation value $V(k)$ is higher and the demand for external finance, $\bar{X} - k$, is smaller, both implying a lower maximum lending rate – the size of the firm imposes a constraint on what the bank can charge on the loan.

Parallel to Assumption 2 in the benchmark environment, we make

**Assumption 3.** $\pi_2 \bar{R}_L(k) + \pi_1 (\theta_1 - \gamma_0) > 1, \forall k.$
That is, for any $k$, the bank is better off lending to the firm at the maximum loan rate $\bar{R}_L(k)$, which implies an average rate of return on lending of $\pi_2 \bar{R}_L(k) + \pi_1 (\theta_1 - \gamma_0)$, than putting the funds on storage.

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

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

where since $V(k) - (E(\theta) - \pi_1 \gamma_0)k > 0$ (which holds for all $k \in [0, \bar{k}]$ from (17)), $R_b(k)$ is larger when $L(k)$ is larger. This is so because a larger loan dilutes the net cost of lending to the firm – the second term of the right hand side of the above equation – which in turn results in a higher average rate of return to the bank.\(^{38}\) Observe the contrast between this and what happened in the benchmark model, where a larger loan implies a lower rate of return on lending.

Given the above then, for any $k \in B$, it is optimal to set $L(k) = \bar{X} - k$, or $Z(k) = \bar{X}$, while the optimal lending rate is set at $R_L(k) = \bar{R}_L(k)$, defined in (37). Remember, with a fixed $R_L$, the bank wants the loans to be of the minimum size. There, by keeping the loans small, the bank lends to more firms, maximizing the use of firm net worth as collateral for enforcing credit repayments. Here, with a flexible $R_L$, the bank is able to make loans larger to minimize its cost of lending per unit of the loan.

Moreover, constraint (35) does not bind,\(^{39}\) and so the bank’s problem is reduced to choosing $B$ to maximize its total profits subject to constraint (22), and the solution has

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

\(^{38}\)The bank’s rate of return on lending is given by

$$R_6(k) = \frac{\pi_1 (\theta_1 - \gamma_0)(L(k) + k) + \pi_2 R_L(k)L(k)}{L(k)} - R_D.$$

Use the binding participation constraint (36) to solve for $R_L(k)$ and plug it into the above expression to get

$$R_6(k) = \frac{\pi_1 (\theta_1 - \gamma_0)(L(k) + k) + \pi_2 \theta_2 (L(k) + k) - V(k)}{L(k)} - R_D,$$

which then gives (38).

\(^{39}\) It is straightforward to show that for any $k \in [0, \bar{k}]$,

$$\bar{R}_L(k) = \theta_2 - \frac{V(k) - \pi_2 \theta_2 k}{\pi_2 (X - k)} > \theta_2 - \frac{V(k) - \pi_2 \theta_2 k}{\pi_2 (Z_0(k) - k)} = R_L \geq R_L.$$  (39)
where

$$\lambda(k) = \frac{(E(\theta) - \pi_1 \gamma_0) \bar{X} - V(k)}{\bar{X} - k} - R_D \quad (40)$$

is the bank’s expected rate of return on the loan to firm $k$, and $\lambda^*$ is determined by

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

That is, to maximize profits, the bank includes in its portfolio firms with the largest $\lambda(k)$s subject to the total funds available, as depicted in Figure 12.

![Figure 12: The bank’s optimal loan portfolio: $B = [\tilde{k}_1, \tilde{k}_2]$](image)

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

A larger $k$ has two effects on $\lambda(k)$. First, it implies a larger $V(k)$ which reduces the returns on lending to firm $k$. Second, it implies a smaller bank loan $(\bar{X} - k)$, which results in a higher average return on lending, increasing $\lambda(k)$. In the appendix (Section 8.12) 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 12, where $B = [\tilde{k}_1, \tilde{k}_2]$. Moreover, given $0 < \tilde{k}_1 < \tilde{k} < \tilde{k}_2 < \tilde{k}$, it follows from Proposition 2 that firms with $k \in [0, \tilde{k}_1)$ seek monitored private finance, and those with $k \in (\tilde{k}_2, \tilde{k}]$ obtain credit by way
of issuing bonds. So removing the lending rate ceiling does not change the model’s prediction that small firms use private loans, medium sized firms are financed with bank loans, and large firms issue bonds.

As is obvious from Figure 13, a larger $D$, by giving a lower $\lambda^*$, results in a lower $\tilde{k}_1$ but a larger $\tilde{k}_2$, implying both less bond finance and less monitored private lending.

![Figure 13: The division of total finance as a function of $D$](image)

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

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

The integral is

As is for $Q(D)$ in (32) in the benchmark case, it is easy to verify that $\tilde{Q}(D)$ is increasing in $\lambda^*$. 

As is for $Q(D)$ in (32) in the benchmark case, it is easy to verify that $\tilde{Q}(D)$ is increasing in $\lambda^*$. 

As is for $Q(D)$ in (32) in the benchmark case, it is easy to verify that $\tilde{Q}(D)$ is increasing in $\lambda^*$. 

As is for $Q(D)$ in (32) in the benchmark case, it is easy to verify that $\tilde{Q}(D)$ is increasing in $\lambda^*$.
The equilibrium bank deposits, denoted $\tilde{D}^*$, then solves
\[ \tilde{Q}(\tilde{D}^*) = M, \]
as depicted in Figure 14.

Obviously, a larger $M$ results in a larger $\tilde{D}^*$ and, from Figure 12, a lower $\lambda^*$ which, in turn, implies a lower $\tilde{k}_1$ and a higher $\tilde{k}_2$. In other words, after removing the lending rate ceiling, any time there is an expansion in the supply of credit in the economy, bank loans would crowd out both monitored private lending and bond finance, as in the benchmark model.

It holds for all $D \in (0, D_0)$ that $\tilde{Q}(D) > Q(D)$.\footnote{See Section 8.13 in the appendix. Remember at these $D$s all three markets are active in the benchmark model.} What happens is that, for any given $D$, removing the lending rate ceiling allows the bank to lend more at a higher interest rate to each individual firm. This reduces the measure of firms obtaining a bank loan, increasing the measure of firms in direct lending and their demand for external finance.

Observe from Figure 14 that $\tilde{D}^* < D^*$. That is, removing the lending rate ceiling results in decreased equilibrium quantity of bank deposits or loans. In addition, given $\tilde{k}_1(\tilde{D}^*) < \tilde{k}_1(D^*) = \tilde{k}$ and $\tilde{k}_2(\tilde{D}^*) < \tilde{k}_2(D^*) < \tilde{k}_2(D^*)$, the equilibrium share of monitored private finance in total lending would decline, but that of bond finance would increase.

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

So removing the lending rate ceiling does not alter the direction in which a change in $M$ affects banking. Consider the story behind (i) of the proposition. After removing the rate ceiling, the bank would want its loans to be larger and charge a higher rate (the $\tilde{R}_L(k)$). With the given $D$, it must then take out from its initial portfolio $B$ a set of larger firms and replace them with a group of smaller firms. This immediately expands the market for bonds – the larger firms, upon leaving the bank, would get finance by way of issuing bonds – and at the same time reduces private lending. The story continues. The adjustment in the bank’s portfolio would result in a net increase in the demand for direct finance, pushing up the interest rate on direct lending. This, in turn, would induce individual investors to substitute bank deposits for direct lending, cutting $D$ and lowering the interest rate on direct lending.

\footnote{We drop $r^*$, which is assumed to be fixed at $R_D$ in this part of the analysis, as an argument of the functions $\tilde{Q}$ and $Q$.}
With the decreased $D$, the bank must again adjust its loan portfolio to make $B$ even smaller, moving more (large) firms into direct lending, pushing up again the interest rate on direct lending, inducing more investors to leave bank deposits and join direct lending. And this goes on, until the market settles at a new and lower equilibrium $D$, the $\bar{D}^*$ in Figure 14, together with an expanded bond market but a smaller market for private lending.

![Figure 14: Equilibrium after removing the lending rate ceiling](image)

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

### 5.2 All lending rate controls removed

In July 2013, the central bank also scraped the policy floor on bank lending rates. The effects of this depends, of course, on whether the floor, $R_L$, binds before being removed. By equation
(39), if the floor is lower than the lending rate in the benchmark model (before the reform), removing it has no effects on the equilibrium outcome 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 $B$ to include some of the larger firms which were not given a bank loan before the reform.

5.3 Removing deposit rate controls

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 $B$, and the size of each loan, $\{Z(k)\}_{k \in B}$, to maximize profits.

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

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

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

subject to (20), (22), (34), (36) and

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

And the solution to the above problem has:

(i) $R_D = r^*$.

(ii) For all $k \in B$, $Z(k) = X$, and $R_L(k) = \bar{R}_L(k)$ (given in (37)).

(iii) $B = \{k : \lambda(k) \geq \lambda^*\}$, where $\lambda(k), k \in [0, \bar{k}]$, is given in (40), and $\lambda^*$ solves

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

---

Notice that what equation (43) describes, namely $D$ as a function of $R_D$, is not continuous and has a non-convex image.
And it follows from (iii) that $B = \left[ \hat{k}_1(D), \hat{k}_2(D) \right]$, as in the case of fixed $R_D$ but flexible $R_L(k)$ (Figure 12). That is, in equilibrium the bank includes in its loan portfolio medium-sized firms with net worth levels that are neither too large nor too small. The largest firms would raise finance from the bond market, and the smallest with monitored private lending.

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 must clear:

$$
\mu \int_0^{\hat{k}_1(D^*)} L_M(k)dG(k) + \mu \int_{\hat{k}_2(D^*)}^k L_N(k)dG(k) = M - D^*. 
$$

(44)

**Proposition 5.** Suppose direct lending and bank loans coexist before the reform. 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, individual firms in the private lending market each raise a smaller amount of finance ($X(k) - k$ smaller) and operate a smaller project.

Lifting the control on $R_D$ increases the competition for funds between the bank and the firms in the direct lending market, and this bids up the returns for investors. The bank, with a new instrument for collecting deposits, is able to expand banking at the expense of direct lending.

Lastly, note that with all the interest rate controls on banking removed, one would think that the bank, being the more efficient monitor, should be able to crowd out monitored direct lending completely. From the above discussion, however, monitored private lending is active in equilibrium if $\hat{k}_1(D^*) > 0$ which, given the non-convexity of the bank’s choices in $D$ (see (43)), is hard to rule out.

6 **Quantitative Analysis**

Does the model make sense quantitatively? In this section, we first take a look at the data regarding the relationship between the supply of credit and the composition of the financial system. We then calibrate the model to China’s financial system and ask how large the private loans market in China is, and what accounts for the rise of bonds over bank loans over the recent decade.
6.1 Credit tightness and the financial system

In the model, the equilibrium composition of the financial system depends on the supply of credit, \( M \), and the measure of firms in the economy, \( \mu \), only through their ratio \( M/\mu \) which, holding the distribution \( G(k) \) of firms fixed, measures the tightness of the credit market. Now, as is straightforward to see from the above analysis, if the tightness of credit stays constant, then the equilibrium 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). And the model predicts that a tighter credit market gives rise to smaller banking but a larger bond market (Figure 11).

There is, however, no direct data on \( M/\mu \) – the tightness of credit that was just defined. We consider instead \( \tilde{M} \) – the ratio of total external finance to total investment (internally plus externally financed):
\[
\tilde{M} \equiv \frac{M}{\mu \int_{0}^{k} kdG(k) + M}.
\]

This is essentially the economy’s average leverage ratio, which is strictly increasing in \( M/\mu \). A proxy for \( \tilde{M} \) is then constructed to be the ratio of “aggregate financing to the real economy” (AFRE) to total fixed investment. According to the PBC, 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.”

The data is displayed in Figure 15, where “banking”, which is the fraction of bank loans in aggregate financing and measures the relative size of banking in the data, is plotted against “credit tightness”, which measures the \( \tilde{M} \) in the data. The data is from CEIC, covering the time period 2002-2015, over which the bank lending rate ceiling was removed in 2004, but the deposit rate control was kept throughout the period. Observe that the movements in banking and credit tightness do look serially correlated, consistent with the prediction of the model. Observe also the steady drop in banking starting from 2004, the year the central

\[\text{43} \text{ Clearly, AFRE does not include private lending. Our interpretation of the data, which is not fully satisfactory, therefore assumes that total private lending and the calculated total finance tend to move in same directions.} \]

\[\text{44} \text{ According to the PBC, aggregate financing to the real economy (AFRE) “refers to the outstanding of financing provided by the financial system to the real economy during the period, where real economy means non-financial enterprises and households.” Clearly, AFRE does not include private lending. Our interpretation of the data, therefore, assumes that total private lending and the calculated total finance tend to move in same directions.} \]

\[\text{45} \text{ Independent evidence from Chen, Ren and Zha (2017) supports our view. They show that the slower growth of M2 after 2009 resulted in a simultaneous fall of bank deposits and loans between 2009 and 2015. In their analysis, tightening monetary policy would push investors into less regulated instruments of finance or shadow banking, and away from bank loans.} \]
bank removed its lending rate ceiling on bank loans – a policy shift which, according to our analysis, should reduce banking.

Figure 15: Banking and the tightness of credit

Source: CEIC.
Note: “banking” is the fraction of bank loans (including shadow banking) in aggregate financing; and “credit tightness” measures aggregate financing as a fraction of total fixed investment, or the $\tilde{M}$.

We are now prepared to take the model more seriously, calibrating it to China’s financial market data. In addition to asking whether the model offers a quantitatively sensible image of the observed parts of the financial system, we also wish to obtain an estimate of the size of the private lending market in China, which is missing from official data but has inspired speculations and concerns from both researchers and the public. Once calibrated, the model also allows us to explore other possibilities for explaining the observed decline in banking relative to bonds over the last fifteen years.
6.2 Calibration

For calibrating the model, the CEIC offers direct information on \( \{D_t, B_t, FA_t, RD_t, RL_t, Y_t\} \), where \( D_t \) and \( B_t \) respectively are newly created bank loans and newly issued bonds to the private non-financial sector; \( FA_t \) is total fixed capital formation; \( RD_t \) and \( RL_t \) are, respectively, the policy deposit and lending rates; and \( Y_t \) is the GDP. All measures are in nominal terms.

We take \( FA_t \) as the data counterpart of the sum of the total internal and external capital in the model, or \( K_t + M_t \), where \( M_t \equiv P_t + D_t + B_t \), where \( P_t \) denotes the amount of private lending, and

\[
K_t \equiv \mu_t \int_{k_0}^{\tilde{k}} kdG(k). \tag{45}
\]

Figure 16: The distribution of firms

Source: China 2008 Economic Census Data.
Note: “data” denotes the kernel density constructed from the data, and “model” denotes the fitted truncated normal distribution with parameters given by (46).
6.2.1 The distribution of firms

In the calibration, we assume that the composition of firms in net worth — the distribution function \( G(k) \) — is constant in time and is truncated log-normal: \( k \sim \ln N(\mu_k, \sigma_k^2) \), with \( k \in [k_0, \bar{k}] \), where the parameters are computed using the firm fixed assets data from the 2008 China Economic Census. Specifically,

\[
k_0 = 1, \quad \bar{k} = \exp(13.96), \quad \mu_k = 6.17, \quad \sigma_k^2 = 5.00.
\] (46)

And the data and model distributions are shown in Figure 16.46

6.2.2 The measure of firms

Given \( P_t \geq 0 \), a theoretical upper bound for the measure of firms can be derived:

\[
\mu_t \leq \bar{\mu}_t \equiv \frac{FA_t - D_t - B_t}{\int_{k_0}^{\bar{k}} kdG(k)}.
\] (47)

Given this, and under the assumption that the composition of firms in \( k \) — function \( G(\cdot) \) — is constant in time, we calibrate the economy’s measure of firms in a period \( t \) as

\[
\mu_t = \min \left\{ \mu_0 \times \frac{\bar{E}_t}{E_0}, \bar{\mu}_t \right\},
\] (48)

where \( \mu_0 \) is a parameter whose value is to be calibrated, \( \bar{E}_t \) is the total equity of manufacturing firms in China,47 which corresponds to the \( K_t \) in the model, and \( \bar{\mu}_t \) is computed from (47), where the distribution \( G(\cdot) \) is given in (46).

6.2.3 The production function

To calibrate the production function, we set the value of \( \pi_{1,t} \) to be the share of non-performing bank loans in year \( t \), whose time series is shown in Figure 17. Notice the great decline in the fraction of non-performing loans over the sample period. We then use the equation \( E(\theta_t)FA_t = Y_t \) to obtain a calibration of the expected value of \( \theta_t \):

\[
E(\theta_t) = \frac{Y_t}{FA_t},
\] (49)

which can be computed using the CEIC data.

46We excluded the largest 0.1% of firms as outliers. The number of firms in the dataset, with a non-missing value in firm fixed assets, is about 2.5 million.
47The data is from CEIC.
To calibrate the firm’s productivity measure $\theta_{1,t}$, we assume a constant growth rate in that value, letting

$$\theta_{1,t} = \theta_{1,1} G_{\theta_1}^{t-1},$$

(50)

where $\theta_{1,1}$ and $G_{\theta_1}$ are constants to be calibrated. Then, given the known $\theta_{1,t}$, for each $t$ the value of the firm’s $\theta_{2,t}$ is given by

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

(51)

where remember $E(\theta_t)$ is given in equation (49).

The time series of the computed $E(\theta_t)$ and the calibrated $\theta_{1t}$ and $\theta_{2t}$ are plotted in Figure 18. Observe the secular and significant decline in the value of $E(\theta_t)$ which, in the calibrated model, measures the average productivity of Chinese firms. Observe also the steady rise of the calibrated $\theta_{1,t}$ and the decline in $\theta_{2,t}$ over the sample period. Combining Figures 17 and 18, the calibration suggests also a decline of the variability in productivity among Chinese
firms over the sample period – the low output is higher and occurs with a lower probability.\textsuperscript{48}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure18.png}
\caption{Calibrated productivity parameters}
\end{figure}

6.2.4 Firm/investment size distribution

Remember $\bar{X}$ is the maximum physical size of the project - the size beyond which any additional investment is not productive. Remember also that after removing the lending rate ceiling, all projects financed with a bank loan would attain the maximum size $\bar{X}$. In the calibration, we deviate from this obviously unrealistic feature of the model. We assume that loan contracting is subject to an additional problem of moral hazard: the firm could take a fraction of the total investment in its project and leave the contract. With this, bank loans must then satisfy the following incentive constraint: $\alpha(k + L_t(k)) = \alpha Z_t(k) \leq V_t(k)$, which says that the value that the contract gives to the firm, which equals $V_t(k)$, must be larger than or equal to $\alpha$ times the size of the project. This constraint then requires that (34) in the bank’s optimization be rewritten, for each $t$, as

\begin{equation}
  k \leq Z_t(k) \leq \min \left\{ V_t(k)/\alpha, \bar{X}_t \right\}, \forall k \in B,
\end{equation}

\textsuperscript{48}That there has been a decline in the average productivity of Chinese firms is observed not just by us. The result of our estimation is consistent with, for example, Wu and Liang (2017), in which the aggregate TFP growth rate over 2001–2007 is estimated to be 1.32%, while that over 2007–2012 is -1.42%.
where $\bar{X}_t$ is to be calibrated. To minimize dimensionality, in the calibration we let

$$\bar{X}_t = \kappa \times \max\{X_t(k), Z_{0,t}(k), \forall k \in B\},$$

(53)

where $X_t(k)$ and $Z_{0,t}(k)$, respectively, are the optimal size of the project when external finance is from direct lending and bank loans, under the fixed lending rate $RL_t$. We let $\kappa \geq 1$ so that the analytical outcomes obtained under the fixed lending rate continues to hold with flexible lending rates.\textsuperscript{49}

6.2.5 Outcome of the calibration

To complete the calibration, we choose optimally the values of $\{\gamma_0, \gamma, \theta_{1,1}, G_{\theta_1}, \mu_0, \alpha, \kappa, \eta\}$ - these are the model’s parameters whose values were not yet determined in the prior discussion - to match the observed bank loans and bonds over 2002-2015. The results are given in Table 3. Figure 19 then shows how the calibrated model performs, in producing the ratio of bank loans over bonds relative to the data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed cost of monitoring parameter</td>
<td>$\gamma_0$</td>
<td>0.057</td>
</tr>
<tr>
<td>Variable cost of monitoring parameter</td>
<td>$\gamma$</td>
<td>0.026</td>
</tr>
<tr>
<td>Monitoring efficiency parameter</td>
<td>$\eta$</td>
<td>0.633</td>
</tr>
<tr>
<td>Initial value of $\theta_1$</td>
<td>$\theta_{1,1}$</td>
<td>0.092</td>
</tr>
<tr>
<td>Growth rate of $\theta_1$</td>
<td>$G_{\theta_1}$</td>
<td>1.022</td>
</tr>
<tr>
<td>Initial measure of firms</td>
<td>$\mu_0$</td>
<td>0.486</td>
</tr>
<tr>
<td>Moral hazard parameter</td>
<td>$\alpha$</td>
<td>0.412</td>
</tr>
<tr>
<td>Maximum investment parameter</td>
<td>$\kappa$</td>
<td>1.081</td>
</tr>
</tbody>
</table>

Table 3: The calibrated parameters

6.2.6 Bonds versus bank loans: the role of productivity variability

Figure 21, which reports the model’s counterfactual under the constant 2002 values of $\theta_1$, $\theta_2$ and $\pi_1$, shows how the ratio of bonds to bank loans would have moved over the sample period – it would not have gone up by that much relative to the data. Specifically, the counterfactual suggests that without the movements in firm productivity (i.e., the reduced

\textsuperscript{49}This obviously is not the only way to calibrate the $\bar{X}_t$. Generally, $\bar{X}_t$ can be calibrated as a free parameter for each $t$, or as a function of $t$ that entails a small number of parameters. Equation (53) offers one technically convenient way for the task.
mean and variability in the realization of $\theta$), the model would only be able to explain about 20% of the observed increase in the bonds/(bonds+bank-loans) ratio.

![Figure 19: Bank-loans/(bonds+bank-loans): model and data](image)

More counterfactuals are computed to breakdown the effect of the reduced variability in productivity on the bonds/bank-loans ratio (Section 8.16 in the appendix). The conclusion is that the rise of $\theta_1$ and the decline in $\pi_1$ contributed positively to the rise of the ratio, while the decline in $\theta_2$ did negatively. These, of course, are consistent with the spirit of the model. Bonds, the instrument that has the least ability in monitoring the borrower and enforcing incentives, expanded in an economy that experienced decreasing variability in productivity and hence reduced incentive problems.

What did the reform of 2004, which lifted the bank lending rate ceiling, do to China’s financial system? Suppose all other things happened the same way they did but the reform never occurred. Then the outcome of the model is that the Chinese economy would not have seen the rise of the markets for bonds and private lending after 2004. That is, bank loans would have been the only lending instrument observed throughout the period.

To summarize, the upshot is that the reform of 2004 and the movements in firm productivity - reduced variability in productivity especially - were responsible for producing the
observed rise of bonds over bank loans in China.

Figure 20: Bonds/(bonds+bank-loans) in the calibrated model

Figure 21: Bonds/(bonds+bank-loans): $\theta_1, \theta_2, \pi_1$ constant at 2002 level
6.2.7 The size of China’s private lending market

How large is China’s private lending market? The calibrated model allows us to back out, for each $t$, the measure of private loans as a fraction of total credit or

$$\frac{P_t}{M_t} = \frac{FA_t - K_t - D_t - B_t}{FA_t - K_t}.$$ 

Remember the private lending market consists of non-delegated monitors, such as relatives, money lenders, and other less delegated monitors including peer-to-peer platforms, and that this market was estimated to be quite large according to some studies. Our model confirms this quantitatively. Figure 22, which plots the time series of $P_t/M_t$ in the calibrated model, shows that over the sample period the size of the private loans market has varied in the rage of 0 to 10 percent of the economy’s total credit.

Observe the sharp reduction in private loans around 2009 and then the steady increase afterwards. Proposition 4, which states that greater credit tightness results in larger markets for bonds and private loans, offer an explanation for this. Remember from Figure 15 the sharp increase and then the steady decline in the tightness of credit ($\tilde{M}$) right after 2009.

![Figure 22: Private loans in total credit in the calibrated model](image-url)
7 Concluding Remarks

In this paper, we have studied a model of China’s financial system where bank loans, corporate bonds, and monitored private lending compete and coexist as means of external finance. The model suggests that the 2004 reform on the bank’s lending rate should have contributed to the observed (relative) decline in banking and the rise of the market for corporate bonds in China. Part of the observed decline in banking, as the model suggests, could also have resulted from a fall in the supply of credit in the Chinese economy. We have calibrated the model to China’s financial system data. This gives us a vehicle for quantitatively evaluating the effects of the movements in some of the economy’s major policy and technology variables. The calibrated model confirms the model’s theoretical prediction on the effects of the 2004 banking reform. A surprising result is that reduced variability in firm productivity contributed substantially to the observed decline in bank loans relative to bonds over the past fifteen years. The calibrated model also reveals that the private lending market has been large after 2008.

To end the paper, we note that the model is “small” and can be extended in potentially many ways for better understanding China’s financial system. For example, by introducing a government that issues public debt, through or not through the banking sector, the model could be used to study the role of public debt in allocating financial resources.
8 Appendix

8.1 Additional data: Figure 23 and Table 4

![Figure 23: Share of banking in markets around the world](image)

Figure 23: Share of banking in markets around the world

Source: BIS (Bank for International Settlements).

Note: This figure shows the end-of-quarter outstanding bank loans as a fraction of total credit to the private non-financial sector around the world.
Table 4: Number of firms in China, by firm size and sources of finance

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

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

(b) Within listed manufacture firms in China, 2011

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

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.

8.2 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

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

subject to

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

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

$$\theta_2 X - r_2 (X - k) - \tilde{C}(X - k, k) \geq \theta_2 X - r_1 (X - k), \quad (54)$$

$$\pi_1 r_1 + \pi_2 r_2 \geq r^*.$$
Let \( \{r^*_1, r^*_2, X^*\} \) be a solution. From constraint (54) 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 firm an extra value of \( \pi_2 \tilde{C}(X^* - k, k) \). A contradiction.

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

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

subject to

\[
0 \leq r_1(X - k) \leq \theta_1 X - \tilde{C}(X - k, k),
\]

\[
0 \leq r_2(X - k) \leq \theta_2 X - \tilde{C}(X - k, k),
\]

\[
\pi_1 r_1 + \pi_2 r_2 \geq r^*.
\]

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 \( r'_1 = r^*_1 - \epsilon, r'_2 = r^*_2 + \frac{\tilde{C}(X^* - k, k)}{X^* - k} \), where \( \epsilon \) is positive and sufficiently small. This new plan is feasible and gives the firm 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 firm a higher value. Again a contradiction.

### 8.3 Proof of Lemma 1

Fixed \( k \in [0, \bar{k}] \). Notice that the participation constraint is binding: \( r_N = r^* \), otherwise \( r_N \) can be reduced to make the firm strictly better off. With this, the firm’s optimization can be rewritten as:

\[
\max_{\{X\}} \{(E(\theta) - r^*)X + r^* k\}
\]

subject to

\[
k \leq X \leq \frac{r^* k}{r^* - \theta_1},
\]

where equation (55) 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 firm’s objective delivers the desired results on the firm’s values.
8.4 Proof of Proposition 2

Let \( \Phi \equiv \{ k \in [0, \bar{k}] \mid V_M(k) > V_N(k) \} \). This is set of firms who prefer monitored direct lending to bond finance. To prove the proposition we need only show \( \Phi = [0, \bar{k}) \) 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 \( V_M(k) = E(\theta)k \leq V_N(k) \), a contradiction to \( k \in \Phi \). Thus the optimal contract has \( X > k \) and so \( \tilde{C}(X - k, k) = C(X - k, X) \). 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 firm 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

\[
V_N(k) \geq E(\theta)X - (\pi_1 r_1 + \pi_2 r_2)(X - k)
\]

\[
\geq E(\theta)X - (\pi_1 r_1 + \pi_2 r_2)(X - k) - \pi_1 C(X - k, X)
\]

\[
= V_M(k),
\]

contradicting to \( k \in \Phi \).

Given the above, the firm’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

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

\[
0 \leq r_2(X - k) \leq \theta_2 X,
\]

\[
\pi_1 r_1 + \pi_2 r_2 = r^*.
\]

54
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

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

where

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

with the firm’s value being

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

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

$$V_{UC}(k) - V_N(k) = \begin{cases} 
> 0, & k \in [0, \bar{k}) \\
0, & k = \bar{k} \\
< 0, & k \in (\bar{k}, k'] 
\end{cases},$$

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

$$\frac{dV_{UC}(k)}{dk} < E(\theta) < \pi_2(\theta_2 - \theta_1)\frac{r^*}{\gamma_0 - \theta_1} = \frac{dV_N(k)}{dk}.$$  

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

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

$$r_1(k) = \frac{\theta_1 X_M(k) - \gamma_0 X_M(k) - (X_M(k) - k)\gamma X_M(k)}{X_M(k) - k},$$

and

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

is optimal conditional on $S(k) = \{\theta_1\}$, and so $V_M(k) = V_{UC}(k)$ for all $k \in [0, \bar{k}]$. This, together with equation (57), gives $\Phi = [0, \bar{k})$, 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

$$V_{UC}(k) = \pi_2(\theta_2 - r_2(k))(X_M(k) - k) > V_N(k) > E(\theta)k,$$

constraint (6) is satisfied. From Assumption 1 we have

$$\theta_1 - \gamma_0 - (X_M(k) - k)\gamma = \theta_1 - \frac{E(\theta) - \pi_1 \gamma k + \pi_1 \gamma_0 - r^*}{2\pi_1}$$

$$> \frac{r^* - (\pi_2 \theta_2 - \pi_1 \theta_1 + \pi_1 \gamma_0)}{2\pi_1}$$

$$> 0,$$

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

8.5 Corollary 6 and proof

Corollary 6. With the optimal contract, the firm’s gross rate of return on equity, $V(k)/k$ is strictly decreasing in $k$ for $k \in [0, \bar{k}]$ and constant in $k$ for $k \in (\bar{k}, \tilde{k}]$.

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

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

and

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

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

8.6 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 $r'_1(k) > 0$, $r'_2(k) < 0$. 56
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,

$$\frac{dr_1(k)}{dk} = \frac{(\theta_1 - \gamma_0)(X_M(k) - 1/2k)}{(X_M(k) - k)^2} - \frac{1}{2\gamma}$$

$$= \frac{2\pi_1\gamma(\theta_1 - \gamma_0)(E(\theta) - \pi_1\gamma_0 - r^*)}{(E(\theta) - \pi_1\gamma k - \pi_1\gamma_0 - r^*)^2} - \frac{1}{2\gamma}$$

$$\geq \frac{2\pi_1\gamma(\theta_1 - \gamma_0)}{E(\theta) - \pi_1\gamma_0 - r^*} - \frac{1}{2\gamma}$$

$$= \frac{2\pi_1\gamma}{2[E(\theta) - \pi_1\gamma_0 - r^*]} (2\pi_1\theta_1 - 2\pi_1\gamma_0 + \pi_1\theta_1 + r^* - \pi_2\theta_2 - \pi_1\gamma_0)$$

$$> 0,$$

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 firm net worth allows the contract to pay the investor more in the state of low output. How a larger $k$ would affect $r_2(k)$ is less obvious. From equation (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 $r_2(k)$.

8.7 Corollary 8 and proof

Corollary 8. With the optimal direct lending contract, $X_M(k) > X_N(k)$, for all $k \in [0, \tilde{k})$.

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

$$E(\theta)X_M(k) - r^*(X_M(k) - k) - \pi_1\gamma_0X_M(k) - \pi_1\gamma X_M(k)(X_M(k) - k)$$

$$> E(\theta)X_N(k) - r^*(X_N(k) - k),$$

which in turn gives $X_M(k) > X_N(k)$.

8.8 Proof for Proposition 3

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 (22) binds. Suppose at the optimum

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

Rewriting the bank’s net profits as

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

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 \( k \in B \) to make the bank strictly better off. A contradiction.

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

\[
\begin{align*}
\max_{B;L(k),k \in B} & \quad C_1 \int_{B} kdG(k) + C_2 \\
\text{subject to} & \quad B \subseteq [0, \bar{k}], \\
& \quad L(k) > 0, \forall k \in B, \\
& \quad \mu \int_{B} L(k)dG(k) = D, \\
& \quad \pi_2 \theta_2 k + \pi_2 (\theta_2 - R_L)L(k) \geq V(k), \forall k \in B,
\end{align*}
\]

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 (60) binds for all \( k \in B \), or the bank provides loans to all firms, \( B = [0, \bar{k}] \). Suppose not. Suppose the bank’s optimal plan is \( \{Z(k) : k \in B\} \) and \( B \subseteq [0, \bar{k}] \), and suppose a subset \( H \subseteq B \) of the firms get higher values than their reservation values through bank loans, where \( H \neq \emptyset \). From (23), \( Z(k) - k > L_0(k), k \in H \). 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 \( F \subseteq [0, \bar{k}] \setminus B \) 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
\[ \mu \int_{F \cup H} [\pi_1(\theta_1 - \gamma_0)(k + L_0(k)) + \pi_2 R_L L_0(k)] dG(k) \]
\[ - \mu \int_{H} [\pi_1(\theta_1 - \gamma_0)Z(k) + \pi_2 R_L (Z(k) - k)] dG(k) \]
\[ = \mu \int_{F} \pi_1(\theta_1 - \gamma_0) [k + L_0(k)] dG(k) - \mu \int_{H} \pi_1(\theta_1 - \gamma_0) [Z(k) - k - L_0(k)] dG(k) \]
\[ = \mu \int_{F} \pi_1(\theta_1 - \gamma_0) \left\{ k dG(k) + \int_{F} L_0(k) dG(k) - \int_{H} [Z(k) - k - L_0(k)] dG(k) \right\} \]
\[ = \mu \int_{F} \pi_1(\theta_1 - \gamma_0) k dG(k), \]
which is strictly positive given \( \theta_1 > \gamma_0 \).

**Step 4** Consider the case where \( D \geq D_1 \). Suppose \( B \subset [0, \bar{k}] \). From **Step 3**, the participation constraint (60) binds for all \( k \in B \). So
\[ L(k) = \frac{V(k) - \pi_2 \theta_2 k}{\pi_2 (\theta_2 - R_L)}. \]

From equation (59) we have
\[ D = \mu \int_{B} \frac{V(k) - \pi_2 \theta_2 k}{\pi_2 (\theta_2 - R_L)} dG(k) < \mu \int_{0}^{\bar{k}} \frac{V(k) - \pi_2 \theta_2 k}{\pi_2 (\theta_2 - R_L)} dG(k) = D_1. \]
A contradiction. So when \( D \geq D_1 \), we have \( B = [0, \bar{k}] \).

Now the total net worth of the firms \( \mu \int_{B} k dG(k) \) is constant. From (58) we know any feasible allocation is optimal. Thus any contract \( \{ B = [0, \bar{k}]; L(k), k \in B \} \) 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 (59) and (60) we have
\[ D \geq \mu \int_{B} \frac{V(k) - \pi_2 \theta_2 k}{\pi_2 (\theta_2 - R_L)} dG(k) \]
Thus \( B \subset [0, \bar{k}] \), which implies resource constraint (60) binds for all \( k \in B \). So
\[ L(k) = \frac{V(k) - \pi_2 \theta_2 k}{\pi_2 (\theta_2 - R_L)} \quad \forall k \in B, \]
or
\[ Z(k) = \frac{V(k) - \pi_2 R_L k}{\pi_2 (\theta_2 - R_L)} \quad \forall k \in B. \]

Now the optimal \( B \) solves the problem

\[ \max_{B \subseteq [0, \bar{k}]} \int_B k dG(k) \quad \text{(61)} \]

subject to

\[ \int_B L(k) dG(k) = \frac{D}{\mu}, \quad \text{(62)} \]

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

Let \( \lambda \) be the Lagrange multiplier of the constraint (62). The Lagrangian for the above problem is

\[ L = \int_B [k - \lambda L(k)] dG(k) + \frac{\lambda D}{\mu}. \]

Thus \( L \) is maximized when \( B \) includes all the \( k \)s that

\[ \frac{k}{L(k)} > \lambda, \]

and part or all of the \( k \)s that

\[ \frac{k}{L(k)} = \lambda. \]

By Corollary 6,

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

is strictly increasing with \( k \) for \( k \in [0, \bar{k}] \) and constant for \( k \in [\hat{k}, \bar{k}] \).

So \( B = [\hat{k}, \bar{k}] \) when \( D \in [D_0, D_1] \), where \( \hat{k} \) satisfies

\[ \mu \int_{\hat{k}}^{\bar{k}} L(k) dG(k) = D. \]

and \( B \subset [\hat{k}, \bar{k}] \) when \( D \in (0, D_0) \). Parts (i) and (ii) of the proposition are now proved.
8.9 $L_0(k)$ and $Z_0(k)$ are increasing in $k$

From (17),

$$V'(k) = \begin{cases} 
V_0'(k) = [E(\theta) + \pi_1 \gamma k - r^* - \pi_1 \gamma_0] / 2 + r^*, & \forall k < \tilde{k} \\
V_0'(k) = \pi_2(\theta_2 - \theta_1)r^*/(r^* - \theta_1), & \forall k \geq \tilde{k}
\end{cases}$$

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

$$L_0'(k) = \frac{V'(k) - \pi_2 \theta_2}{\pi_2 (\theta_2 - R_L)} > 0, \forall k \in [0, \bar{k}].$$

8.10 Lemma 9, proof, and Figure 25

**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 $k^* \in (0, \bar{k})$ and solves $Z_0(k^*) = X(k^*)$.

**Proof.** From equation (26), $Z_0(k)$ satisfies

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

(63)

From Proposition 2 and Corollary 7,

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

(64)

From (63) and (64),

$$Z_0(k) \geq X(k) \iff R_L \geq r_2(k).$$

(65)

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 \( r_2(\bar{k}) = R_D < R_L \). So there exists some \( k^* \in [0, \bar{k}] \) so that

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

which, given (65), proves part (ii) of the lemma.

Figure 25 depicts what Lemma 9 states for the two cases, \( R_D > \bar{R}_D \) and \( R_D \leq \bar{R}_D \), respectively.

![Figure 24: The demand function under \( R_D \geq \bar{R}_D \)](image-url)
Figure 25: The optimal size of the project: direct and bank lending
8.11 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 \( R_D < \tilde{R}_D \), it follows from Lemma 9 that \( Z(k) \geq Z_0(k) > X(k), \forall k \in [0, \bar{k}] \). And from Proposition 3, \( \tilde{k}_1(D) \) is weakly decreasing in \( D \) and \( \tilde{k}_2(D) \) weakly increasing in \( D \). Given equation (32), the lemma is proved. \( \square \)

Suppose \( R_D \geq \tilde{R}_D \). Then \( Q(D) \) is decreasing in \( D \) over the interval \([D'_1, D_1]\), where \( D'_1 = \mu \int_{k^*}^{k} L_B(k)dG(k) \) and \( k^* \) is defined in part (ii) of Lemma 9. This case is depicted in Figure 24 below.

8.12 The monotonicity of \( \lambda(k) \) in Figure 12

From equations (17) and (40) we have

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

Then for \( k \in (0, \bar{k}) \),

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

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

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

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

8.13 \( \tilde{Q}(D) > Q(D), \forall D \in (0, D_0) \)

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

\[
Q(D) = \mu \int_0^k [X(k) - k]dG(k) + \mu \int_k^{k_1} [Z_0(k) - k]dG(k) + \mu \int_{k_1}^k [X(k) - k]dG(k),
\]

and

\[
\tilde{Q}(D) = \mu \int_0^{\tilde{k}_1} [X(k) - k]dG(k) + \mu \int_{\tilde{k}_1}^{\tilde{k}_2} [\bar{X} - k]dG(k) + \mu \int_{k_2}^k [X(k) - k]dG(k),
\]

64
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

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

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

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

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

$$\int_{\tilde{k}_1}^{\tilde{k}_2} [Z_0(k) - k]dG(k) = (Z_0(k'') - k'') \int_{\tilde{k}_1}^{\tilde{k}_2} dG(k).$$

Given $k' < k''$ and so $\bar{X} - k' > Z_0(k'') - k''$, (66) and the above two equations give

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

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

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

### 8.14 After removing all interest rate regulations

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

From Figure 12, it holds that $\lambda(k) > R_0(k, L_0(k))$ for all $k \in [0, \bar{k}]$, and so $\lambda(k) > \max\{R_0(k, L_0(k')), k' \in [0, \bar{k}]\} > 0$, 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 $\bar{k}$ and goes to $-\infty$).

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

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

(67)

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 (44),

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

Suppose the economy is initially in the state of the benchmark model with $R_D = r^* = R_D$. Given $\lambda(k) > 0$ for all $k \in [0, \bar{k}]$ then, the optimal choice of the bank has $D = M$ and $B = [\tilde{k}_1(M), \tilde{k}_2(M)]$.

But given $Q(R_D) < M < Q_0(R_D)$ we have $\tilde{k}_2(M) < \bar{k}$. This implies the total demand for credit in the direct lending market is larger than the supply of credit (see equation (44)). This will increase $r^*$ and then $R_D$. Thus in equilibrium $R_D = r^* > R_D$. 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 (31)).

8.15 Formulating the equilibrium in a system of equations

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

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

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

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

A contradiction.
that solves the following system of equations:

(I) \((r^*, D^*)\) satisfy:
\[
r^* \geq R_D, \quad \text{and} \quad D^* = 0 \text{ if } r^* > R_D.
\]

(II) \(\tilde{k}\) solves
\[
\frac{(E(\theta) + \pi_1 \gamma \tilde{k} - r^* - \pi_1 \gamma_0)^2}{4 \pi_1 \gamma} + \tilde{k}r^* = \pi_2 (\theta_2 - \theta_1) \frac{\tilde{k}r^*}{r^* - \theta_1}
\]
and \(X(k)\) and \(V(k)\) are given by
\[
X(k) = \begin{cases} 
X_M(k) = \frac{E(\theta) + \pi_1 \gamma k - r^* - \pi_1 \gamma_0}{2 \pi_1 \gamma}, & \forall k < \tilde{k} \\
X_N(k) = kr^*/(r^* - \theta_1), & \forall k \geq \tilde{k}
\end{cases}
\]
and
\[
V(k) = \begin{cases} 
V_M(k) = \frac{(E(\theta) + \pi_1 \gamma k - r^* - \pi_1 \gamma_0)^2}{4 \pi_1 \gamma} + kr^*, & \forall k < \tilde{k} \\
V_N(k) = \pi_2 (\theta_2 - \theta_1) kr^*/(r^* - \theta_1), & \forall k \geq \tilde{k}
\end{cases}
\]

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

(a) \(B = [0, \tilde{k}]\) if \(D^* \geq D_1\). In this case,
\[
Z(k) \geq \frac{V(k) - \pi_2 R_L k}{\pi_2 (\theta_2 - R_L)}, \quad \forall k \in [0, \tilde{k}],
\]
and
\[
\mu \int_0^k Z(k) dG(k) = D^* + \mu \int_0^k kdG(k).
\]

(b) \(B = [\hat{k}, \tilde{k}]\) where \(\hat{k} \in (0, \tilde{k})\) if \(D^* \in (D_0, D_1)\). In this case,
\[
Z(k) = \frac{V(k) - \pi_2 R_L k}{\pi_2 (\theta_2 - R_L)}, \quad \forall k \in [\hat{k}, \tilde{k}],
\]
and
\[
D^* = \mu \int_{\hat{k}}^{\tilde{k}} \frac{V(k) - \pi_2 \theta_2 k}{\pi_2 (\theta_2 - R_L)} dG(k).
\]
(c) \( B \subset [\bar{k}, \bar{k}] \) if \( 0 \leq D^* < D_0 \). In this case,
\[
Z(k) = \frac{V(k) - \pi_2 R_L k}{\pi_2 (\theta_2 - R_L)}, \quad \forall k \in B,
\]
and
\[
\mu \int_B \frac{V(k) - \pi_2 \theta_2 k}{\pi_2 (\theta_2 - R_L)} dG(k) = D^*.
\]

(IV) The market for finance clears:
\[
\mu \int_{k \in [0, \bar{k}] \setminus B} [X(k) - k]dG(k) + \mu \int_{k \in B} [Z(k) - k]dG(k) = M.
\]

8.16 Bonds against bank loans: more counterfactuals

Figure 26 shows that if \( \theta_1 \) stayed constant at its 2002 level, the model would have missed about 80% of the increase in the bonds to bank loans ratio over 2002 - 2015. Figure 27 shows that if \( \pi_1 \) stayed constant at its 2002 level, the model would have produced at 2015 a bonds/(bonds+bank-loans) ratio that is 30% more than the observed level. Lastly, Figure 28 shows that if \( \theta_2 \) were to stay constant at its 2002 level, the model would have produced, at 2015, too much, about 50% more than the observed increase in the bonds to bank loans ratio.

Figure 26: Bonds/(bonds+bank+loans): \( \theta_1 \) constant at 2002 level
Figure 27: Bonds/(bank loans+bonds): $\pi_1$ constant at 2002 level

Figure 28: Bonds/(bank loans+bonds): $\theta_2$ constant at 2002 level
Reference


72