The London School of Economics and Political Science

Essays on Banking in Macroeconomics

Yu Yi

A thesis submitted to the Department of Economics
for the degree of Doctor of Philosophy

August 2, 2023

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of roughly 26,522 words.

Statement of inclusion of previous work

I can confirm that Chapter 3 is extended from a previous study for the Master of Research (MRes) degree in Economics at the LSE, awarded in July 2019.

Acknowledgment

I would like to express my sincere gratitude to my supervisor, Shengxing Zhang, for his invaluable guidance and mentoring. My work and PhD life benefited a lot from his continuous support. I am also very grateful to my mentors Benjamin Moll, Ricardo Reis, and Wouter Den Haan for their insightful advice and generous help.

I appreciate helpful comments and inspirations from participants at LSE Macro work-in-progress seminars and faulty at PHBS. Particularly, I have benefited a lot from fruitful conversations with Maarten De Ridder, Matthias Doepke, Jonathon Hazell, Kai Li, John Moore, Martin Oehmke, Mark Schankerman, Pengfei Wang, Zhiwei Xu, Hongda Zhong.

My thank also goes to my friends, especially Yuxiao Hu, Runhong Ma, Tiancheng Sun, Yifan Wang, Linchuan Xu, and friends in Sidney Webb House 2018-2019, for our discussions and mutual support during my PhD life.

Last but not the least, I thank my parents and my sister for their lifelong love.

Abstract

How does bank capital affect the relationship between bank concentration and risk taking? Chapter 1 presents a tractable dynamic model that incorporates heterogeneous financially constrained entrepreneurs and an imperfectly competitive banking sector. When the bank capital ratio exceeds the minimum requirement, reducing bank concentration leads to more risk taking; otherwise, the concentration-risk relationship is ambiguous. To explain the equilibrium characterization, I propose two mechanisms, a *net margin mechanism* and a *risk shifting mechanism*, whose direction depends on banks' optimal decisions regarding loan quantity and the accumulation of excess bank capital. Considering the *risk shifting mechanism* and the non-binding capital constraint, the model suggests a non-monotonic relationship between bank concentration and loan rate, which is supported by the micro-level evidence in the U.S. The two mechanisms also jointly establish a non-monotonic relationship between bank concentration and allocative efficiency. I discuss how efficiency and stability can be enhanced simultaneously.

Chapter 2 streamlines and refines the theoretical framework presented in Chapter 1, with a particular emphasis on efficient allocation. The model demonstrates that an increase in bank concentration leads to an increase in bank capital and potentially a non-binding capital constraint. I use the model to understand how bank concentration affects misallocation through the interaction between bank concentration and bank capital when the financial market is imperfect, which is referred to as the *"bank capital channel"*. This channel suggests that banks over-accumulate equity capital in terms of allocative efficiency.

Chapter 3 presents a general equilibrium model that incorporates the effects of collateral constraints and screening costs. The model reveals that a transitory productivity shock is amplified and prolonged through collateral constraints and a countercyclical average screening cost. This countercyclicality depends on firms' optimal decision between screened and unscreened capital. Moreover, higher screening costs serve to mitigate amplification while enhancing persistence.

Contents

List of Figures

List of Tables

Chapter 1

Clarifying the Relationship Between Bank Concentration and Risks: Role of Bank Capital

The relationship between bank concentration[1] and risk taking has been a widely researched topic in both theoretical and empirical literature, and remains an issue of debate among policymakers and academics. Many argue that reducing bank concentration encourages risk taking by squeezing bank profits and lowering franchise values (Corbae and Levine (2018)). Alternatively, some researchers contend that a more concentrated banking sector carries more risks (Carlson and Mitchener (2009)). In this regard, higher bank concentration leads to higher loan rates, which subsequently induce firms to take on additional risks (Boyd and De Nicolo (2005)).

The question whether bank concentration and risk taking are positively or negatively correlated remains pertinent as markedly different policies are implied from different perspectives. A positive correlation between concentration and risk taking prompts policymakers to remove barriers to competition in order to bolster efficiency and lower risk. Conversely, proponents of a negative concentration-risk relationship emphasize the trade-off between efficiency and low risks, highlighting concerns about a heavily concentrated banking sector and the need for alternative policy instruments to enhance efficiency while mitigating economic risk.

[1] In many countries, bank concentration has experienced a notable surge. In the United States, for instance, the number of banks decreased from 10,000 in 1997 to 5,000 in 2017. By contrast, the top three asset share, represented as the assets of the three largest banks as a percentage of total commercial banking assets, rose from 20% in 1997 to 35% in 2017.

This paper emphasizes the role of bank capital in shaping the relationship between concentration and risk-taking. Research has established a substantial correlation between bank capital and concentration in the United States (Yi (2022)). Additionally, bank capital plays a pivotal role in mitigating economy-wide risks. In light of this, the Basel III regulatory framework implemented stricter capital requirements in response to the 2008 financial crisis.

In this paper, I build a tractable dynamic model to explore the impact of bank concentration on risks and allocative efficiency by introducing the *net margin mechanism* and *risk shifting mechanism*. The analysis reveals that the relationship between bank concentration and risk-taking depends on whether banks hold excess capital above the minimum required level. Specifically, when the bank capital constraint is binding, the effect of bank concentration on risk-taking is ambiguous, and it leads to allocative inefficiency. In contrast, when the bank capital ratio exceeds the minimum requirement, a less concentrated banking environment motivates entrepreneurs to take risks, and the relationship between bank concentration and output becomes non-monotonic, with the two mechanisms moving in opposite directions with respect to efficiency.

The model incorporates two key agents: heterogeneous entrepreneurs and bankers. Entrepreneurs are short lived and protected by limited liability. They have access to two distinct types of projects, namely, prudent and gambling projects. The former guarantees a certain return, while the latter provides an excess return only upon successful completion. With limited enforcement and commitment, bankers facilitate the flow of credit among different entrepreneurs and compete in both loan and deposit markets à la Cournot.

In equilibrium, there are four types of entrepreneurs based on their levels of productivity: borrowing entrepreneurs who engage in risky ventures, borrowing entrepreneurs who exercise caution, lending entrepreneurs who provide credit to others, and autarky entrepreneurs who stay financially inactive. Entrepreneurs at the top of the productivity scale borrow funds and produce goods and services at full capacity, while also selecting the optimal investment project. Conversely, those situated at the lower end of the productivity scale will typically choose to deposit their endowments in banking institutions. Due to imperfect competition in the banking sector, there is a positive net margin between the loan and deposit rates, which encourages some entrepreneurs (autarky entrepreneurs) to withdraw from the credit market. Instead, they use their initial holdings to engage in production.

The model demonstrates that borrowing entrepreneurs who possess lower levels of productivity are more prone to engage in risky projects. This is a result of asymmetric information between bankers and entrepreneurs. Particularly, bankers lack access to knowledge regarding the productivity and investment preferences, and apply identical repayment rates to all the borrowers. The intuition behind this result lies in the fact that highly productive borrowing entrepreneurs tend to receive a larger portion of loan returns, motivating them to prioritize safer investment projects with higher expected profit margins. In contrast, borrowing entrepreneurs with low productivity are more inclined to invest in gambling projects and benefit from the limited liability protection afforded to them when their projects fail. This trade-off serves as a micro foundation for the *risk shifting mechanism* in the partial equilibrium, where higher loan rates lead to increased funding costs, and a greater proportion of risky loans and gambling projects. The *risk shifting mechanism* was initially proposed by Boyd and De Nicolo (2005), which demonstrate that higher bank concentration leads to increased loan rates and, consequently, heightened risks.

This paper contributes that, in the general equilibrium, bankers internalize entrepreneurs' best responses when deciding on the optimal loan quantities to be issued. Higher levels of bank concentration may lead to increased market power and, consequently, higher loan rates. However, the *risk shifting mechanism* in the partial equilibrium could mitigate the inclination of powerful bankers to increase loan rates. Due to the asymmetric information between bankers and entrepreneurs, bank capital and loan quantity are the only available instruments in the loan market. In highly concentrated banking sectors, accumulation of bank capital can not only reduce the moral hazard problem by lowering the loan rate, but it can also lead to maximum profit. In such scenarios, bankers choose to hold excess capital above the minimum requirement. As a result, the relationship between bank concentration and loan rate may become negative.

To examine the non-monotonic relationship between bank concentration and loan rates, this paper uses quarterly data on U.S. commercial banks spanning the period from 1994 to 2020. For each branch under a bank institution, loan rates for two specific types, namely 6-year auto loans and business loans, are matched with the Herfindahl-Hirschman Index (HHI) at the county level where each branch is located. The empirical findings reveal two key observations. Firstly, there exists a negative correlation between bank concentration and loan rate at HHI levels of approximately 0.7. Notably, a significant decrease of 0.023% in loan rates is observed when HHI increases from 0.6 to 0.7. This finding can be attributed to the interplay of the *risk shifting mechanism* and the presence of non-binding capital constraints. Secondly, when bank capital ratios are below the 80th quantile, an increase in HHI from 0 to 1 is associated with a significant 0.3% increase in loan rates. However, when bank capital ratios are high, no discernible correlation between bank concentration and loan rates is observed. The rationale behind this pattern lies in the *risk shifting mechanism*, wherein banks have reduced incentive to charge high loan rates when bank concentration is high. More specifically, banks tend to lower loan rates when they prioritize the accumulation of higher capital ratios.

In the partial equilibrium of the model, the *risk shifting mechanism* implies that higher loan rates lead to higher risk taking and lower efficiency. However, the interplay between the *risk shifting mechanism* and bank capital induces the relationship between bank concentration and risk to be dependent on the bank capital constraint. Specifically, when the bank capital constraint is binding, a more concentrated banking sector is associated with lower efficiency and higher risk taking. Conversely, increasing bank concentration may enhance efficiency and reduce risks when bankers accumulate excess capital above the required minimum. In contrast to Boyd and De Nicolo (2005), the direction of *risk shifting mechanism* in the general equilibrium is inherently uncertain and can exhibit divergent outcomes on risks.

The relationship between bank concentration and risk taking is influenced not only by the *risk shifting mechanism* but also by the *net margin mechanism*. As the banking sector becomes more concentrated, the wedge between the loan rate and deposit rate widens, leading to a greater proportion of autarky entrepreneurs. Despite their inefficiencies, these entrepreneurs tend to invest in prudent projects due to their reliance on internal financing. Consequently, as bank concentration increases, the *net margin mechanism* leads to lower risks and inefficiency.

When considering both the *risk shifting mechanism* and *net margin mechanism*, I find that a less concentrated banking sector is associated with greater efficiency and ambiguous risk taking when the bank capital constraint is binding, as the two mechanisms have opposite effects on risks. Under the calibration to match U.S. moments, the magnitudes of the two mechanisms are quantitatively similar. Conversely, when the bank capital constraint is non-binding, a more concentrated banking sector leads to increased risk taking and a hump-shaped output, with the two mechanisms having opposite effects on efficiency.

According to the model, reducing bank concentration is considered relatively safe when the bank capital constraint is binding. In this scenario, the reduction in bank concentration is expected to increase efficiency without significantly affecting risks when the bank capital ratio is close to the minimum requirement. However, when the capital ratio exceeds the minimum requirement, enhancing efficiency and lowering risks may require a simultaneous reduction in bank concentration and an increase in the minimum bank capital requirement.

Related Literature

This paper contributes to the literature on the relationship between bank concentration and risk taking, which remains unsettled. It is important to note that bank competition and bank concentration are distinct concepts, although the latter is often considered to be suggestive of the former. A strand of literature suggests that there exists a positive correlation between bank concentration and stability (Hellmann et al. (2000); Beck et al. (2003); Agoraki et al. (2011); Tabak et al. (2012); Jiang et al. (2017); Carlson et al. (2022); Beck et al. (2013)), where they provide related empirical evidence with indirect measures of bank competition and stability. Another viewpoint, known as the concentration-fragility view, posits that increased bank competition can lead to greater economic stability (De Nicolò et al. (2004); Beck et al. (2006); Carlson and Mitchener (2009); Craig and Dinger (2013)). This paper re-examines the relationship between bank concentration and risks, and demonstrates that this relationship is contingent upon the binding nature of bank capital constraint. Specifically, I show that the correlation between these two variables is ambiguous when the bank capital constraint is binding, but that increasing bank concentration can reduce risks when the bank capital ratio exceeds the minimum requirement. The model implications are partially in line with the concentration-stability view.

The paper is most related to theories developed by Boyd and De Nicolo (2005), Corbae and Levine (2018), and Martinez-Miera and Repullo (2010). Boyd and De Nicolo (2005) support the concentration-fragility view by arguing that lower lending rates will reduce entrepreneurs' borrowing costs and motivate them to take less risk. In this paper, I provide a micro-foundation for the *risk shifting mechanism* and suggest that it might not be the dominant mechanism that determines the relationship between bank concentration and risk taking in the general equilibrium. According to Corbae and Levine (2018), however, banks take more risks in a more competitive market when their profit margins are squeezed and their franchise values fall. However, they fail to acknowledge the existence of a loan market. Martinez-Miera and Repullo (2010) find a U-shaped relationship between bank concentration and stability when the model allows for imperfect correlations among loan defaults. Unlike prior studies, this paper considers the excess accumulation of bank capital, and the impact of bank concentration on risks depends on whether the capital constraint is binding or not.

The relationship between bank competition and efficiency has been empirically examined by Jayaratne and Strahan (1996), Black and Strahan (2002), Diez et al. (2018), and Joaquim et al. (2019). This paper contributes to the existing literature by describing how the *net margin mechanism* and *risk shifting mechanism* work together to determine the impact of bank concentration on the real economy, which turns out to be non-monotonic. The *risk shifting mechanism* is the key driver of the local optimum of output.

This theoretical work is related to the heterogeneous agent models. The entrepreneurs' side of the model is built on Angeletos (2007), Kiyotaki and Moore (2019) and Moll (2014). In particular, Moll (2014) eases the *i.i.d.* assumption of productivity and shows how the persistence of idiosyncratic productivity shock affects misallocation. This paper builds upon their models by incorporating the bankers' perspective in this setting and examining how bank concentration affects efficiency and risks when the financial market is imperfect.

Numerous studies have explored the concept of imperfect competition in the banking sector, including those by Drechsler et al. (2017), Lagos and Zhang (2022), Van Hoose et al. (2010), Corbae et al. (2021), and Head et al. (2022). Drechsler et al. (2017) adopt the framework of Dixit and Stiglitz (1977) by assuming the representative household substitutes deposits across banks imperfectly. Lagos and Zhang (2022) incorporate bargaining power into the model to account for imperfect bank competition. Based on Burdett and Judd (1983), Head et al. (2022) examine the impact of bank concentration on the transmission of monetary policy. Corbae et al. (2021) develop a market structure where big banks interact with small fringe banks. This paper is closely related with Van Hoose et al. (2010), in which both papers assume that banks compete à la Cournot. Nonetheless, there are noticeable distinctions in at least two aspects. Firstly, imperfect competition is present in both the deposit and loan markets. Secondly, the elasticities of loan demand and deposit supply are determined endogenously through the decisions of entrepreneurs.

This is not the first theory that examines bank capital. Some papers focus on static models where bank capital is not a choice but rather a fixed parameter (Brunnermeier and Koby (2018)). Other papers impose an exogenous law of motion on bank capital (Li (2019)). Meanwhile, some papers assume that bank capital constraints are always binding (Repullo (2004)). In contrast, bank capital is endogenously determined in this model by optimizing dividend payouts and retained earnings. This setup enables the examination of the relationship between bank concentration and bank capital, as well as the potential for a non-binding capital constraint.

A substantial body of literature has been dedicated to exploring non-binding capital constraints. According to empirical evidence, banks voluntarily hold more capital than what is required by capital regulations and adjust their capital ratio independently. For example, Alfon et al. (2004) demonstrate that banks in the U.K. increased their capital ratios in the last decade, despite a reduction in the minimum capital requirement. Flannery and Rangan (2008) find that the U.S. banking sector experienced a significant capital buildup, with half of the large bank holding companies more than doubling their equity ratios over the same period. This paper finds that the non-binding capital constraint is instrumental in elucidating the impact of bank concentration on risks, as well as the non-monotonic relationship between bank concentration and loan rate. The theoretical underpinnings of why banks accumulate excess capital are akin to Yi (2022), which highlights the substitution effect between bank capital and deposits. However, this paper highlights the *risk shifting mechanism* that motivates banks to hold even more capital.

The rest of the chapter is organized as follows. In section 1.1, I lay out the model environment. Section 1.2 characterizes the symmetric model equilibrium and discusses the implication of the *risk shifting mechanism* and *net margin mechanism.* Section 1.3 calibrates the model quantitatively, under which setting I study how the two mechanisms shape the impact of bank concentration on efficiency and risks. In section 1.4, I present micro-data evidence on the relationship between bank concentration and loan rate, which supports *risk shifting mechanism* in the model. Further, I give policy implications. Section 1.5 concludes the chapter. Section 1.6 presents the Appendix.

1.1 Environment

Consider a model economy with discrete time and infinite horizon, where time is indexed by $t = 0, 1, 2, \cdots$. The model aims to capture the credit structure of an economy comprising three distinct types of agents, namely entrepreneurs, bankers, and capital suppliers. Entrepreneurs are short lived, while bankers and capital suppliers are long lived. The economy consists of two distinct goods: consumption goods and capital goods. The available endowment of capital goods is depleted in each time period to generate the production of consumption goods. During each period, bankers intermediate resources among a continuum of ex-ante heterogeneous entrepreneurs, while capital suppliers provide capital to both bankers and entrepreneurs.

1.1.1 Entrepreneurs

There is a continuum of short lived entrepreneurs, who are indexed by their productivity z. The productivity of entrepreneurs is assumed to follow an exogenous distribution $G(z)$ in the domain of $[z_{min}, z_{max}]$, which is identically and independently distributed ($i.i.d.$). Entrepreneurs are risk neutral and thus maximize the expected consumption

$$E_{t-1}[c_t]$$

At period t, entrepreneurs of this generation are endowed with two production technologies, namely, a prudent project and a gambling project. The former generates a return of z per unit of capital input, while the latter yields αz with probability p, and nothing otherwise. The success of the gambling project depends on the realization of an idiosyncratic shock. Following Hellmann et al. (2000):

Assumption 1.1. $\alpha > 1$ *and* $\alpha p < 1$.

The aforementioned assumption suggests that, in the event of success, the gambling project would provide a higher return compared to the prudent project, but it would result in a lower expected return overall. Entrepreneurs who invest in gambling projects are shielded by limited liability, thereby ensuring that they die with nothing if the project fails.

At the middle of each period, some entrepreneurs prefer to borrow, while others may lend. I assume that borrowers are unable to commit, and lenders are unable to enforce their promises. To make banks function as financial intermediaries, I assume that bankers have the ability to enforce and commit. The entrepreneurs may obtain external financing from the bankers, and repay their debt at r_t^b once the project has been successfully completed. On the other hand, entrepreneurs deposit their resources in banks and receive a return of r_t^d. r_t^b and r_t^d represent the loan rate and deposit rate, respectively, both measured in units of capital goods. Entrepreneurs will give birth to offspring after producing and trading in the loan and deposit markets. Entrepreneurs of this generation will consume a certain percentage (s) of their net returns and invest the remainder in capital. The capital is then transferred to the next generation of entrepreneurs and distributed equally among them. Different from Moll (2014), there is no heterogeneity of wealth among the entrepreneurs of the same generation. This homogeneity of wealth is not a necessary assumption, but it is one way to ensure a non-zero endowment for every entrepreneur, even if their parents leave nothing for them.

Additionally, entrepreneurs face a borrowing constraint

$$k_t \leq \lambda a_t, \ \lambda \geq 1 \tag{1.1}$$

Finite λ implies an imperfect financial market, which captures the intuition that entrepreneurs are constrained by their initial endowment when borrowing. The parameter λ measures the efficiency of the financial market. In the extreme case where $\lambda = 1$, the financial market is shut down and all the entrepreneurs remain autarky. When λ converges to ∞, the financial market is perfect. I denote $\theta_t = \frac{k_t}{a_t}$, where θ_t represents the entrepreneurs' actual leverage ratio. Entrepreneurs' decisions are then characterized by θ_t and p.

1.1.2 Bankers

The key assumption in the banking sector is imperfect competition. To characterize this, I assume there are $1, 2, \cdots, M$ long-lived bankers in the economy[2], each of whom competes for the quantity of loans Q_{it}^L and deposits Q_{it}^D à la Cournot[3]. When $M = 1$, the economy consists of a monopoly bank and when M converges to infinity, the banking sector is perfectly competitive. At the beginning of each period, each banker i is endowed with some equity capital N_{it}. Bankers are risk neutral and derive utility from dividend payouts

$$\sum_{t=0}^{\infty} \beta^t c_{it}^b$$

Bankers act as financial intermediaries and facilitate lending and borrowing between lending entrepreneurs and borrowing entrepreneurs. Using equity capital and deposits, the banker issues a loan contract, which could be either safe or risky. The fraction of risky loans is denoted by v_{rt}. Bank equity capital is accumulated only through retained earn-

[2]The parameter M serves as a unified measure encompassing both deposit market power and loan market power. As shown in subsequent sections of this chapter, I incorporate both deposit market power and loan market power to establish a quantitative non-monotonic relationship between bank concentration and loan rates. Deposit market power plays a crucial role in this context as it serves as the dominant mechanism underlying the accumulation of excess capital in highly concentrated banking sectors. Moreover, loan market power holds significance as it directly influences bankers' decision-making pertaining to loan rates. By considering both aspects, the model captures the comprehensive dynamics of the relationship between bank concentration and loan rates.

[3]Following Van Hoose et al. (2010), I model imperfect bank competition using Cournot competition. It is a simple approach to examine the banking sector between the extremes of perfectly competitive banking ($M = \infty$) and monopoly banking ($M = 1$). The extreme cases under Cournot competition are equivalent to those when applying Bertrand competition (monopoly banking with $M = 1$ and perfectly competitive banking with $M > 1$). However, to generate an intermediate market structure with price competition, additional frictions may be required.

ings.[4] Balance sheet identity of banker i then follows

$$Q_{it}^L = Q_{it}^D + N_{it} \tag{1.2}$$

The balance sheet items at the beginning of the period t are summarized in Table 1.1. I assume that each banker can fully diversify the idiosyncratic risks and analyze the equilibrium in regions in which no bankers default on deposits. At the end of period t, bankers' dividend payouts and retained earnings are funded by the return of their operations in the loan and deposit market. The intratemporal decision is simplified to a standard consumption and saving problem, where banker i faces a budget constraint

$$c_{it}^b + q_t N_{it+1} \leq (1 + r_t^b) q_t (1 - v_{rt}) Q_{it}^L + p(1 + r_t^b) q_t v_{rt} Q_{it}^L - (1 + r_t^d) q_t Q_{it}^D \tag{1.3}$$

RHS terms represent banker i obtains income—returns from issuing safe and risky loan contracts, minus the repayment back to the depositors, which is used for financing the LHS variables—consumption of dividends and accumulation of banker's equity capital. The price of capital is q_t. To simplify equation (1.3), I define a new variable

$$p_t^e = (1 - v_{rt}) \cdot 1 + v_{rt} \cdot p, \tag{1.4}$$

the intuition of which is the expected probability of loan repayment. By construction, p_t^e represents a weighted average of repayment probabilities between safe loans and risky loans. Equation (1.3) then becomes

$$c_{it}^b + q_t N_{it+1} \leq q_t \{(1 + r_t^b) p_t^e Q_{it}^L - (1 + r_t^d) Q_{it}^D\}. \tag{1.3'}$$

Assets	Liabilities
Safe loans ($(1 - v_{rt}) Q_{it}^L$)	Deposits (Q_{it}^D)
Risky loans ($v_{rt} Q_{it}^L$)	Equity capital (N_{it})

Table 1.1: Banker's Balance Sheet

There is asymmetric information between entrepreneurs and bankers, who are uninformed of the types of entrepreneurs, including their productivity and project choices. As a result of this, bankers charge a loan rate that applies to the entire population of entrepreneurs. However, the amount of loans they can issue is limited by the minimum capital requirement

[4]One can allow for the equity issuance by new equity holders, which does not alter the mechanisms of the model.

$$N_{it} \geq \kappa Q_{it}^{L} \tag{1.5}$$

where κ measures the flexibility of the minimum capital requirement. According to the minimum capital requirement, at least a fraction κ of bank loans should be financed by capital. The Basel Committee on Banking Supervision introduced the first framework for the minimum capital requirement for controlling market risk at the end of the twentieth century. This constraint was imposed to ensure that banks maintained a sufficient level of regulatory capital to absorb economic losses. The capital constraint here is simplified from a minimum requirement over capital to a risk-weighted asset ratio. Incorporating the capital to risk-weighted asset ratio does not alter the main mechanism of the model.

1.1.3 Capital Supplier

There is a continuum of capital suppliers, who are endowed with $\overline{K}$ units of capital. At the end of each period t, capital suppliers provide capital to entrepreneurs and bankers in a perfectly competitive capital market.

1.2 Equilibrium Characterization

This section presents the model equilibrium characterization and uses the results to discuss how bank concentration impacts risk taking through two channels: a *"net margin mechanism"* and a *"risk shifting mechanism"*.

1.2.1 Entrepreneurs' Side

I will first derive the conditions under which gambling exists in equilibrium. Entrepreneurs, whose objectives are to maximize their expected consumption, choose to borrow when gambling, so as to benefit from limited liability. Otherwise, they prefer to invest in the prudent project that will provide them with a higher expected return. Because of the linearity of the production function, borrowing entrepreneurs who gamble always borrow up the borrowing limits. The incentive compatible condition then follows that borrowing entrepreneurs who gamble should obtain a higher expected return than if they self-finance and produce [5]

$$p(\alpha z\lambda a - q(1+r^b)(\lambda-1)a) \geq za,$$

which yields a lower bound on productivity (z_2)

$$z \geq \frac{(\lambda-1)p}{\lambda\alpha p - 1} q(1+r_b) \equiv z_2 \tag{1.6}$$

Further, borrowing and gambling should dominate borrowing and staying prudent

$$p(\alpha z\lambda a - q(1+r^b)(\lambda-1)a) \geq z\lambda a - q(1+r^b)(\lambda-1)a,$$

which gives a upper bound on productivity (z_3) by a rearrangement of the above inequality:

$$z \leq \frac{(\lambda-1)(1-p)}{\lambda(1-\alpha p)} q(1+r_b) \equiv z_3 \tag{1.7}$$

When the two incentive compatible conditions are met, borrowing entrepreneurs are motivated to gamble. In the equilibrium, the two conditions simultaneously hold if and only if $z_2 < z < z_3$, which gives the following assumption:

Assumption 1.2. $\frac{(\lambda-1)p}{\lambda\alpha p-1} < 1$.

According to Assumption 1.2, entrepreneurs with productivity between z_2 and z_3 choose to gamble. There are three parameters involved in Assumption 1.2, namely α, p and λ. The inequality is more likely to hold when α is larger, λ is larger or p is larger. Intuitively, the marginal benefit of gambling is higher when excess return α or success probability p is higher. When the asset pledgeability λ is higher, heterogeneity among borrowing entrepreneurs declines. Consequently, bankers are able to extract more profit from borrowers, inducing them to gamble.

Given the deposit and loan rate, entrepreneurs' financial decisions (borrow or lend) are fully characterized in Proposition 1.3.

Proposition 1.3. *There are three productivity cutoffs z_1, z_2 and $\overline{z_3}$, which characterize*

- *The capital demand for individual entrepreneur is:*

$$k = \begin{cases} \lambda a & z \geq z_2 \\ a & z_1 \leq z \leq z_2 \\ 0 & z \leq z_1 \end{cases}$$

[5] I neglect the time index here for simplification

- *The entrepreneurs with productivity between z_2 and z_3 will gamble, while those with $z > \overline{z_3}$ and $z_1 \leq z \leq z_2$ will invest in the prudent project.*

The productivity cutoffs are defined by $z_1 = q(r^d + 1)$, $\overline{z_3} = Min\{z_3, z_{max}\}$, and z_2 follows Equation (1.6).

The cutoff property relies heavily on the constant return to scale of the production function. The optimal capital demand decision is at corners according to Proposition 1.3: it is zero for entrepreneurs with low enough productivity ($z < z_1$), maximum amount allowed by the borrowing constraint for those with high enough productivity ($z > z_2$), and initial wealth for those with intermediate level productivity ($z_1 < z < z_2$). Capital demand distinguishes two types of marginal entrepreneurs. For the entrepreneurs with productivity z_1, the return on each additional unit of capital $\frac{z}{q}$ equals the opportunity cost of not depositing that capital in the bank $r^d + 1$. The entrepreneurs with productivity z_2, however, are indifferent between gambling and using their own capital to engage in production. Assumption 1.2 indicates that $z_2 < q(1 + r^b)$, therefore, the borrowing entrepreneurs with z_2 will not invest in the prudent project.

Entrepreneurs with productivity levels exceeding z_2 may select different projects. While investing in the prudent project will yield a higher expected return, limited liability protection is not provided. It follows from Proposition 1.3 that entrepreneurs with productivity above $\overline{z_3}$ will invest in prudent projects, whereras those with $z_2 < z < \overline{z_3}$ will gamble. Since bankers are unable to observe the productivity types and projects selected by entrepreneurs, they set a uniform interest rate for all entrepreneurs on the loan market. Considering that entrepreneurs with productivity above $\overline{z_3}$ receive a high fraction of the net return from a loan contract, they prefer to invest in a project that yields a higher expected return. Entrepreneurs with $z_2 < z < \overline{z_3}$ will, however, gamble because it is unlikely that they will receive much benefit from the loan issuance. An extreme case is that those with productivity of $q(1 + r^b)$ obtain nothing from the loan contract if they invest in the product project, but receive a positive return when the project is successful if they gamble.

It is now sensible to refer to the entrepreneurs with productivity below z_1 as the lending entrepreneurs, those with productivity above z_2 as the borrowing entrepreneurs, and those with productivity in between as the autarky entrepreneurs. The productivity of lending entrepreneurs is so low that it is not worthwhile for them to produce and instead deposit all their endowment in banks. Borrowing entrepreneurs are willing to borrow

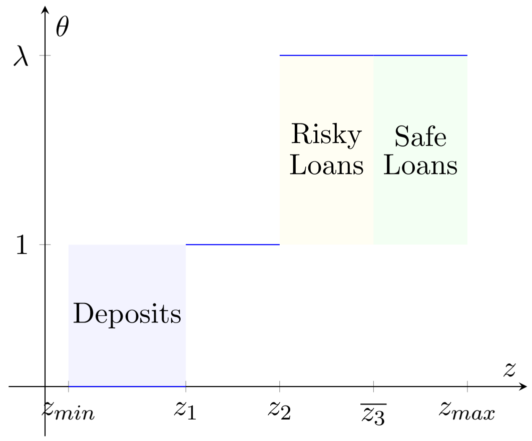


Figure 1.1: Entrepreneurs' Leverage and Project Choice

up to the asset pledgeability λ with their productivity being high. The introduction of imperfect competition in the banking sector results in the emergence of the third type of entrepreneur. Generally, banks charge a positive net margin between their loan and deposit rates, which induces some entrepreneurs to opt neither to borrow nor to lend. In this regard, entrepreneurs who have productivities between z_1 and z_2 are referred to as autarky entrepreneurs. Because the autarky entrepreneurs employ their own funds to engage in production, they choose the prudent project with a higher expected return.

Entrepreneurs' financial and intertemporal decisions generate an endogenous loan demand and deposit supply, as well as a law of motion of aggregate entrepreneurial capital, as shown in Figure 1.1. Loan contracts could either be risky or safe, depending on the project choice of borrowers. There is an extreme case ($\overline{z_3} > z_{max}$) in which all loans are risky. In the next section, I will discuss how the equilibrium behaves if there are only risky loans, as well as if there are both risky and safe loans. As entrepreneurs of each generation have the same initial wealth, the aggregate entrepreneurial capital demand equals the individual demand.

Lemma 1.4. *Denote Q_t^L and Q_t^D as the loan size and deposit size respectively. Aggregate quantities $\{Q_t^L, Q_t^D, a_{t+1}\}$ satisfy:*

$$Q_t^L = (1 - G(\overline{z_{3t}}))(\lambda - 1)a_t \tag{1.8}$$

$$Q_t^D = G(z_{1t})a_t \tag{1.9}$$

$$
\begin{aligned}
q_t a_{t+1} =& s\{\int_{z_{min}}^{z_{1t}} q_t(1+r_t^d)dG(z_t) + \int_{\overline{z_{3t}}}^{z_{max}} \left[\lambda[z_t - q_t(1+r_t^b)] + q_t(r_t^b+1)\right] dG(z_t) \\
&+ \int_{z_{1t}}^{z_{2t}} z_t dG(z_t) + p\int_{z_{2t}}^{\overline{z_{3t}}} \left[\alpha\lambda z_t - (\lambda-1)q_t(r_t^b+1)\right] dG(z_t)\} a_t
\end{aligned} \tag{1.10}
$$

Equation (1.8) implies that the total loan demand depends on three elements: the proportion of borrowing entrepreneurs, the amount each entrepreneur borrows and entrepreneurs' initial capital holdings. Similarly, the deposit supply is given by lending entrepreneurs' total capital holding as described by Equation (1.9). Equation (1.10) captures the law of motion for the aggregate entrepreneurial capital demand, where the wealth of entrepreneurs in the next generation $q_t a_{t+1}$ depends on their saving rate s and net return.

1.2.2 Bankers' Side

Rewriting the first-order conditions yields the optimal loan and deposit rate as a function of the mark-up (-down) on banker i's marginal cost (benefits)

$$
1 + r^d + Q_i^D \frac{\partial r^d}{\partial Q_i^D} = \mu_i \tag{1.11}
$$

$$
p_e[(1+r^b) + Q_i^L \frac{\partial r^b}{\partial Q_i^L}] + (1+r^b) Q_i^L \frac{\partial p_e}{\partial r^b}\frac{\partial r^b}{\partial Q_i^L} = \mu_i + \kappa\chi_i \tag{1.12}
$$

where $q\mu_i$ is the Lagrangian multiplier on the balance sheet identity and $q\chi_i$ is the Lagrangian multiplier on the bank capital constraint. Equation (1.11) indicates that the deposit rate depends on the elasticity of deposit supply and the multiplier on the balance sheet identity, which captures the marginal cost and benefit of deposits, respectively. As shown in Equation (1.12), the marginal cost of issuing loans is a tightening of both the balance sheet identity and bank capital constraint by κ. The benefits of issuing more loans are influenced by the elasticity of loan demand and the expected probability of loan repayment (second term on the LHS of Equation (1.12)).

Proposition 1.5 (*Risk Shifting Mechanism*). *Assume $\frac{zg(z)}{1-G(z)}$ is increasing, then*

$$
\frac{\partial v_r}{\partial r^b} \geq 0 \;\&\; \frac{\partial p_e}{\partial r^b} \leq 0,
$$

where the equality holds when $\overline{z_3} = z_{max}$

Based on Proposition 1.5, in partial equilibrium, when the lending rate is higher, the fraction of risky loans is higher, whereas the expected probability of loan repayment is lower. Following a spike in loan rates, more entrepreneurs are motivated to gamble, resulting in a higher proportion of risky loans. The *risk shifting mechanism* here is very similar to that proposed by Boyd and De Nicolo (2005). Their risk-incentive mechanism is completely based on the functional assumption of project return, whereas I provide a micro-foundation from the perspective of the entrepreneurs. When $\overline{z_3} = z_{max}$, all loans are risky, $v_r = 1$ and $p_e = p$, and the *risk shifting mechanism* is shut down. As I will demonstrate in the following section, even though *risk shifting mechanism* acts, it is not necessarily the dominant effect in the equilibrium.

Denote the aggregate loan demand elasticity $\epsilon^b = -\frac{\partial log Q^L}{\partial log(1+r^b)}$, the aggregate deposit supply elasticity $\epsilon^d = \frac{\partial log Q^D}{\partial log(1+r^d)}$, and market share of loans and deposits that each banker holds as s_i^b and s_i^d respectively. Equations (1.11) and (1.12) then become:

$$1 + r^d = \frac{\epsilon^d}{\epsilon^d + s_i^d}\mu_i \tag{1.13}$$

$$p_e(1 + r^b) = \frac{\epsilon^b}{\epsilon^b - s_i^b[1 + \frac{\partial log p_e}{\partial(1+r^b)}]}(\mu_i + \kappa\chi_i) \tag{1.14}$$

The above two equations indicate that the optimal loan (deposit) rate represents a mark-up (-down) over the marginal cost (benefit) of issuing the loan (deposit). As long as the bankers obtain a greater share of either the loan or deposit market, the markup or markdown will be higher. In a perfectly competitive banking sector where s_i^b and s_i^d converge to zero, there will be no mark-up (-down). The *risk shifting mechanism*, however, generates a new term in Equation (1.14) that lowers the markup. The reason for this is that bankers know that if they set a loan rate too high, entrepreneurs are more likely to gamble. When the banking sector becomes highly concentrated, bankers' motives to raise loan rates will be mitigated.

There is a standard Euler equation derived from the optimal condition for bank capital

$$q_t = \beta q_{t+1}(\mu_{it+1} + \chi_{it+1}) \tag{1.15}$$

Accumulating one unit of bank capital today costs q_t, which relaxes the balance sheet identity and bank capital requirement by multipliers tomorrow.

1.2.3 Steady State Equilibrium

In this section, I will turn to the general equilibrium, where I focus on the symmetric equilibrium throughout the chapter.

Definition 1.6 (Symmetric Equilibrium)**.** A *Symmetric Equilibrium* in the economy consists of a sequence of policy function of bankers' consumption, banker's equity capital holding $\{c^b_{it+1}, N_{it+1}\}_{t=0}^{\infty}$, a sequence of aggregate quantities $\{a_{t+1}, Q^D_t, Q^L_t\}_{t=0}^{\infty}$, a sequence of interest rates $\{r^b_t, r^d_t\}_{t=0}^{\infty}$, and a sequence of price $\{q_t\}_{t=0}^{\infty}$ such that:

1. Entrepreneurs maximize expected life-time utility given loan rate, deposit rate and the price of capital;
2. Bankers maximize their life-time utility given constraints (1.2) (1.3) (1.5) by competing for loans and deposits;
3. Bankers choose the same quantities for all assets and liabilities;
4. Market clearing condition for
   - loan market: $\sum_{i=1}^{M} Q^L_{it} = Q^L_t$;
   - deposit market: $\sum_{i=1}^{M} Q^D_{it} = Q^D_t$;
   - capital market: $\sum_{i=1}^{M} N_{it} + a_t = \overline{K}$.

The term symmetry refers to the fact that in equilibrium, there is no heterogeneity among the bankers. To begin with, it would be interesting to examine how symmetric equilibrium with a perfectly competitive banking sector differs from the benchmark scenario where there is no risk taking involved. ($\alpha = p = 1$).

Corollary 1.7. *Assume $\alpha p \lessapprox 1$. When $M \to \infty$ and $\kappa = 0$, there is a positive net margin ($r^b > r^d$) and a non-zero fraction of autarky entrepreneurs ($z_2 > z_1$), where:*

$$1 + r^b = \frac{1 + r^d}{p} \tag{1.16}$$

$$z_2 = qp(1 + r^b)\frac{\lambda - 1}{\lambda \alpha p - 1} > q(1 + r^d) = z_1 \tag{1.17}$$

When the banking sector is perfectly competitive and there is no risk taking, the model equilibrium is equivalent to Moll (2014) without labor. There is no positive margin and a single cutoff determines who will be creditors and lenders. With a slight deviation ($\alpha p \lessapprox 1$) from the benchmark, bankers will charge a positive wedge between the loan rate and deposit rate, which I interpret as risk premium. Additionally, due to the inefficiency of the gambling project ($\alpha p < 1$), there will be some autarky entrepreneurs. The risk taking motive is therefore undesirable, not only because the gambling project is inefficient, but also because more resources are allocated to inefficient producers.

The next step is to discuss how imperfect competition in the banking sector impacts risks. An indicator of entrepreneurial risk taking in the model economy is the amount of capital invested in the gambling project, which I refer to as risky capital. The risky capital at period t is denoted as rc_t, whose size in equilibrium is as follows:

$$rc = v_r[\overline{K} - (1 - v_a)(\overline{K} - N)] \tag{1.18}$$

where v_a is defined as the fraction of autarky entrepreneurs. The relationship between bank concentration and risky capital depends on three factors:

$$\frac{\partial rc}{\partial M} = (1 - v_a)\overline{K}\frac{\partial v_r}{\partial M} - v_r\overline{K}\frac{\partial v_a}{\partial M} + v_a v_r \frac{\partial N}{\partial M} \tag{1.19}$$

The sign of the second element in equation (1.19) depends on how the proportion of autarky entrepreneurs is affected by bank concentration. If the banking sector is more concentrated, the net margin is wider, and thus there are more autarky entrepreneurs. I refer to this as the *"net margin mechanism"*. The validity of this channel has been established in Yi (2022) by assuming that there is no risk-taking motive and a uniform distribution of productivity. This chapter will demonstrate the *net margin mechanism* quantitatively in the following section. Autarky entrepreneurs who use their own money for production always invest in prudent projects. Consequently, the *net margin mechanism* leads to a negative correlation between bank concentration and risk taking.

The first term in equation (1.19) relates to how bank concentration affects the percentage of risky loans. According to Proposition 1.5, bank concentration affects the loan rate, which in turn affects the proportion of risky loans through the *risk shifting mechanism.* However, there is still a great deal of uncertainty regarding the impact of bank concentration on loan rates. The third element in equation (1.19) relies on the sensitivity of bank capital to the number of bankers. In the following section, I will clarify that this term is positive while quantitatively negligible.

Therefore, the concentration-risk relationship is an issue of how bank concentration affects the loan rate and the magnitude of the *risk shifting mechanism* and the *net margin* *mechanism*. When the loan rate is higher in a highly concentrated banking sector, bank concentration has a positive impact on risk through the *risk shifting mechanism*, while it has a positive impact through the *net margin mechanism*. With the two mechanisms operating in opposite directions, the overall effect is ambiguous. If a higher bank concentration results in a lower loan rate, the *risk shifting mechanism* and the *net margin mechanism* both lead to a negative relationship between bank concentration and risks.

Role of Bank Capital. Bankers lack knowledge of entrepreneurs' productivities and project choices, so they only rely on two instruments: bank capital and loan rate. Due to the *risk shifting mechanism*, bankers are not motivated to raise loan rates too high in an imperfectly competitive banking sector. Instead, the expected probability of loan repayment will be higher as bankers accumulate excess capital above the minimum capital requirement when the bank concentration is large. The binding nature of the bank capital constraint will directly determine how loan rates and risk taking are affected by bank concentration through *risk shifting mechanism*. I will provide a quantitative explanation of the concentration-risk relationship in the following section.

1.3 Quantitative Analysis

This section first calibrates the parameters in the model, followed by a quantitative analysis of how bank concentration impacts risks through the *risk shifting mechanism* and *net margin mechanism*. I will discuss the case where there are only risky loans ($\overline{z_3} = z_{max}$) and the case where both safe and risky loans are present ($\overline{z_3} < z_{max}$). Quantifying the two mechanisms allows me to further study the impact of bank concentration on efficiency.

1.3.1 Calibration

I choose parameters to match several key moments of the U.S. economy in years between 1994 and 2020. The focus of the calibration is primarily on the distribution of productivity, the level of bank competition, and the quality of U.S. financial institutions (asset pledgeability λ).

Bank concentration ($\frac{1}{M}$) in the model economy is measured by the average HHI in the U.S. over years between 1994-2020. By definition of HHI:

$$HHI = \sum_{i=1}^{M} {s_i^d}^2 = \sum_{i=1}^{M} (\frac{1}{M})^2 = \frac{1}{M} \tag{1.20}$$

where the second equality follows that in the steady state of the symmetric equilibrium, bankers represent a market share of $1/M$ in the deposit market. Following Drechsler et al. (2017), the average deposit market HHI in the U.S. from 1994 to 2020, which amounts to 0.1342, is calculated as the weighted average of branch-level HHI, using branch deposits for weights. According to Equation (1.20), M is approximately 7.45. [6]

I have not focused on a particular distribution of productivity in the above sections. Nevertheless, the assumption that firm productivity follows a Pareto distribution has become widely accepted, building on Melitz (2003). In this chapter, I use the bounded Pareto distribution instead of the Pareto distribution to satisfy Assumption 1.2. The bounded Pareto distribution is characterized by the shape parameter γ, the maximal value z_{max}, and the minimal value z_{min}. z_{min} is normalized to 1. I calibrate z_{max} and γ to match the dispersion of productivity and markups for the US in the sample years. As illustrated in Hsieh and Klenow (2009), the difference between the 75th and 25th percentiles of TFPR is 0.53[7]. The cumulative density distribution function of $log(z)$ is $\frac{z_{min}^{-\gamma}-e^{-\gamma z}}{z_{min}^{-\gamma}-z_{max}^{-\gamma}}$ if the productivity z follows a bounded Pareto distribution[8]. γ is set to be 1.5 to keep the markup around 20% following Liu and Wang (2014).[9]

Based on the value of M, λ is chosen so that the model matches the bank capital to asset ratio similar to that of the US in years between 2001 and 2017. A higher λ corresponds to a more efficient financial market, which is further reflected in a higher bank capital to asset ratio. According to FRED, the average bank regulatory capital to risk-weighted assets for the U.S. in years between 2001 and 2017 is 13.71%. Given the value of M, the implied λ is approximately 15.

In accordance with Basel III, the parameter κ is used to generate the implied policy requirement. Basel III requires a minimum Total Capital Ratio of 8%. With the addition of the capital conservation buffer, a financial institution is required to hold at least 10.5% of risk-weighted assets in capital. As I do not include the risk-based capital constraint in

[6] M should be an integer in the model economy. I do not approximate the number of banks to 7 or 8 because the estimation of some other parameters is based on a precise calibration of bank concentration. In the comparative statics, however, Ms are set to be integers.

[7] Hsieh and Klenow (2009) distinguish between TFPQ and TFPR, where the use of the plant-specific deflator yields TFPQ and the use of the industry deflator yields TFPR. Due to the normalization of the price of consumption good, TFPQ and TFQR are equivalent by definition.

[8] Assume there is a random variable X which follows a bounded Pareto distribution with parameter L, H and γ, where γ denotes the shape parameter, L denotes the minimum, and H denotes the maximum. Define $Y = log(X)$. The cumulative distribution function (cdf.) of X is $F_X(x) = Pr(X \leq x) = \frac{L^{-\gamma}-x^{-\gamma}}{L^{-\gamma}-H^{-\gamma}}$. Then the cdf. of Y is $F_Y(x) = Pr(Y \leq x) = Pr(log(X) \leq x) = Pr(X \leq e^x) = \frac{L^{-\gamma}-e^{-\gamma x}}{L^{-\gamma}-H^{-\gamma}}$. The probability distribution function is therefore $\frac{\gamma e^{-\gamma x}}{L^{-\gamma}-H^{-\gamma}}$.

[9] Different from Liu and Wang (2014), I introduce an imperfect competition in the banking sector. In response, the markup has been raised such that the shape parameter γ does not have to be as large as in their chapter.

the benchmark model, I simply value κ at 0.08.

One period in my model corresponds to one year. Following Gali and Monacelli (2005) and Christiano et al. (2005), the discount factor β is calibrated at 0.96, which implies a riskless annual rate of about 4% in the steady state. I assume that entrepreneurs are more patient so that $s = 0.98$ (Gentry and Hubbard (2000)). The aggregate capital capacity $\overline{K}$ is normalized to 1. Table 1.2 summarizes the calibration of all the parameters.

Parameters	Values	Description
β	0.96	Risk-free interest rate*
λ	15	Bank capital to asset ratio*
M	7.45	Average HHI between 1994-2020*
z_{max}	3	Hsieh and Klenow (2009)*
z_{min}	1	Normalized to 1
γ	1.5	Markup of 20%*
s	0.98	saving rate
κ	0.08	Basel III regulations*
$\overline{K}$	1	Normalized to 1

Calibrated Parameter Values

Table 1.2: * indicates that the parameter is calculated to match moments from data

1.3.2 Equilibrium with Only Risky Loans

There is a scenario in which all entrepreneurs prefer gambling projects. As a consequence, all loans are risky. An extreme case follows that:

Corollary 1.8. *Assume $\alpha p \lessapprox 1$. All loans are risky in the equilibrium.*

The proof directly follows Equation (1.7), where z_3 converges to infinity when αp is close to 1. In contrast to a prudent project, the gambling project is more costly because the expected return is lower. Under the assumption of $\alpha p \lessapprox 1$, the gap between the expected return of gambling projects and that of prudent projects narrows, causing all borrowing entrepreneurs to gamble.

I set $p = 0.9$ and $\alpha = 1.05$. Figure 1.2 illustrates the effect of bank concentration on risk taking. As can be seen in Panel (b) of Figure 1.2, all loans are risky. Panel (a) of Figure 1.2 shows the bank capital to asset ratio, which far exceeds the minimum capital requirement when the banking sector is highly concentrated. However, the minimum bank capital requirement is binding when the the banking sector is less concentrated. As mentioned in Yi (2022), the empirical evidence indicates that, on the one hand, the bank capital ratio has far surpassed the minimum capital requirements in the U.S., and on the other hand, bank concentration is positively correlated with bank capital. The intuition is straightforward: a more concentrated banking sector will result in a lower deposit rate, as well as a lower deposit supply. On the bank's balance sheet, both deposits and equity capital are liabilities. In a more concentrated banking sector, the bank capital ratio increases due to the substitution effect between the two objects.

Panel (c) of Figure 1.2 illustrates how intensifying competition among banks leads to an increase in capital allocated to gambling projects. There is a significantly negative relationship between bank concentration and risky capital regardless of whether the capital constraint is binding. With all loans being risky, the *risk shifting mechanism* is shut down, leaving only the *"net margin mechanism"* to shape the relationship between bank concentration and risk taking. In a more concentrated banking sector, as shown in panels (e) and (f) of Figure 1.2, bankers charge a higher spread, which in turn results in an increased percentage of autarky entrepreneurs. Those autarky entrepreneurs who use their initial endowments to engage in production would choose prudent projects. Therefore, bank concentration and risks are negatively correlated.

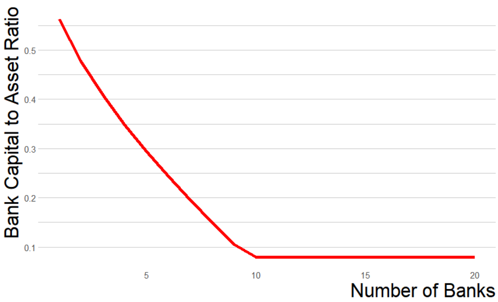
This *"net margin mechanism"* also explains how bank concentration distorts the allocation, as illustrated in panel (d) of Figure 1.2. As bank concentration climbs up, both the net margin and the fraction of autarky entrepreneurs rise. It is autarky entrepreneurs that are the most inefficient producers, to whom more capital is allocated through an extensive margin when the banking sector is more concentrated.

1.3.3 Equilibrium with Both Safe and Risky Loans

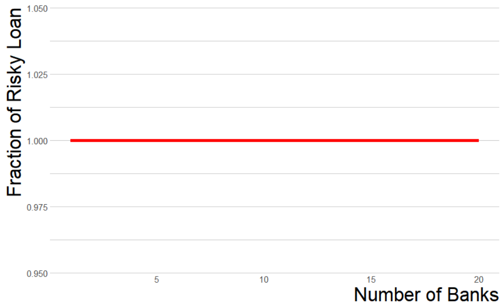
A more general case of the equilibrium is that loans could be either risky or safe. In addition to the *net margin mechanism*, the *risk shifting mechanism* as described in Proposition 1.5 will reshape the relationship between bank concentration and risks. I set $\alpha = 1.05$ and $p = 0.7$ in this section to enable gambling and prudent projects to coexist in the steady state [10]. Figure 1.3 presents how bank concentration affects the loan rate.

When considering the *risk shifting mechanism*, it is noteworthy that the correlation between bank concentration and loan rate is non-monotonic. The loan rate could rise as a result of profit maximization when the bank concentration is large. More specifically, as the banking sector becomes more concentrated, the elasticity of loan demand declines. As a consequence of the *risk shifting mechanism*, however, bankers who charge a higher

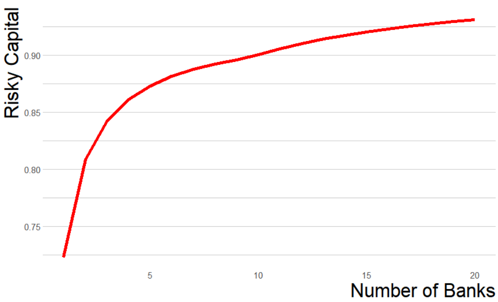
[10] Calibration is performed to match moments from data under this parameter setting.



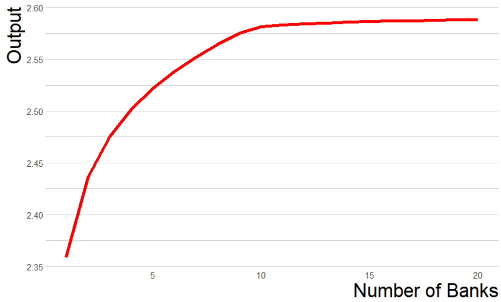
Panel (a): Bank Capital to Asset Ratio



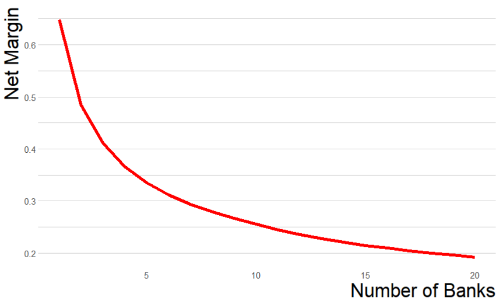
Panel (b): Fraction of Risky Loan



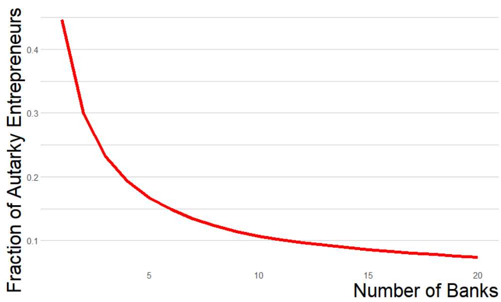
Panel (c): Risky Capital



Panel (d): Output



Panel (e): Net Margin



Panel (e): Fraction of Autarky Entrepreneurs

Comparative Statics of Model with Only Risky Loan

Figure 1.2: This plot presents the relationship between bank concentration (number of bankers) and endogenous variables: bank capital to asset ratio, fraction of risky loan, risky capital, output, net margin, and fraction of autarky entrepreneurs, when all loans are risky. I focus on the comparative statics in the steady state.

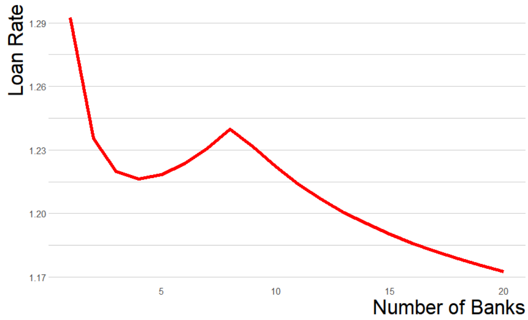


Figure 1.3: Bank Concentration and Loan Rate

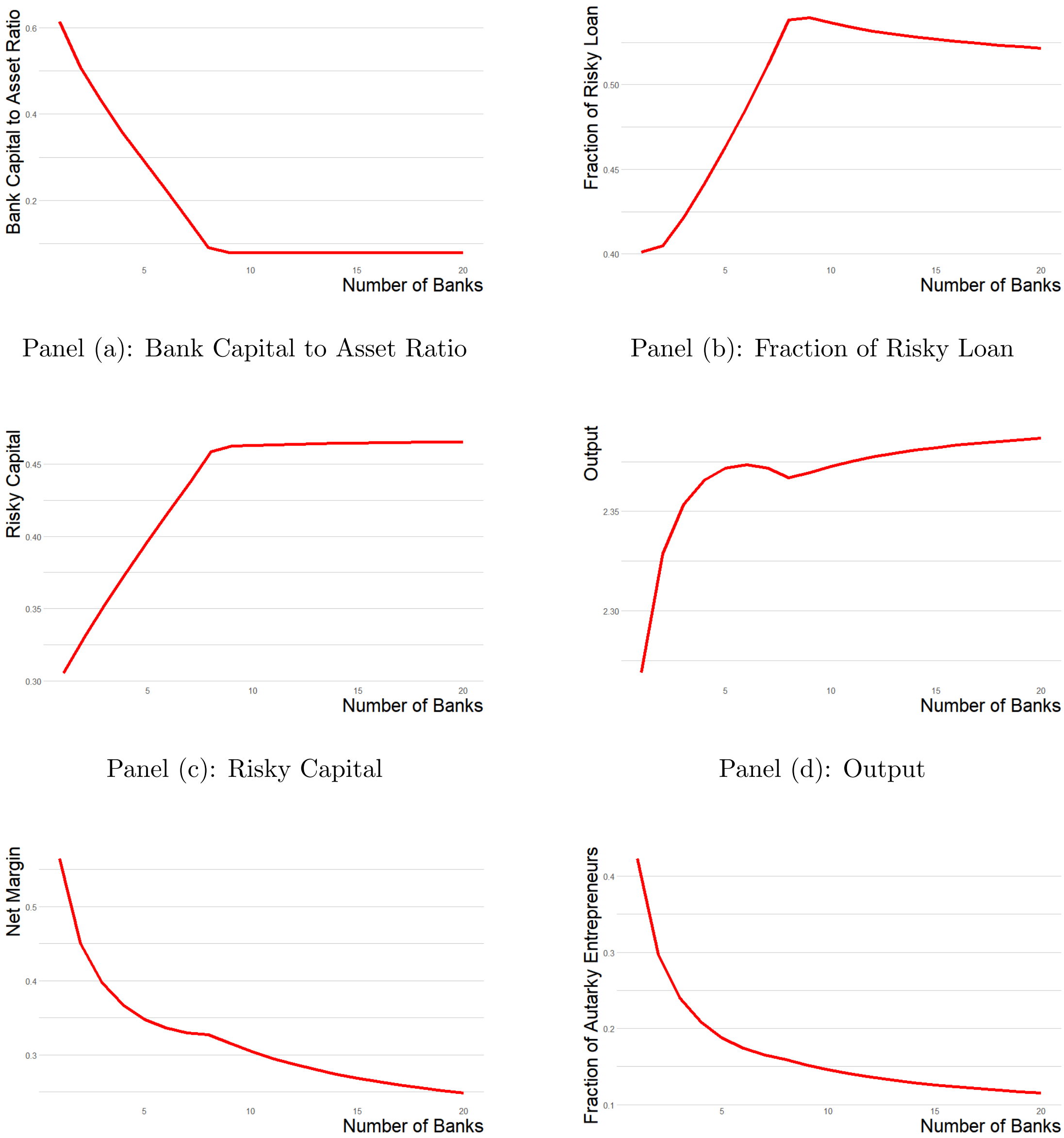
loan rate observe more gambling entrepreneurs and lower expected probability of loan repayment. In response, they will internalize the best response of entrepreneurs and will not set a rate that is too high. As shown in Figure 1.3, there is a surprisingly negative correlation between bank concentration and loan rate when there are approximately 4 to 8 banks in the economy.

However, when the banking sector is highly concentrated, there is a positive correlation between bank concentration and the loan rate. The intuition follows that, in general equilibrium, bank concentration reduces output and thus lowers capital demand. Consequently, the price of capital q drops, which pushes up the loan rate.

The kink in Figure 1.3 corresponds to the point at which banks start to accumulate excess capital, as shown in the panel (a) of Figure 1.4. Bankers with large market power are motivated to hold a capital ratio above the minimum requirement. This motive is further enhanced when considering the *risk shifting mechanism*. Those bankers with more capital may issue more loans, lowering the rate and the fraction of risky loans.

Panel (b) of Figure 1.4 shows the relationship between bank concentration and the fraction of risky loans. As the banking sector becomes more concentrated, the fraction of risky loans rises when the bank capital constraint is binding, while declines when the bank capital constraint is non-binding. When the bank capital constraint binds, bankers with a high level of market power charge high interest rates, resulting in a higher proportion of risky loans through the *risk shifting mechanism*. When bankers hold excess capital above the required minimum, large bank concentration leads to lower effective loan rates and a smaller fraction of risky loans.

It is relevant to note that the relationship between fraction of risky loan and loan rate in the general equilibrium is non-monotonic. When bank concentration is extremely high, the general equilibrium effect through the price of capital induces the loan rate to rise.



Panel (a): Bank Capital to Asset Ratio

Panel (b): Fraction of Risky Loan

Panel (c): Risky Capital

Panel (d): Output

Panel (e): Net Margin

Panel (f): Fraction of Autarky Entrepreneurs

Comparative Statics of Model with Both Safe and Risky Loan

Figure 1.4: This plot presents the relationship between bank concentration (number of bankers) and endogenous variables: bank capital to asset ratio, fraction of risky loan, risky capital, output, net margin, and fraction of autarky entrepreneurs, when loans are either risky or safe. I focus on the comparative statics in the steady state.

However, the return rate of loans in terms of consumption (numeraire) $q(1+r^b)$ is positively correlated with the fraction of risky loans v_r in the steady state. The relationship between $q(1+r^b)$ and M is shown in the appendix. In fact, increasing the price of capital also leads to a higher fraction of risky loans because doing so raises entrepreneurs' external funding cost.

Bank Concentration and Risk Taking

The relationship between bank concentration and risks (entrepreneurs' risk taking) is illustrated in panel (c) of Figure 1.4. When the bank capital constraint is non-binding, the level of risky capital is negatively correlated with bank concentration, while it is somewhat uncorrelated when the bank capital constraint is binding.

The overall level of risk is influenced by two key factors: the total loan size and the proportion of risky loans. Higher risk can arise from either a larger total loan size or an increased fraction of risky loans. When the bank capital constraint is binding, the loan rate should be higher in a more concentrated banking sector. According to the *risk shifting mechanism*, bank concentration should be positively correlated with risk because of higher proportion of risky loans, but this relationship is not evident in panel (c). In this regard, the first element in Equation (1.19)—the *net margin mechanism*— gives a compensating effect in the opposite direction: as bank concentration increases, the proportion of autarky entrepreneurs will be higher, which in turn leads to lower loan size. Autarky entrepreneurs always invest in prudent projects, which results in a decline in risky capital. Based on parameters calibrated to match U.S. moments, the magnitude of the two effects is so similar that there is almost no correlation between bank concentration and risks. As long as the capital ratio exceeds the minimum requirement, however, both mechanisms lead to a negative correlation between bank concentration and risk taking.

Bank Concentration and Output

Panel (d) of Figure 1.4 illustrates how bank concentration affects output. There is a non-monotonic relationship between the two objects when both the *net margin mechanism* and the *risk shifting mechanism* are considered.

Due to the *"net margin mechanism"*, output should have surged when bank competition is more intense. When bank concentration is high, bankers charge a wider wedge between the loan and deposit rate. In this way, autarky entrepreneurs, who are also the least efficient producers, are allocated with more resources through extensive margins. In fact, Joaquim et al. (2019) demonstrate empirically that if the lending spread falls to the world average, Brazilian output will increase by five percent. Meanwhile, higher bank concentration leads to a higher bank capital ratio in the equilibrium region where bank capital constraint is non-binding. In light of the *risk shifting mechanism*, bankers have even greater incentives to accumulate excess capital ratios, thus reducing the proportion of risky loans. Considering Assumption 1.1, gambling projects have a lower expected payoff, which mitigates the negative impact of bank concentration on output. The negative relationship between bank concentration and output is even reversed when the number of bankers is approximately 6 to 8. As the number of banks M is calibrated at 7.45 in the above section, this local optimum may be quantitatively significant.

1.4 Discussions

In this section, I provide supporting evidence based on the model predictions. In line with the model, I observe a non-monotonic relationship between bank concentration and loan rates in the U.S. Furthermore, the model characterization allows for regulations to improve efficiency and reduce risks.

1.4.1 Supporting Evidence

Using U.S. data, I document new empirical evidence regarding the non-monotonic relationship between bank concentration and loan rate. While these results do not directly address the impact of bank concentration on risks, they illustrate how the *risk shifting mechanism* operates through the loan rate.

Data Description

The analysis combines three different data sources: (i) Summary of Deposits from the Federal Deposit Insurance Corporation (FDIC), (ii) bank balance sheet items from U.S. Call Reports provided by the Federal Reserve Bank of Chicago, (iii) branch level rate data from RateWatch. In this section I discuss the main characteristics of each dataset.

Deposit Quantity The data on deposit quantities from the FDIC contains all the U.S.

bank branches at an annual frequency from 1994 to 2020. The dataset provides information on branch characteristics, ownership details, and deposit quantities at county-year level. I use the unique FDIC bank identifier to match it with other datasets.

Bank Balance Sheet The bank data is from U.S. Call Reports provided by the Federal Reserve Bank of Chicago, from March 1994 to March 2020. The data covers quarterly data on the balance sheet items of all U.S. commercial banks. I match the Call Reports to the FDIC data using the FDIC bank identifier. In the Appendix, I will show the non-monotonic correlation between bank concentration and loan rate at bank level with a local polynomial smoothing.

RateWatch RateWatch data covers monthly loan rates at the branch level. My sample is from 1994 to 2021. For loan rates, I use one of the most common loans in the sample: auto loans (72 months)[11]. Using this strategy, I am able to eliminate issues associated with observed (and unobserved) heterogeneity among loan products. I will focus on the branches that are actively involved in setting the loan rate.

Following Drechsler et al. (2017), I use HHI to measure bank concentration. I first construct a county-year level HHI, which is measured by the sum of each bank institution's squared deposit market share by county for each year (Equation (1.20)). To obtain a bank-level HHI, I calculate the weighted average HHI of all the branches under the same bank institution, using branch deposit sizes for weights.

Bank Concentration and Loan Rate Revisited

According to the model, there is a non-monotonic relationship between bank concentration and the loan rate. This section examines the empirical evidence for the non-monotonicity by conducting a fixed effect regression of loan rate on branch-level HHI. I begin by estimating the following regression:

$$LoanRate_{kt} = \sum_{i=1}^{10} \beta_i HHI_{c(k)t} * \mathbb{1}(HHI_{c(k)t} \in (\frac{i-1}{10}, \frac{i}{10}]) + \alpha_{j(k)} + \alpha_t + \alpha_{s(k)t} + \epsilon_{jt} \quad (1.21)$$

where $LoanRate_{kt}$ is the loan rate for branch k at quarter t, $\alpha_{j(k)}$ is the fixed effect associated with branch k belonging to institution j, α_t is the quarter fixed effect, $\alpha_{s(k)t}$ is

[11] In the model, the borrowers are those entrepreneurs who use external funds to produce. In this chapter, I use the auto loans because: 1. auto loan is the loan type with the most observations in the dataset; 2. households are not obviously the only type of agent who borrows money to buy car, some firms will also buy autos to promote business. In the Appendix, I will rerun the regressions using the business loan as a robustness check.

the state-time fixed effect and $HHI_{c(k)t}$ is the branch-level (county-level) HHI for branch k at quarter t. The inclusion of $\alpha_{s(k)t}$ in the regression is based on Rice and Strahan (2010), where they construct a state-by-state deregulation index.[12] I cluster the standard error at bank level. The main coefficients of interest in the regression are β_i, where i=1, 2, $\cdots$, 10. The coefficients capture the differential effect of bank concentration on loan rate within deciles of bank concentration. For example, a positive β_1 implies a positive correlation between HHI and loan rate when HHI falls within $(0, 0.1]$. The model predicts that β_i will have both positive and negative values.

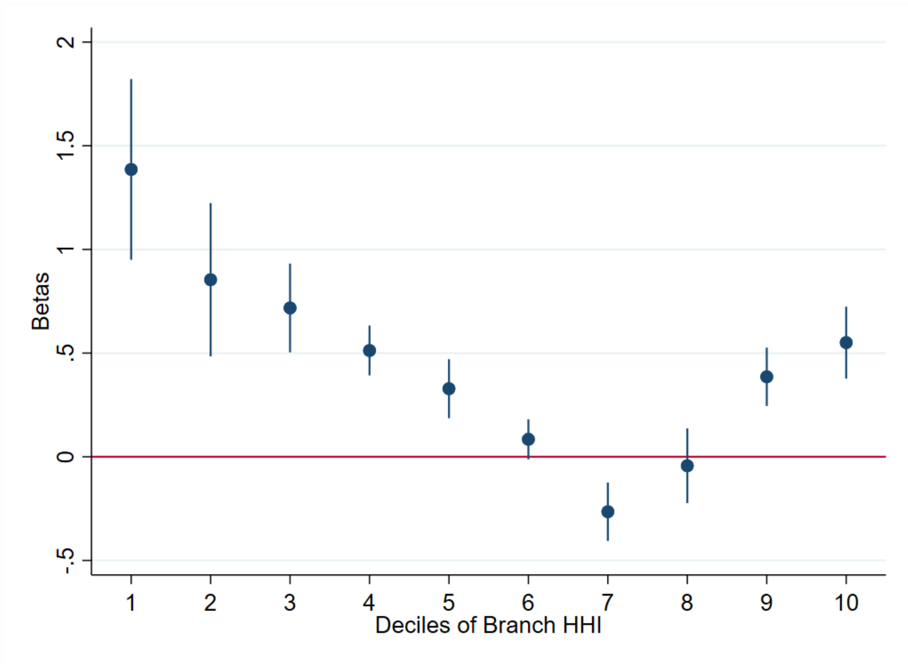
I control for bank fixed effect and time fixed effect in Figure 1.5. As illustrated in Figure 1.5, the coefficients are positive and significant when HHI lies in $(0, 0.6]$. This implies a positive correlation between bank concentration and loan rate in this region. In contrast, the coefficient becomes negative and significant at 1% level as HHI increases. More specifically, β_7 is -0.27. When HHI increases from 0.6 to 0.7, the loan rate drops by 0.027%. In accordance with the model prediction, this negative correlation can be attributed to *risk shifting mechanism* and non-binding capital constraint. Rising interest rates increase entrepreneurs' motivation to gamble. Banks internalize entrepreneurs' decisions and dislike too high a loan rate, even in a highly concentrated banking sector. β_9 and β_{10} are significantly positive. The non-monotonic relationship between HHI and loan rate is consistent with the model equilibrium. In Appendix, I show more specifications in Table 1.6.

Interaction between Bank Concentration and Bank Capital

In accordance with model predictions and Table 1.6, the relationship between bank concentration and loan rate is not monotonic. The non-monotonicity depends on whether the bank capital constraint is binding: when the bank capital constraint is binding, increasing bank concentration always results in higher loan rates, as the elasticity of loan demand falls; when banks accumulate excess capital above the minimum capital requirement, the relationship between bank concentration and loan rate is ambiguous as a result of the *risk shifting mechanism* and the general equilibrium effect. To support the model characterization, I conduct the following regression:

$$LoanRate_{kt} = \beta_1 HHI_{c(k)t} + \beta_2 HHI_{c(k)t} * Low\ Capital_{jt} + \alpha_{j(k)} + \alpha_t + \alpha_{s(k)t} + \epsilon_{jt} \quad (1.22)$$

[12] Interstate branching was not allowed in the U.S. until the Riegle-Neal Act was enacted in 1994. To mitigate the risks associated with financial institutions, the Dodd-Frank Act entered into force in 2010. However, states are allowed to use the four key provisions contained in IBBEA to restrict or increase the cost of out-of-state entry, based on which Rice and Strahan (2010) construct a bank deregulation index ranging from 0 to 4, with 4 for states with the most strict requirement for entry of out-of-state banks. By adding the state-time fixed effect, I rule out the issues of different deregulation policies across states.



Branch-level HHI and Loan Rate (Auto Loan)

Figure 1.5: This plot shows how the relationship between HHI and loan rate varies with HHI. I control for bank fixed effect and time fixed effect in this figure. The X axis represents the ordinal deciles of branch-level HHI, and the Y axis represents the coefficients of interaction between HHI and the indicator of HHI being in different deciles (β_is in Equation 1.21). The figure shows pointwise estimates and the 95 % confidence interval. When HHI is extremely low or high, the pointwise estimate is significantly positive, whereas when it lies in the 7th decile, the pointwise estimate is significantly negative. More specifications will be shown in Table 1.6.

where $LoanRate_{kt}$ is the loan rate for branch k at quarter t, $\alpha_{j(k)}$ is the fixed effect associated with branch k belonging to institution j, α_t is the quarter fixed effect, $\alpha_{s(k)t}$ is the state-time fixed effect and $HHI_{c(k)t}$ is the HHI for county $c(k)$ at quarter t. $Low\ Capital_{jt}$ is a dummy variable that indicates whether the bank-level capital ratio is below the 80th quantile. β_2 is the main coefficient of interest, which measures the heterogeneous effect of bank concentration across groups with different capital ratios. I cluster the standard error at bank level.

I show the regression results in Table 1.3, where each column controls for different fixed effects. As shown in Table 1.3, the effect of bank concentration ($HHI_{c(k)t}$) on the loan rate ($LoanRate_{kt}$) is insignificant when the capital ratio is high. Under different specifications, this result remains robust. In contrast, the coefficient on the interaction between $HHI_{c(k)}t$ and $Low\ Capital_{jt}$ is positive and significant when I control for either the state or state-time fixed effect. As shown in columns 2 and 3, β_2 is approximately 0.279 and 0.308, respectively. This indicates that there is a significantly different effect of bank concentration between groups with high and low bank capital ratios. Moreover, $\beta_1 + \beta_2$ measures the impact of bank concentration on loan rate when the bank capital ratio is low, the estimate of which is positive and significant at 5% level.

According to Table 1.3, when bank capital ratios are low (high), the effect of bank concentration on loan rate is positive (ambiguous). Observations with high capital ratios are indicative of bankers who accumulate excess capital over the minimum capital requirement in the model. Consequently, $\beta_1 + \beta_2$ being positive and significant is due to a decrease in the elasticity of loan demand when the bank capital constraint is binding. When the bank capital constraint is non-binding, the *risk shifting mechanism* and the general equilibrium effect result in a U-shaped correlation between bank concentration and loan rate, making β_1 insignificant.

1.4.2 Policy Implications

Based on the model equilibrium characterization, it is apparent that the bank capital constraint is a significant factor in determining the relationship between bank concentration, risks, and efficiency. I will discuss the policy implications of how to improve efficiency and reduce risks simultaneously in this section.

Variables	(1) OLS	(2) OLS	(3) OLS
Branch-HHI	-.00114	0.0738	0.0737
	(0.104)	(0.111)	(0.108)
Branch-HHI*Low Capital	0.242	0.279*	0.308*
	(0.152)	(0.168)	(0.170)
Constant	5.42***	5.41***	5.41***
	(0.0180)	(0.0162)	(0.0192)
Bank Fixed-effect	Yes	Yes	Yes
Quarter Fixed-effect	Yes	Yes	Yes
State Fixed-effect	No	Yes	No
State-Year Fixed-effect	No	No	Yes
R-Squared	0.775	0.781	0.791
Observations	82,065	82,065	82,065

Bank Concentration and Loan Rate in Low/High-Capital-Ratio Group (Auto Loan)

Table 1.3: This table shows the heterogeneous effect of branch-level HHI on loan rate in high/low-capital-ratio groups. The data is at the branch-quarter level and cover from January 1994 to March 2021. The standard errors are clustered at bank level. Compared to column 1, I additionally control for the state fixed effect in the second column and the state-time fixed effect in the third column. *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level.

A relevant question is whether it is sufficient to simply remove the barriers to competition. As long as the bank capital constraint is binding, scaling down the bank concentration would boost efficiency, yet has a negligible effect on risks, as the *risk shifting mechanism* and the *net margin mechanism* operate in opposite directions. When banks accumulate a capital ratio well above the minimum requirement, it is no longer sufficient to reduce concentration, since there will be a higher degree of fragility simultaneously. To enhance efficiency and lower risk, it would be prudent to reduce bank concentration and raise the minimum capital requirement at the same time. A higher level of bank capital would not only expand the region where the bank capital constraint is binding and thereby, risk is insensitive to bank competition, but it would also reduce the level of risk taken by entrepreneurs by lowering the interest rate on loans.

Reduced bank concentration contributes to lower efficiency and higher risk when the number of banks is approximately 6 to 8, as illustrated in Figure 1.4. The short-term effects of intensifying bank competition are therefore not always favorable. There is a local optimum rather than a global optimum when the number of banks reaches approximately 7. Accordingly, policymakers should be confident in reducing the obstacles to bank competition even if they observe a short-term loss of welfare. The argument is relevant because the calibrated number of banks in the U.S. using HHI is 7.45, which means a small deviation would have a negative but temporary impact.

1.4.3 Aggregate Uncertainty

In this section, I consider the model where the success of the gambling project depends on the realization of an aggregate shock. In the context of aggregate uncertainty, analyzing the steady state becomes challenging within the dynamic framework. The occurrence of adverse states, where gambling projects fail, introduces positive default probabilities for both entrepreneurs and bankers. Consequently, the family of entrepreneurs may confront a lack of endowment to overcome financial frictions in subsequent periods, while bankers may cease to exist. Furthermore, the inclusion of new entrants among entrepreneurs and bankers would alter the underlying intuition regarding the relationship between bank capital and the concentration-risk dynamics.

To shed light on the underlying mechanisms in the presence of aggregate uncertainty, I conduct an analysis within a static framework. Specifically, I assume that the bank capital is exogenously determined at a predetermined value of c. Within this setting, all entrepreneurs are endowed with a fixed amount of capital, denoted as K, and the price of capital remains constant at 1. It is important to note that all other aspects of the model remain unchanged in this static setup. Notably, all the propositions and lemmas remain valid within Section 1.2 of this chapter. It is pertinent to acknowledge that bank default occurs when:

$$(1 + r^b)p^e Q^L (1 + r^d)Q^D < 0 \tag{1.23}$$

The *net margin mechanism* pertains to a scenario where an increase in bank concentration results in a reduced probability of bank default. This is attributed to two key factors: firstly, enhanced profitability for the bank due to a rise in loan rates and a decline in deposit rates; secondly, a greater reliance on capital as a funding source denoted by the increase in the ratio of bank capital to aggregate deposits ($\frac{c}{Q^D}$). Consequently, the *net margin mechanism*, as observed in the baseline model, continues to indicate a negative relationship between bank concentration and risks.

Conversely, the *risk shifting mechanism* manifests when an increase in bank concentration leads to higher loan rates and a decrease in the expected probability of loan repayment (p^e). Consequently, this results in an elevated probability of bank default.

Within this static framework, it is important to note that the *risk shifting mechanism* and *net margin mechanism* consistently operate in opposite directions regarding the relationship between bank concentration and the probability of bank default. This corresponds to

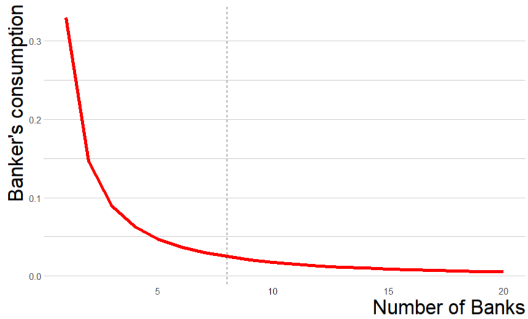


Figure 1.6: Number of Banker and Bankers' Consumption

the findings of the baseline model in instances where the bank capital constraint is binding. Given the absence of an interaction between bank concentration and bank capital within this static setting, the direction of the risk shifting mechanism remains unchanged.

1.4.4 Exogenous Variation of Bank Concentration

In the baseline model, I include an exogenous number of bankers (M) to capture the impact of bank concentration on risk taking. Based on the equilibrium characterization, the relationship between bank concentration ($\frac{1}{M}$) and risk taking (rc) depends on whether the bank capital constraint is binding. Nevertheless, the number of banks in the real world is determined endogenously by other market conditions, such as switching costs, entry costs, etc.

In this section, I endogenize the number of bankers M by allowing the entry cost to vary. When bankers decide to enter the banking sector, they are expected to pay a constant amount of τ. In the steady state of a symmetric equilibrium, the free entry condition implies:

$$\frac{1}{1-\beta}c_{it}^{b} = \tau,$$

which equalizes the lifetime utility derived from consumption with the entry cost. It would be useful to examine the relationship between bankers' consumption and the implied number of bankers so that I understand how entry costs affect bank concentration. As illustrated in Figure 1.6, there is a negative correlation between bankers' consumption and the number of bankers. Consequently, rising entry costs cause few bankers to participate in the banking sector, as a highly concentrated sector assures a higher benefit to entry.

In light of the monotonic correlation between bankers' consumption and the number of bankers, the extension to generate an endogenous M is completely equivalent to the baseline model. Accordingly, the number of bankers M in the baseline model can be interpreted as an exogenous variation in bank concentration accompanied by varying entry costs.

1.5 Conclusion

Throughout this chapter, I develop a tractable dynamic model to investigate how bank capital affects the relationship between bank concentration and risk taking. Accumulating excess bank capital when the banking sector is highly concentrated not only enables banks to maximize their profits, but also minimizes the effect of the *risk shifting mechanism.* As a result of the *risk shifting mechanism* together with the *net margin mechanism*, there is a kinked relationship between bank concentration and risk taking, which depends on whether the minimum capital requirement is binding. The model suggests a negative correlation between bank concentration and risk when the capital ratio exceeds the minimum requirement; otherwise, bank concentration has an ambiguous but quantitatively negligible impact on risk. This chapter raises concerns about future empirical studies that examine the effects of bank concentration without considering bank capital levels.

The purpose of this chapter is to explore idiosyncratic risk, which is often believed to be associated with financial stability. However, for a more concrete analysis, it would be valuable and necessary to explicitly model aggregate risk, in which case I could analyze the financial crisis, financial distress, etc. To keep the model tractable, all terms are real. This model with such rich heterogeneity would be useful for studying monetary policy by introducing price rigidity. I leave all these extensions for future research.

1.6 Appendix

1.6.1 Proofs

Proof of Proposition 1.3. All the entrepreneurs are risk neutral and maximize their expected consumption today. Since the saving rate of entrepreneurs is exogenous given, consumption follows:

$$c_t = s_t \Pi_t$$

where Π_t is the net return of the generation t. The functional form of Π_t is different in the following 3 cases:

Case 1. If the entrepreneurs choose to deposit part of their wealth ($k_t \leq a_t$), then

$$\Pi_t = z_t k_t + q_t(r_t^d + 1)(a_t - k_t) = [z_t - q_t(r_t^d + 1)]k_t + q_t(r_t^d + 1)a_t \tag{1.24}$$

where q_t is the price of capital, r_t^d is the deposit rate and k_t is the capital that is used in production. Note that entrepreneurs who do not borrow will not invest in gambling projects. The reason for this is that they prefer projects with a higher expected return.

The above equation implies that k_t equals to 0 or a_t, which depends on whether the productivity is above $q_t(r_t^d + 1)$.

Case 2. Suppose that the entrepreneur becomes a borrower and chooses the prudent project. Denote her leverage ratio as θ with $\theta \leq \lambda$, the net profit is then:

$$\Pi_t = z_t \theta a_t - q_t(r_t^b + 1)(\theta - 1)a_t = [z_t - q_t(r_t^b + 1)]\theta a_t + q_t(r_t^b + 1)a_t \tag{1.25}$$

where r_t^b is the loan rate. Following the above equation, the value of θ equals to 1 or λ, which depends on whether the productivity is above $q_t(r_t^b + 1)$.

Case 3. Suppose that the entrepreneur becomes a borrower while invests in the gambling project. Denote her leverage ratio as θ with $\theta \leq \lambda$, the net profit is then:

$$\Pi_t = p\{\alpha z_t \theta a_t - q_t(r_t^b + 1)(\theta - 1)a_t\} = p\{[\alpha z_t - q_t(r_t^b + 1)]\theta a_t + q_t(r_t^b + 1)a_t\} \tag{1.26}$$

Following the above equation, the value of θ equals to 1 or λ, which depends on whether the productivity is above $\frac{q_t(r_t^b+1)}{\alpha}$. Since α is greater than 1, there is a region of productivity in which borrowing entrepreneurs might want to start a gambling project rather than a prudent one.

The remaining calculation is to identify the border of each case. If borrowing and gambling exists in the equilibrium, the benefit of doing so should dominate that of staying autarky, as well as borrowing and investing in the prudent project. The condition is derived in Equation (1.6) and (1.7) that:

$$\frac{(\lambda-1)p}{\lambda\alpha p-1}q(1+r^b) = z_2 < z < z_3 = \frac{(\lambda-1)(1-p)}{\lambda(1-\alpha p)}q(1+r^b) \tag{1.27}$$

Further, $\frac{(\lambda-1)p}{\lambda\alpha p-1} > \frac{1}{\alpha}$ following Assumption 1.1. Therefore, under the condition implied by Equation (1.27), entrepreneurs will borrow up to the borrowing limits λ and invest in the gambling project.

By Assumption 1.2, $\frac{(\lambda-1)(1-p)}{\lambda(1-\alpha p)} > 1$ and entrepreneurs borrow and invest in the prudent project if and only if $z > z_3$. In an extreme when $z_3 > z_{max}$, there are no borrowing entrepreneurs who stay prudent in the equilibrium.

When $q(1+r^d) < z < z_2$, $k = a$, which means that entrepreneurs will use their internal finance to produce. When $z < q(1+r^d)$, $k = 0$, so that the entrepreneurs deposit all their money in banks. □

Proof of Lemma 1.4. Equations (1.8) and (1.9) are directly obtained from Proposition 1.3, given that borrowing entrepreneurs borrow up to the borrowing limit and lending entrepreneurs deposit all their capital in banks.

For the lending entrepreneurs, their net return becomes:

$$\Pi_t = (r_t^d + 1)q_t a_t$$

For the borrowing entrepreneurs who invest in the prudent project, their net return becomes:

$$\Pi_t = \lambda(z_t - (r_t^b + 1)q_t)a_t + (r_t^b + 1)q_t a_t$$

For the borrowing entrepreneurs who gamble, their net return becomes:

$$\Pi_t = p\{\lambda(\alpha z_t - (r_t^b + 1)q_t)a_t + (r_t^b + 1)q_t a_t\}$$

For the autarky entrepreneurs, their net return becomes:

$$\Pi_t = z_t a_t$$

Given the constant saving rate, I have:

$$\begin{aligned} q_t a_{t+1} =& \beta\{\int_{z_{min}}^{z_{1t}} q_t(1+r_t^d)dG(z_t) + \int_{\overline{z_{3t}}}^{z_{max}} \lambda[(z_t - q_t(1+r_t^b)] + q_t(r_t^b+1))dG(z_t) \\ &+ \int_{z_{1t}}^{z_{2t}} z_t dG(z_t) + p\int_{z_{2t}}^{\overline{z_{3t}}} \alpha\lambda z_t - (\lambda-1)q_t(r_t^b+1)dG(z_t)\}a_t \end{aligned}$$

by simply aggregating all the entrepreneurs of different productivities. □

Proof of Proposition 1.5. The Bellman equation for the banker i is:

$$V(N_{it}) = \max_{\{c_{it}^b, Q_{it}^L, Q_{it}^D\}} \{c_{it}^b + \beta V(N_{it+1})\}$$

subject to the balance sheet identity (1.2), the budget constraint (1.3) and the minimum capital requirement (1.5). The Lagrangian function for banker i becomes:

$$L_{it} = q_t\{(1+r_t^b)p_t^e Q_{it}^L - (1+r_t^d)Q_{it}^D\} - q_t N_{it+1} + \mu_{it}(Q_{it}^D + N_{it} - Q_{it}^L) + \chi_{it}(N_{it} - \kappa Q_{it}^L) \quad (1.28)$$

by substituting the budget constraint into the utility function, where μ_{it} is the multiplier of the bank's balance sheet identity. χ_{it} is the multiplier of the bank capital constraint. Deriving the first order condition, I obtain Equations (1.11) and (1.12).

By definition, $v_r = \frac{G(z_3)-G(z_2)}{1-G(z_2)}$. I denote $\frac{(\lambda-1)p}{\lambda\alpha p-1}q = a_2$ and $\frac{(\lambda-1)(1-p)}{\lambda(1-\alpha p)}q = a_3$. Therefore:

$$\frac{\partial v_r}{\partial r^b} = \frac{[g(z_3)a_3 - g(z_2)a_2](1-G(z_2)) + (G(z_3)-G(z_2))g(z_2)a_2}{(1-G(z_2))^2} \quad (1.29)$$

The second element in the numerator is equivalent to $\{-[1-G(z_3)] + [1-G(z_2)]\}g(z_2)a_2$, so that Equation (1.29) becomes:

$$\begin{aligned} \frac{\partial v_r}{\partial r^b} &= \frac{[g(z_3)a_3 - g(z_2)a_2](1-G(z_2)) + \{-[1-G(z_3)] + [1-G(z_2)]\}g(z_2)a_2}{(1-G(z_2))^2} \\ &= \frac{g(z_3)a_3(1-G(z_2)) - (1-G(z_3))g(z_2)a_2}{(1-G(z_2))^2} \\ &= \frac{1}{(1+r^b)(1-G(z_2))^2}(g(z_3)z_3(1-G(z_2)) - (1-G(z_3))g(z_2)z_2) \\ &= \frac{1}{g(z_3)g(z_2)z_3z_2(1+r^b)(1-G(z_2))^2}(\frac{(1-G(z_2))}{g(z_2)z_2} - \frac{(1-G(z_3))}{g(z_3)z_3}) \end{aligned}$$

Since $\frac{zg(z)}{(1-G(z))}$ is increasing, $\frac{\partial v_r}{\partial r^b} \geq 0$. Further, p_e is a decreasing function of v_r so that $\frac{\partial p_e}{\partial r^b} \leq 0$.

□

1.6.2 Robustness Checks with Other Loan Type

In this section, I will rerun regressions that are similar to Equation 1.21 and Equation 1.22 with secured business loans in the RateWatch dataset. The total number of observations for secured business loans is 17,282, which is substantially less than the total number of observations for auto loans. Due to the limited data size, I run the following regression

$$LoanRate_{kt} = \sum_{i=1}^{5} \beta_i HHI_{c(k)t} * \mathbb{1}(HHI_{c(k)t} \in (\frac{i-1}{5}, \frac{i}{5}]) + \alpha_{j(k)} + \alpha_t + \alpha_{s(k)t} + \epsilon_{jt}, \quad (1.30)$$

where I divide the entire sample into five equal parts and include the interaction terms between HHI and quintile indicators in the regression. Table 1.4 illustrates that bank concentration has a positive and significant effect on the loan rate when branch-level HHI falls into the second or fifth quintile. Conversely, in other quintiles, there is no significant association between bank concentration and the loan rate.

Based on the model predictions, the effect of bank concentration on the loan rate is more likely to be significantly positive either when bank concentration is low or high. This model explains the positive correlation by considering the channel of the elasticity of loan demand as well as the general equilibrium effect of capital price. Due to the *risk shifting mechanism* in the model, however, the correlation between bank concentration and loan rate should be negative when the bank concentration is in between. The reason for not obtaining negative estimates in Table 1.4 might be that the dataset contains too much noise. There is a significant dispersion in the estimate due to the limited number of business loans. The correlation between bank concentration and the loan rate may be significantly negative if the quality of business loans is as good as that of auto loans.

To explain the mechanisms behind the non-monotonicity between bank concentration and

Variables	(1) OLS	(2) OLS	(3) OLS
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0, 0.2]$)	-0.0142 (0.649)	-0.291 (0.698)	-0.256 (0.669)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.2, 0.4]$)	0.743* (0.408)	0.657 (0.470)	0.734* (0.430)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.4, 0.6]$)	0.216 (0.361)	0.173 (0.373)	0.121 (0.433)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.6, 0.8]$)	0.250 (0.181)	0.0852 (0.295)	-0.0002 (0.391)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.8, 1]$)	1.10*** (0.363)	1.12*** (0.399)	0.921** (0.424)
Constant	7.10*** (0.0560)	7.13*** (0.0626)	7.13*** (0.0595)
Bank Fixed-effect	Yes	Yes	Yes
Quarter Fixed-effect	Yes	Yes	Yes
State Fixed-effect	No	Yes	No
State-time Fixed-effect	No	No	Yes
R-Squared	0.637	0.650	0.672
Observations	17,282	17,282	17,282

Bank Concentration and Loan Rate (Business Loan)

Table 1.4: This table shows the relationship between branch-level HHI and loan rate (Secured Business Loan) within different quintiles of HHI. The data is at the branch-quarter level and cover from January 1994 to March 2021. Rows 1-5 show the coefficients on the interaction term between HHI and the indicator of HHI within different quintiles. The 5 coefficients reflect the heterogeneous effect of HHI on the loan rate within different quintiles. The standard errors are clustered at bank level. Compared to column 1, I additionally control for the state fixed effect in the second column and the state-time fixed effect in the third column. *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level.

Variables	(1) OLS	(2) OLS	(3) OLS
Branch-HHI*High Capital	0.700* (0.373)	0.635* (0.369)	0.609 (0.401)
Branch-HHI*Low Capital	0.539** (0.216)	0.510** (0.243)	0.455* (0.251)
Constant	7.07*** (0.0272)	7.07*** (0.0297)	7.08*** (0.0305)
Bank Fixed-effect	Yes	Yes	Yes
Quarter Fixed-effect	Yes	Yes	Yes
State Fixed-effect	No	Yes	No
State-Year Fixed-effect	No	No	Yes
R-Squared	0.639	0.652	0.676
Observations	16,714	16,714	16,698

Bank Concentration and Loan Rate in Low/High-Capital-Ratio Group (Business Loan)

Table 1.5: This table shows the heterogeneous effect of branch-level HHI on loan rate in high/low-capital-ratio groups. The data is at the branch-quarter level and cover from January 1994 to March 2021. The standard errors are clustered at bank level. Compared to column 1, I additionally control for the state fixed effect in the second column and the state-time fixed effect in the third column. *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level.

loan rate, I run the following regression:

$$LoanRate_{kt} = \beta_1 HHI_{c(k)t} * High\ Capital_{jt} + \beta_2 HHI_{c(k)t} * Low\ Capital_{jt} + \alpha_{j(k)} + \alpha_t + \alpha_{s(k)t} + \epsilon_{jt}, \tag{1.31}$$

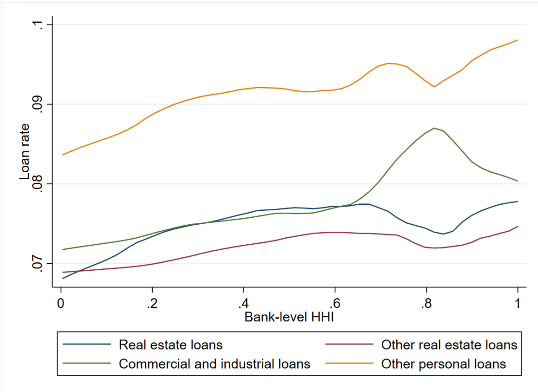
which is similar to Equation 1.22. Nevertheless, β_2 in Equation 1.31 represents the effect of branch-level HHI on the loan rate when the bank capital ratio is low. The correlation between bank concentration and loan rate is more significant when the capital ratio is low, as shown in Table 1.5. It is consistent with the model predictions and the results presented in Table 1.5.

1.6.3 Evidence at Bank-level

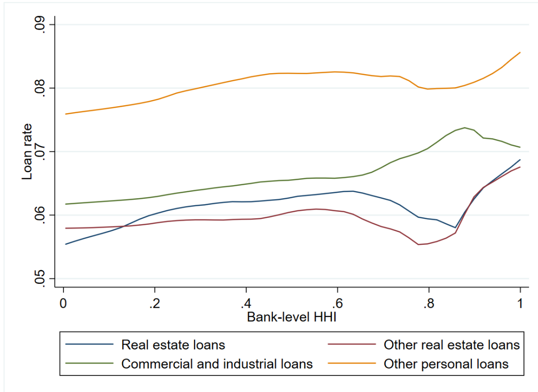
Using the FDIC and the Call Reports data, I examine the relationship between HHI and loan rate at the bank level. I compute the branch-level HHI as the sum of the squared deposit market share of each bank institution by county for each year. To obtain the bank-level HHI, I calculate the weighted average branch-level HHI of all branches belonging to the same bank institution, using branch deposits as weights. I calculate the loan rate by dividing the interest income over the loan size. What I do is running a local polynomial smoothing, and visualizing the non-linear correlation between the two objects in Figure 1.7. There are four lines in each sub-figure, where the yellow line represents other personal loans; the green line represents commercial and industrial loans; the blue line represents the real estate loans and the purple lines represents other real estate loans. As illustrated in Figure 1.8, these four loan types accounted for more than 80 percent of the total loan size.

The four sub-figures capture the relationship between bank-level HHI and loan rate in years 2008, 2012, 2016, 2020, where I partially control for the time fixed effect. It is observed from the figure that the loan rate for personal loans is higher than for other loan types. Moreover, the correlation between bank-level HHI and loan rate is non-monotonic. When the bank concentration is large, there is a region where the correlation is negative. The model prediction and branch-level evidence support this non-monotocity. The intuition follows the *risk shifting mechanism* that banks internalize the best response of entrepreneurs and prefer not to raise too high a loan rate.

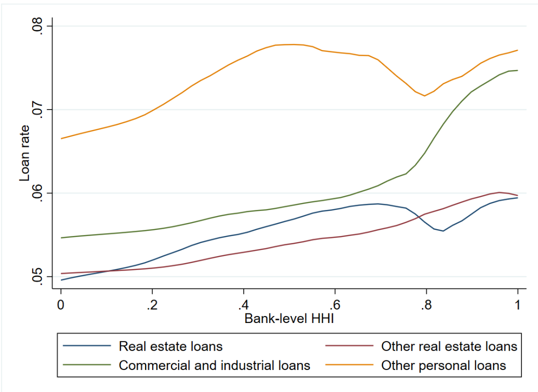
1.6.4 Additional Tables and Figures



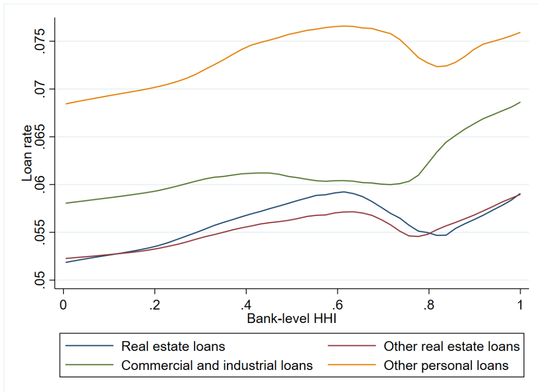
Panel (a): Year 2008



Panel (b): Year 2012



Panel (c): Year 2016



Panel (d): Year 2020

Bank Concentration and Loan Rate at Bank Level

Figure 1.7: This plot presents the correlation between bank concentration and loan rate. There are four lines in the graph, where the yellow line represents other personal loans; the green line represents commercial and industrial loans; the blue line represents the real estate loans, and the purple line represents other real estate loans.

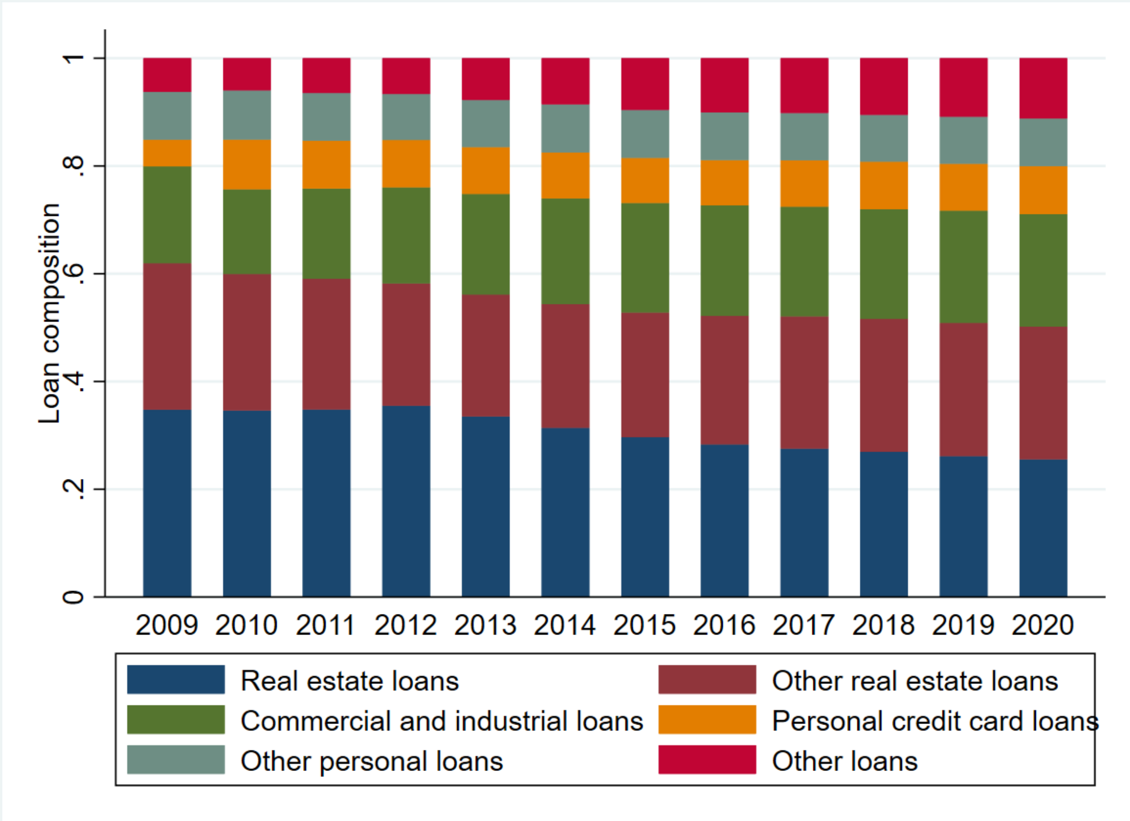
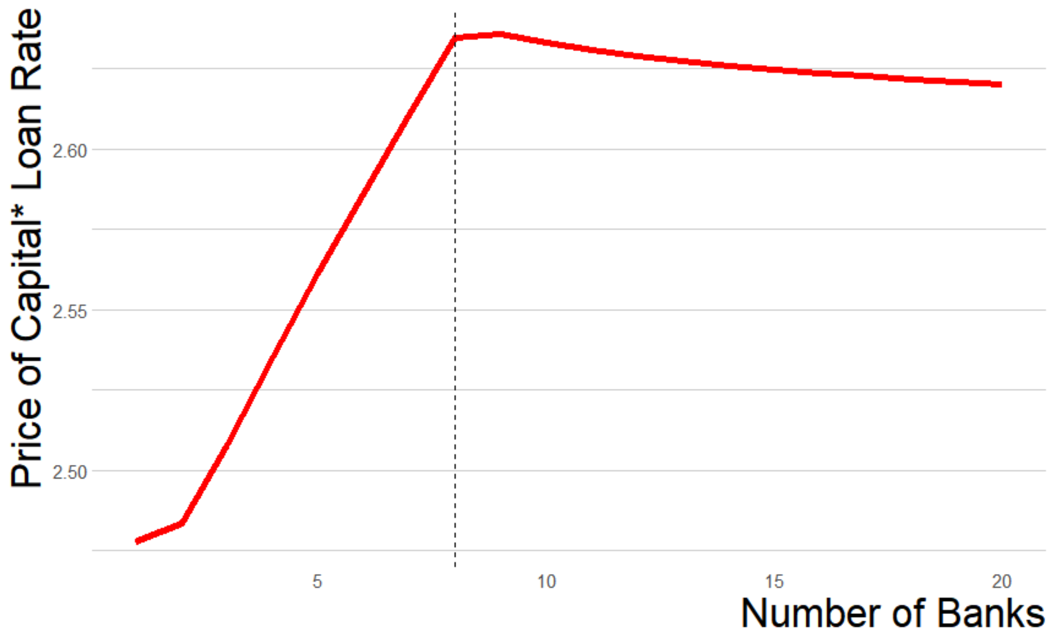


Figure 1.8: Loan Composition in the U.S.



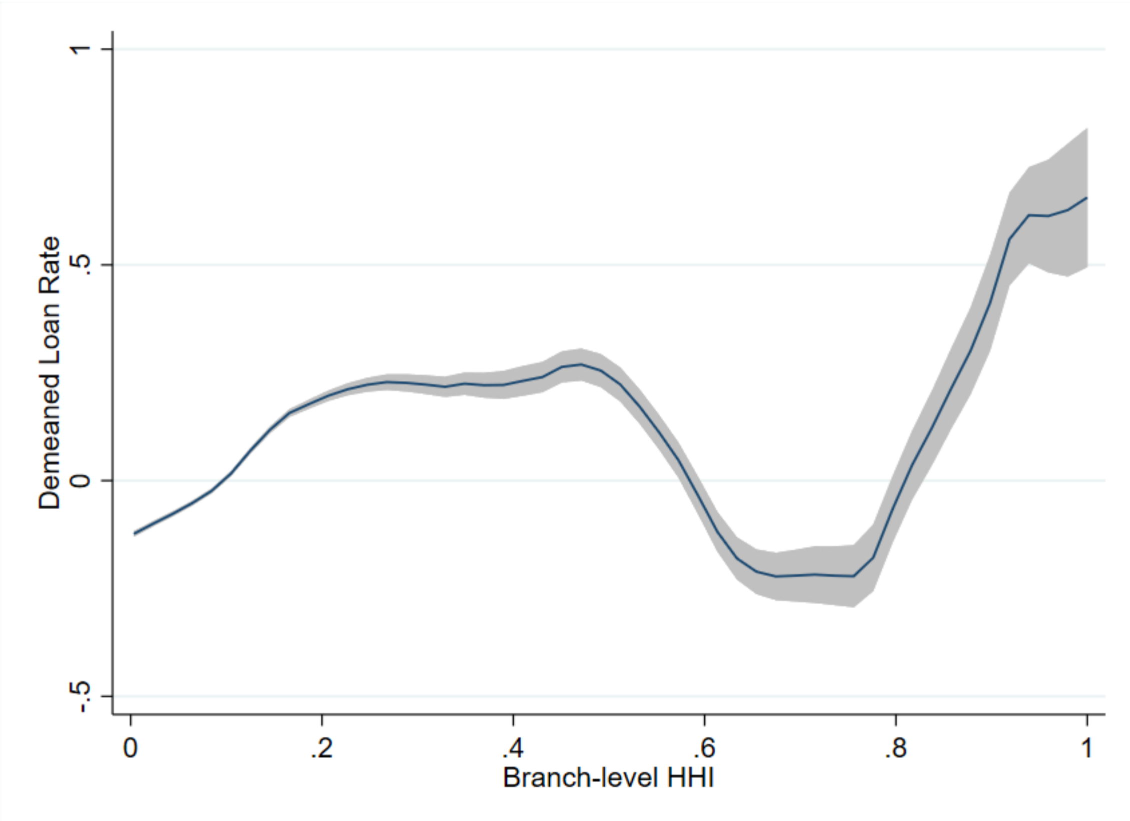
Price of Capital Times Loan Rate V.S. Number of Banks

Figure 1.9: This figure shows the price of capital q times loan rate $1 + r^b$ in the steady state. $q(1 + r^b)$ shapes exactly the same as the fraction of risky loan.

Variables	(1) OLS	(2) OLS	(3) OLS
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0, 0.1]$)	1.386***	0.573***	0.569***
	(0.222)	(0.153)	(0.178)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.1, 0.2]$)	0.854***	0.494***	0.491***
	(0.188)	(0.114)	(0.129)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.2, 0.3]$)	0.718***	0.448***	0.469***
	(0.109)	(0.0705)	(0.0886)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.3, 0.4]$)	0.513***	0.299***	0.279***
	(0.0613)	(0.0900)	(0.0665)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.4, 0.5]$)	0.328***	0.231***	0.200**
	(0.0726)	(0.0867)	(0.0848)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.5, 0.6]$)	0.0843*	0.0171	0.00593
	(0.0493)	(0.0519)	(0.0543)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.6, 0.7]$)	-0.265***	-0.249***	-0.228***
	(0.0718)	(0.0647)	(0.0563)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.7, 0.8]$)	-0.0433	-0.0207	-0.0768
	(0.0918)	(0.106)	(0.112)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.8, 0.9]$)	0.386***	0.291***	0.438***
	(.0717)	(0.0563)	(0.0561)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.9, 1]$)	0.551***	0.435***	0.440***
	(0.088)	(0.0610)	(0.121)
Constant	4.75***	4.80***	4.80***
	(0.0138)	(0.00755)	(0.00986)
Bank Fixed-effect	Yes	Yes	Yes
Quarter Fixed-effect	Yes	Yes	Yes
State Fixed-effect	No	Yes	No
State-time Fixed-effect	No	No	Yes
R-Squared	0.772	0.778	0.783
Observations	166,864	166,864	166,864

Bank Concentration and Loan Rate (Auto Loan)

Table 1.6: This table shows the relationship between branch-level HHI and loan rate (Auto 6 years) within different deciles of HHI. The data is at the branch-quarter level and cover from January 1994 to March 2021. Rows 1-10 show the coefficients on the interaction term between HHI and the indicator of HHI within different deciles. The 10 coefficients reflect the heterogeneous effect of HHI on the loan rate within different deciles. From top to bottom, the coefficients are positive, negative, and then positive again, which indicates a non-monotonic relationship between bank concentration and the loan rate. The standard errors are clustered at bank level. Compared to column 1, I additionally control for the state fixed effect in the second column and the state-time fixed effect in the third column. *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level.



Local Polynomial Smoothing between Bank Concentration and Loan Rate

Figure 1.10: This figure shows the non-monotonic relationship between branch-level HHI and loan rate (Auto 72 loan). The data is at the branch-quarter level and cover from January 1994 to March 2021. The loan rate is demeaned at quarter level. A local polynomial smoothing is conducted between the demeaned loan rate and HHI. The shaded area represents the 95% confidence interval.

Bibliography

Agoraki, M.-E. K., Delis, M. D., and Pasiouras, F. (2011). Regulations, competition and bank risk-taking in transition countries. *Journal of Financial Stability*, 7(1):38–48.

Alfon, I., Argimon, I., and Bascuñana-Ambrós, P. (2004). *What determines how much capital is held by UK banks and building societies?* Financial Services Authority London.

Angeletos, G.-M. (2007). Uninsured idiosyncratic investment risk and aggregate saving. *Review of Economic dynamics*, 10(1):1–30.

Beck, T., De Jonghe, O., and Schepens, G. (2013). Bank competition and stability: Cross-country heterogeneity. *Journal of financial Intermediation*, 22(2):218–244.

Beck, T., Demirguc-Kunt, A., and Levine, R. (2003). Bank concentration and crises.

Beck, T., Demirgüç-Kunt, A., and Levine, R. (2006). Bank concentration, competition, and crises: First results. *Journal of Banking & Finance*, 30(5):1581–1603.

Black, S. E. and Strahan, P. E. (2002). Entrepreneurship and bank credit availability. *The Journal of Finance*, 57(6):2807–2833.

Boyd, J. H. and De Nicolo, G. (2005). The theory of bank risk taking and competition revisited. *The Journal of Finance*, 60(3):1329–1343.

Brunnermeier, M. K. and Koby, Y. (2018). The reversal interest rate. Technical report, National Bureau of Economic Research.

Burdett, K. and Judd, K. L. (1983). Equilibrium price dispersion. *Econometrica*, pages 955–969.

Carlson, M., Correia, S., and Luck, S. (2022). The effects of banking competition on growth and financial stability: Evidence from the bational banking era. *Journal of Political Economy*, 130(2):462–520.

Carlson, M. and Mitchener, K. J. (2009). Branch banking as a device for discipline: Competition and bank survivorship during the great depression. *Journal of Political Economy*, 117(2):165–210.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, 113(1):1–45.

Corbae, D., D'Erasmo, P., et al. (2021). Capital buffers in a quantitative model of banking industry dynamics. *Econometrica*, 8.

Corbae, D. and Levine, R. (2018). Competition, stability, and efficiency in financial markets. *Unpublished manuscript.*

Craig, B. R. and Dinger, V. (2013). Deposit market competition, wholesale funding, and bank risk. *Journal of Banking & Finance*, 37(9):3605–3622.

De Nicolò, G., Bartholomew, P., Zaman, J., and Zephirin, M. (2004). Bank consolidation, internationalization, and conglomeration: Trends and implications for financial risk. *Internationalization, and Conglomeration: Trends and Implications for Financial Risk.*

Diez, M. F., Leigh, M. D., and Tambunlertchai, S. (2018). *Global market power and its macroeconomic implications.* International Monetary Fund.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *The American Economic Review*, 67(3):297–308.

Drechsler, I., Savov, A., and Schnabl, P. (2017). The Deposits Channel of Monetary Policy*. *The Quarterly Journal of Economics*, 132(4):1819–1876.

Flannery, M. J. and Rangan, K. P. (2008). What caused the bank capital build-up of the 1990s? *Review of finance*, 12(2):391–429.

Gali, J. and Monacelli, T. (2005). Monetary policy and exchange rate volatility in a small open economy. *The Review of Economic Studies*, 72(3):707–734.

Gentry, W. M. and Hubbard, R. G. (2000). Entrepreneurship and household saving.

Head, A. C., Kam, T., Ng, S., and Pan, G. (2022). Money, credit and imperfect competition among banks. *Working Paper.*

Hellmann, T. F., Murdock, K. C., and Stiglitz, J. E. (2000). Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? *American Economic Review*, 90(1):147–165.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448.

Jayaratne, J. and Strahan, P. E. (1996). The finance-growth nexus: Evidence from bank branch deregulation. *The Quarterly Journal of Economics*, 111(3):639–670.

Jiang, L., Levine, R., and Lin, C. (2017). Does competition affect bank risk? Technical report, National Bureau of Economic Research.

Joaquim, G., van Doornik, B. F. N., Ornelas, J. R., et al. (2019). *Bank competition, cost of credit and economic activity: evidence from Brazil.* Banco Central do Brasil.

Kiyotaki, N. and Moore, J. (2019). Liquidity, business cycles, and monetary policy. *Journal of Political Economy*, 127(6):2926–2966.

Lagos, R. and Zhang, S. (2022). The limits of onetary economics: On money as a constraint on market power. *Econometrica*, 90(3):1177–1204.

Li, J. (2019). Imperfect banking competition and financial stability. Technical report, Working Paper.

Liu, Z. and Wang, P. (2014). Credit constraints and self-fulfilling business cycles. *American Economic Journal: Macroeconomics*, 6(1):32–69.

Martinez-Miera, D. and Repullo, R. (2010). Does competition reduce the risk of bank failure? *The Review of Financial Studies*, 23(10):3638–3664.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Moll, B. (2014). Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, 104(10):3186–3221.

Repullo, R. (2004). Capital requirements, market power, and risk-taking in banking. *Journal of financial Intermediation*, 13(2):156–182.

Rice, T. and Strahan, P. E. (2010). Does credit competition affect small-firm finance? *The Journal of Finance*, 65(3):861–889.

Tabak, B. M., Fazio, D. M., and Cajueiro, D. O. (2012). The relationship between banking market competition and risk-taking: Do size and capitalization matter? *Journal of Banking & Finance*, 36(12):3366–3381.

Van Hoose, D. et al. (2010). *the Industrial Organization of Banking.* Springer.

Yi, Y. (2022). Bank concentration, bank capital and misallocation. *Working Paper.*

Chapter 2

Bank Concentration, Bank Capital, and Misallocation

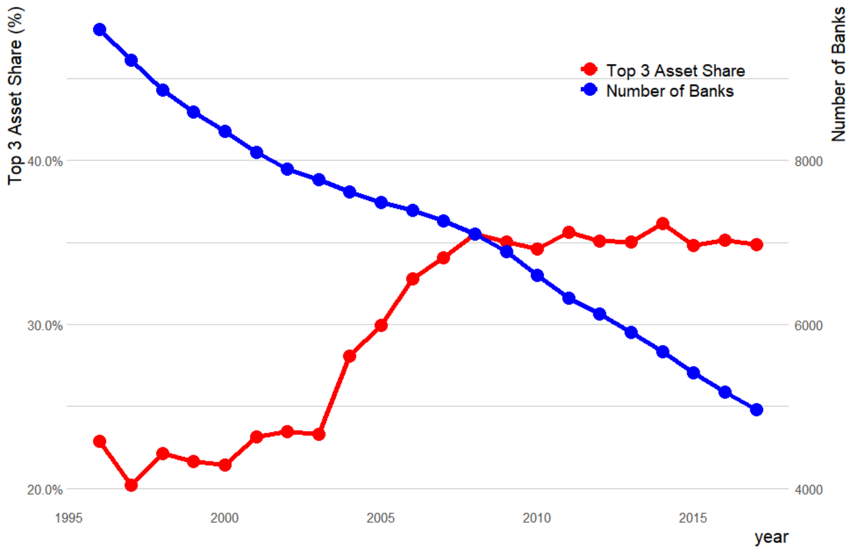
2.1 Introduction

Chapter 1 examines the relationship between bank concentration and entrepreneurs' risk taking. Building upon the theoretical foundations laid in Chapter 1, Chapter 2 streamlines and refines the model presented in Chapter 1, aiming to enhance its tractability. Additionally, Chapter 2 concentrates on the correlation between bank concentration and bank capital, and its implication on efficient capital allocation.

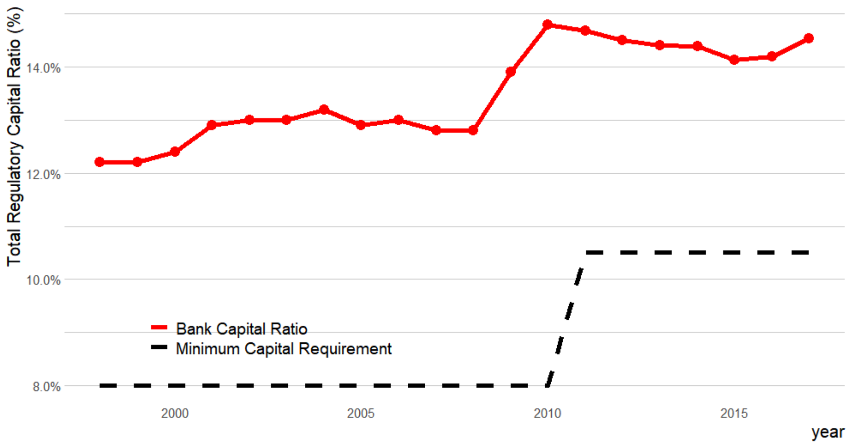
In fact, bank concentration and bank capital are two key concepts in the banking literature, while little work has been done to illuminate their relationship. In the United States, both bank concentration and the regulatory bank capital ratio have been increasing simultaneously, as observed in Figure 2.1. Specifically, panel A of the figure shows a decline in the total number of banks from 9,600 to 5,000 between 1996 and 2017, with the top three asset share increasing from 20% to 35% during the same period[1]. Conversely, as demonstrated in panel B, the total regulatory capital ratio in the United States has consistently risen over time and surpassed the minimum capital requirement represented by the black dashed line. This chapter presents a dynamic model of an imperfectly competitive banking sector with heterogeneous entrepreneurs to analyze the relationship between bank concentration and bank capital. The proposed model can also be leveraged to explore the

[1]The degree of market concentration in the banking sector can also be estimated using markup measures, as demonstrated in previous studies such as Bresnahan (1989), Berry et al. (1995), De Loecker and Warzynski (2012), De Loecker et al. (2020). In this chapter, I will estimate bank concentration using the Herfindahl index and detail the estimation procedure in subsequent sections.

impact of bank concentration on efficient allocation, considering various channels such as the accumulation of excess bank capital.



Panel A: Number of banks and top 3 asset shares



Panel B: Bank regulatory capital to risk-weighted assets in U.S.

Trend of Bank Concentration and Bank Capital Ratio in the U.S.

Figure 2.1: In panel (a), the blue line shows the number of banks in the U.S. over years, while the red line shows the assets of three largest commercial banks as a share of total commercial banking assets over years; in panel (b), the red line illustrates the total regulatory capital ratio in the United States and the black dashed line is the minimum capital requirement.(Source: FRED)

The model considers banks as the exclusive intermediaries for resource allocation among entrepreneurs with varying levels of productivity and wealth. Banks compete in the deposit and loan markets à la Cournot while adhering to a capital requirement, where they must maintain a specified level of capital with respect to their loan portfolio size. The presence of productivity heterogeneity allows me to discuss the distribution of resources among entrepreneurs of different levels of efficiency.

Chapter 2 presents two primary findings. Firstly, a higher level of bank concentration leads to a potentially non-binding capital requirement and a higher actual bank capital ratio. This is mainly a result of deposit market concentration. This prediction is supported by micro-level data from the US, which shows a positive correlation between deposit market Herfindahl index and risk-based bank capital to asset ratio. Secondly, the excess accumulation of bank capital exacerbates the distortive effect of bank concentration on efficient allocation, which is referred to as the *"bank capital channel"*.

There are three types of entrepreneurs in the equilibrium, who are classified as borrowing entrepreneurs, lending entrepreneurs, and autarky entrepreneurs, depending on their productivities. Entrepreneurs with the highest productivity level produce and borrow up to their limits, while those at the bottom prefer to hold all their resources in banks. Imperfect competition in the banking sector results in a positive net margin between the loan and deposit rates, which prompts some entrepreneurs (autarky entrepreneurs) to neither borrow nor lend. Instead, they use their initial holdings to engage in production activities.

The model generates two empirically verified outcomes concerning autarky entrepreneurs. Firstly, as bank competition falls, the net margin rises, causing a rise in the fraction of autarky entrepreneurs. Secondly, the increase in the proportion of autarky entrepreneurs has a distortionary effect on output as these entrepreneurs are the least efficient producers. Thus, bank concentration affects the efficient allocation of production resources via a *"net margin channel"*, directing more capital towards the autarky entrepreneurs through an extensive margin.

The model predicts that higher levels of bank concentration are associated with a potentially non-binding capital constraint and an increased bank capital to asset ratio. This positive relationship is largely due to deposit market concentration, which leads to a decline in deposit rates charged by banks, resulting in a lower deposit supply. Bank equity capital and deposits are the main sources of funding for banks, and this substitution effect between the two liabilities leads to an increase in bank capital. The borrowing constraint is identified as another factor that affects bank capital, as a higher borrowing limit motivates banks to accumulate more capital by raising the loan rate, which in turn increases the productivity of marginal entrepreneurs. As a result, banks achieve a higher marginal return on holding capital.

The model also examines the impact of bank concentration on the optimal allocation of resources in production, considering the interaction between bank concentration and bank capital. By solving the social planner's problem, the chapter identifies the optimal allocation between entrepreneurial initial capital and bank equity capital. A higher bank capital results in less endowment held by autarky entrepreneurs (benefit) but a lower average productivity of both autarky entrepreneurs and borrowing entrepreneurs (cost). The model indicates that banks are over-accumulating capital in terms of allocative efficiency, as the market solution involves a level of bank capital higher than that which maximizes total factor productivity (TFP). The mechanism that distorts the optimal allocation in production through the interaction between bank concentration and bank capital is referred to as the *"bank capital channel"*. The conflict between banks and the social planner arises from the fact that accumulating bank capital reduces banks' incentives to issue deposit and the associated costs, while the social planner values output and ignores profit allocation between bankers and entrepreneurs.

I conduct a quantitative evaluation of the efficacy of bank regulations in reducing bank capital ratios to optimal levels and enhancing efficiency. Three regulatory mechanisms, namely deposit rate floor, capital requirement ceiling, and raising transaction cost of bank capital, are compared and assessed. It is argued that the deposit rate floor is superior to the capital requirement ceiling and raising transaction cost of bank capital. While the capital requirement ceiling sustains the centralized equilibrium, the deposit rate floor is more effective in improving efficiency by reducing the proportion of autarky entrepreneurs. In contrast, introducing transaction costs of bank capital raises the fraction of autarky entrepreneurs by increasing the loan rate. The analysis finds that raising the deposit rate floor from 2.5% to 2.87% leads to a 1% increase in output and meets the minimum capital ratio requirement.

Related Literature

This chapter contributes to the existing literature on bank market power. While bank concentration is a suggestive indicator of bank market power, it is important to distinguish between the two concepts. Within the realm of bank market power, scholars have pursued two main avenues of research: examining the impact of bank market power on the real economy and exploring the implications for the transmission of monetary policies. To answer the first question, Drechsler et al. (2017), Wang et al. (2020), Scharfstein and Sunderam (2016), Ulate (2021) provide insights on how bank market power in either deposit market or loan market affects the transmission to monetary policies. Meanwhile, the relationship between bank market power and real economy has been empirically examined by Jayaratne and Strahan (1996), Black and Strahan (2002), Diez et al. (2018), and Joaquim et al. (2019). Chapter 2 contributes to the literature by examining a previously unexplored channel through which bank market power influences resource allocation and output. Specifically, the chapter focuses on the role of "bank capital" as a key determinant in this relationship.

The theoretical work is related to the heterogeneous agent models. The entrepreneurs' side of the model is built on Angeletos (2007), Kiyotaki and Moore (2019) and Moll (2014). Angeletos (2007) examines the effect of incomplete markets à la Bewley without a borrowing constraint. Kiyotaki and Moore (2019) include the borrowing constraint and study its effect on aggregate fluctuations. Moll (2014) relaxes the assumption of independently and identically distributed (i.i.d.) productivity shocks made in two previous studies and demonstrates the impact of productivity shock persistence on resource misallocation. Building upon their work, I incorporate the problem faced by bankers in this framework and examine how bank concentration is linked to resource misallocation in the presence of imperfect financial markets

This chapter adds to the literature on the effects of micro distortions on macroeconomic outcomes (Hsieh and Klenow (2009), Bartelsman et al. (2013)). Particularly, Hsieh and Klenow (2009) reveal significant discrepancies in the productivity of labor and capital among various agents in China and India. This capital and labor misallocation leads to a reduction in the manufacturing Total Factor Productivity (TFP). In this chapter, I identify two major factors contributing to capital misallocation: financial frictions and bank concentration.

This is not the first theory that examines bank capital. Some models adopt static frameworks that treat bank capital as a parameter rather than a choice (Brunnermeier and Koby, 2018). Other models assume exogenous law of motion for bank capital (Li, 2019). A further class of models considers the cost of bank capital to be prohibitively high, leading to binding capital constraints (Repullo, 2004). In contrast, my model endogenously determines bank capital by optimizing the trade-off between dividend payouts and equity capital issuance. This specification enables me to analyze the relationship between bank concentration and bank capital, while also considering non-binding capital constraints.

An extensive body of literature has developed that pertains to nonbinding capital constraints. Empirical evidence suggests that banks willingly hold more capital than the minimum required and modify their capital ratios regardless of capital regulations. For example, Alfon et al. (2004) reveal that banks in the U.K. increased their capital ratios despite a decrease in the minimum capital requirement. Flannery and Rangan (2008) report that the U.S. banking industry underwent a dramatic capital accumulation, with a half large bank holding companies doubling their equity ratios in the past decade. From the theoretical perspective, this chapter is related and complementary to recent studies explaining non-binding capital constraints. Allen et al. (2011) attribute the positive capital ratio to asset discipline, with bank capital and loan rates serving as two tools to encourage banks to monitor, and banks favoring bank capital in specific regions. Additionally, Corbae et al. (2021) propose a dynamic quantitative model and suggest that "capital ratios are above what regulation defines as well capitalized suggests a buffer stock motive". Other papers, such as Blum and Hellwig (1995), Bolton and Freixas (2006) and Van den Heuvel (2008), describe a similar "capital buffer". In this chapter, I present a supplementary explanation for why banks accumulate positive capital even in the absence of risk. Bank concentration could be another, but not the only force that drives the buildup of bank capital. Indeed, Flannery and Rangan (2008) report that there is not a significant correlation between portfolio risk and capitalization from 1986 to 2001.

A unique feature of the model is the emergence of autarky entrepreneurs because of imperfect competition in the banking industry. These entrepreneurs are akin to on-account workers in the labor literature who are self-employed and do not employ others. A considerable body of literature exists on intermediation costs, on-account workers, and real outcomes, with cross-country evidence indicating a negative relationship between the proportion of on-account workers and per capita income (Gindling and Newhouse (2014)). Cavalcanti et al. (2021) and Gu (2021) show that a higher share of on-account workers results from a larger intermediation cost caused by financial frictions. This chapter contributes to this literature in two significant ways: by emphasizing capital market allocations over labor market allocations and by exploring the effect of financial friction and bank concentration.

2.2 More Stylized Facts

2.2.1 Data Description

In this chapter, I employ a combination of three distinct data sources to perform our analysis. Firstly, I utilize the Summary of Deposits data from the Federal Deposit Insurance Corporation (FDIC). Secondly, I draw upon bank balance sheet data from U.S.

Call Reports, which is made available by the Federal Reserve Bank of Chicago. Finally, I extract additional bank-specific characteristics from the Research Information System (RIS) Database, also provided by the FDIC. In this section, I outline the salient features of each of these datasets.

Deposit Quantity The dataset on deposit quantities is obtained from the Federal Deposit Insurance Corporation (FDIC), encompassing all U.S. bank branches from 1994 to 2020. The data provides information on a variety of branch characteristics, including ownership details and deposit quantities at the county level. To facilitate analysis, the unique FDIC bank identifier is employed to link this dataset with other relevant datasets.

Bank Balance Sheet The bank data is from U.S. Call Reports provided by the Federal Reserve Bank of Chicago, spanning from March 1994 to March 2020. The Call Reports provide quarterly balance sheet information on all U.S. commercial banks, including details on assets, deposits, various loan types, and equity capital, etc. The Call Reports are matched with the FDIC data using the FDIC bank identifier.

More Bank Characteristics Other bank characteristics are obtained from RIS Database, FDIC. It contains financial data and history of all entities filing the Call Report at a quarterly frequency from March 1984 to June 2021. It includes crucial capital ratio variables based on diverse criteria. The RIS data is linked to the previously mentioned datasets using the FDIC bank identifier.

In the empirical analysis, two essential variables that require identification are bank concentration and bank capital. Consistent with prior literature, I use the Herfindahl-Hirschman Index (HHI) as a standard measure of market concentration in the banking industry (Drechsler et al., 2017). Specifically, I compute the branch-level HHI as the sum of the squared deposit market share of each bank institution by county for each year. To obtain the bank-level HHI, I calculate the weighted average branch-level HHI of all branches belonging to the same bank institution, using branch deposits as weights. Note that I use the time-varying bank-level HHI, which differs from the main analysis in Drechsler et al. (2017). To address outliers, both bank-level HHI and branch-level HHI are winsorized at the 1% and 99% levels.

Capital ratio is defined as the risk-based capital ratio at the bank level under Prompt Corrective Action (PCA), a regulatory framework that evaluates a bank's capital adequacy and supervisory rating to determine whether it is at a heightened risk of stress or failure. Tier 1 risk-based capital ratio is used as a proxy for bank capital, although total risk-based capital ratio is considered as an alternative measure. The capital ratio variables are also winsorized at the 1%- and 99%- level to remove outliers.

2.2.2 Bank Capital and Bank Concentration

I conduct a fixed-effect regression to examine the relationship between the Herfindahl index (HHI) and the risk-based capital to asset ratio. The regression model is specified as follows:

$$CAR_{it} = \alpha_i + \alpha_t + \gamma HHI_{it-1} + \beta Controls_{it-1} + e_{it} \tag{2.1}$$

where CAR_{it} represents the Tier 1 (Total) risk-based capital to asset ratio for bank i in quarter t, α_i and α_t are the bank and quarter fixed effects, respectively, and HHI_{it-1} denotes the bank-level HHI for bank i in quarter $t-1$. To address potential endogeneity issues, I use lagged values of the bank-level HHI and controls. Standard errors are clustered at bank level. The main coefficient of interest is γ, which measures the correlation between bank HHI and the risk-based capital to asset ratio. Additionally, the return on assets (ROA) is included as a control variable in the regression to proxy for earnings.

Under different specifications, a statistically significant positive coefficient (γ) is found for the relationship between bank concentration and risk-based capital ratios. Specifically, in Table 2.1, columns (1) and (2) analyze the relationship between bank concentration and total capital to risk-weighted asset ratio, yielding an estimated γ of approximately 0.04 that is statistically significant at the 1% level. This finding implies that a transition from a bank-level HHI of 0 to 1 leads to a 0.04% increase in the bank capital ratio. A similar positive relationship between bank concentration and Tier 1 capital to risk-weighted asset ratio is observed in columns (3) and (4) of the same table. These results provide evidence that bank concentration is positively correlated with bank capital.

The empirical results discussed above provide the impetus for me to develop a model that examines the interplay between bank concentration and bank capital. To this end, I build a model that extends Moll (2014) by incorporating imperfect competition within the banking sector.

Bank Concentration and Capital to Risk Weighted Asset Ratio

Variables	Total Capital to RWA Ratio		Tier 1 Capital to RWA Ratio	
	(1)	(2)	(3)	(4)
Bank-level HHI	0.0393***	0.0394***	0.0392***	0.0395***
	(0.0064)	(0.0064)	(0.0064)	(0.0064)
Return on Assets		-1.26***		-1.25***
		(0.110)		(0.110)
Bank Fixed-effect	Yes	Yes	Yes	Yes
Quarter Fixed-effect	Yes	Yes	Yes	Yes
Observations	763018	763018	763018	763018
R-squared	0.0460	0.0482	0.0462	0.0484

Table 2.1: This table presents an estimation of the relationship between bank concentration and bank capital, using data at the bank-quarter level covering the period from 1994 to 2020. Specifically, columns (1) and (2) report results using the total capital to risk weighted asset ratio as the dependent variable, while columns (3) and (4) use the Tier 1 capital to risk weighted asset ratio. I additionally control for the return of assets in columns (2) and (4). The Standard errors are clustered at bank level. *** indicates significance at the 1% level.

2.3 Model Environment

Consider a discrete time economy with infinite horizon, where time is indexed by $t = 0, 1, 2, \cdots$. The model describes the credit structure in an economy consisting of three types of agents, namely entrepreneurs, bankers and capital suppliers. At each period, bankers intermediate resources among a continuum of ex-ante heterogeneous entrepreneurs, while capital suppliers supply capital to both bankers and entrepreneurs.

2.3.1 Entrepreneurs

There is a continuum of infinitely lived entrepreneurs, who are indexed by their initial capital a and productivity z. Productivity z is assumed to follow an exogenously given distribution $G(z)$ that is identically and independently distributed (*i.i.d.*). I assume the law of large numbers so that the distribution of entrepreneurs of a specific productivity is deterministic at each period. Entrepreneurs have preferences

$$\sum_{t=0}^{\infty} \beta^t log(c_t)$$

At period t, entrepreneurs are endowed with a linear production technology, which allows them to use capital as an input in production with return z_t:

$$y_t = z_t k_t$$

Capital is assumed to fully depreciate after production.

During the middle of each period, entrepreneurs participate in the loan and deposit market. Entrepreneurs have the option to borrow from the bankers and repay the loan at an interest rate of r_t^b, or to deposit funds in the bank and withdraw them at a return of r_t^d. Following the financial market transactions and production, each entrepreneur optimally decides the amount to consume and invests the remaining resources to purchase capital from the capital suppliers at the end of the period. The entrepreneur's budget constraint is therefore

$$c_t + q_t a_{t+1} \leq \Pi_t \equiv \begin{cases} z_t k_t - q_t(r_t^b + 1)(k_t - a_t) & k_t \geq a_t \\ z_t k_t + q_t(r_t^d + 1)(a_t - k_t) & k_t \leq a_t \end{cases} \tag{2.2}$$

where q_t is the price of the capital. Each entrepreneur generates income by producing output and earning interest on deposits or paying interest on loans from bankers. This income is used for consumption and investment in capital.

Additionally, entrepreneurs face a borrowing constraint that limits the amount of funds they can borrow

$$k_t \leq \lambda a_t, \ \lambda \geq 1 \tag{2.3}$$

The parameter λ captures the degree of market imperfection in the financial market, where higher values of λ indicate greater efficiency of the market. When λ is infinite, the financial market is perfect, whereas when λ is 1, the financial market is shut down and all entrepreneurs remain in autarky. The actual leverage ratio of the entrepreneur is denoted by $\theta_t = k_t/a_t$.

2.3.2 Bankers

The banking sector is characterized by assuming the presence of imperfect competition. Specifically, the economy is assumed to have a total of $M \geq 1$ bankers[2], each competing

[2] M is an integer. The parameter M serves as a unified measure encompassing both deposit market power and loan market power. As shown in subsequent sections of this chapter, my primary objective is to establish a positive relationship between bank concentration and bank capital by incorporating deposit market power. However, in order to achieve a more refined analytical solution, I introduce both deposit market power and loan market power in the baseline model.

for the quantity of loans Q^L_{it} and deposits Q^D_{it} à la Cournot. The case where $M = 1$ represents a monopoly bank, whereas in the limit as M approaches infinity, the banking sector is perfectly competitive. At beginning of each period, each banker i is endowed with some equity capital N_{it}. Bankers are risk neutral and have preferences over dividend payouts

$$\sum_{t=0}^{\infty} \beta^t c^b_{it}$$

Banker serves as a financial intermediary and facilitates borrowing and lending between entrepreneurs. The loans are the sole asset on bankers' balance sheet and are financed by equity capital and deposits. Bank equity capital is accumulated through retained earnings. Table 2.2 summarizes the balance sheet items at the start of each period t.

Assets	Liabilities
Loans (Q^L_{it})	Deposits (Q^D_{it})
	Equity capital (N_{it})

Table 2.2: Bankers' Balance Sheet

Banker i's balance sheet identity can be expressed as

$$Q^L_{it} = Q^D_{it} + N_{it} \tag{2.4}$$

The dividend payouts and equity capital accumulation of the banker through retained earnings can be simplified to a standard consumption and savings problem in the model. Banker i faces a budget constraint given by

$$c^b_{it} + q_t N_{it+1} \leq (1 + r^b_t) q_t Q^L_{it} - (1 + r^d_t) q_t Q^D_{it} \tag{2.5}$$

The right-hand side terms in the above equation represent the banker's income, which is the return from investing in the loans market, minus the repayment to depositors. The left-hand side terms in the equation denote the banker's consumption of dividends and accumulation of equity capital.

Bankers also face a minimum capital requirement

$$N_{it} \geq \kappa Q^L_{it}, \tag{2.6}$$

where κ represents the extent to which the minimum capital requirement is adjustable. This requirement mandates that a proportion of bank loans be financed through capital, and was first introduced by the Basel Committee on Banking Supervision in 1996 to prevent banks from being vulnerable to losses arising from changes in the economic landscape. I integrate this requirement into the model to shed light on the empirical observations that the capital requirement may not be binding.

2.3.3 Capital Supplier

There is a continuum of capital suppliers, who are endowed with $\overline{K}$ units of capital. At the end of each period, the entrepreneurs and bankers have the opportunity to purchase capital from these suppliers in a perfectly competitive capital market.

2.4 Equilibrium Characterization

This section presents the model equilibrium and uses the results to domonstrate the positive relationship between bank concentration and bank capital. Additionally, I discuss the ways in which imperfect banking competition leads to efficiency losses through two channels: the *"net margin channel"* and the *"bank capital channel"*.

2.4.1 Entrepreneurs' Side

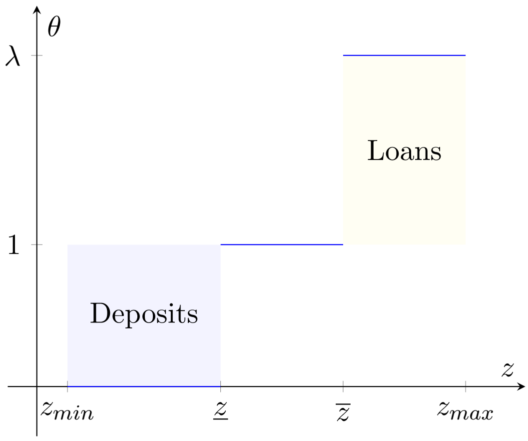
Owing to the existence of imperfect competition in the banking sector, a net margin $r_t^b - r_t^d$ is levied on the transactions conducted by the bankers. Given the deposit and loan rate, entrepreneurs' financial decisions regarding borrowing or lending are characterized in Lemma 2.1.

Lemma 2.1. *There are two productivity cutoffs $\underline{z_t}$ and $\overline{z_t}$ and the capital demand for individual entrepreneur is:*

$$k_t = \begin{cases} \lambda a_t & z_t \geq \overline{z_t} \\ a_t & \underline{z_t} \leq z_t \leq \overline{z_t} \\ 0 & z_t \leq \underline{z_t} \end{cases}$$

The productivity cutoff is defined by $\underline{z_t} = q_t(r_t^d + 1)$ and $\overline{z_t} = q_t(r_t^b + 1)$.

The cutoff property relies heavily on the constant return to scale of the production function. According to Lemma 2.1, the optimal capital demand decision is at corners: it is zero for entrepreneurs with low enough productivity, maximum amount allowed by the



Leverage Ratio for Different Entrepreneurs

Figure 2.2: The blue area is the deposit size and the yellow area is the loan size.

borrowing constraint for those with high enough productivity and initial wealth for those with intermediate level productivity. There are two types of marginal entrepreneurs. For the entrepreneurs with productivity $\underline{z_t}$, the return of each additional unit of capital investment $\frac{\underline{z_t}}{q_t}$ equals the opportunity cost of not depositing that in the bank $r_t^d + 1$; while for those with productivity $\overline{z_t}$, the return of each additional unit of capital investment $\frac{\overline{z_t}}{q_t}$ equals the cost of acquiring that unit $r_t^b + 1$. This heterogeneity in productivity among entrepreneurs generates an endogenous loan demand and deposit supply in the economy, as illustrated in Figure 2.2.

It is now sensible to call the entrepreneurs with productivity below $\underline{z_t}$ as lending entrepreneurs, those with productivity above $\overline{z_t}$ as borrowing entrepreneurs, and those with productivity in between as autarky entrepreneurs. Lending entrepreneurs possess such low levels of productivity that investing all their capital in banks appears more viable than engaging in production activities. Conversely, borrowing entrepreneurs exhibit productivities that surpass the effective loan rate, rendering borrowing from banks a profitable venture. Additionally, imperfect competition in the banking sector engenders a third category of entrepreneurs. Bankers impose a net margin between the loan rate and deposit rate, allowing some entrepreneurs to opt for production activities without borrowing. These entrepreneurs are referred to as autarky entrepreneurs.

The model has not explicitly modeled the distribution of initial wealth, as the assumption of independent and identically distributed productivity has been made. However, in order to establish a clear definition of entrepreneurs' aggregate capital, it is necessary to assume a joint distribution of (a, z) at time t, denoted as $h_t(a_t, z_t)$. Therefore, entrepreneurs' aggregate capital K_t is as follows

$$K_t = \int a_t dH_t(a_t, z_t) \tag{2.7}$$

To characterize the aggregates, the share of wealth held by productivity type z is

$$\omega(z_t, t) \equiv \frac{1}{K_t}\int_0^\infty a_t h_t(a_t, z_t) da_t = g(z_t)$$

where the first equality is following definition presented in Kiyotaki (1998) and Moll (2014), and the second equality follows by the independence between a_t and z_t.

The financial decisions and intertemporal optimization of entrepreneurs lead to an endogenous demand for loans and supply of deposits, alongside a law of motion for entrepreneurs' aggregate capital.

Lemma 2.2. *Denote Q_t^L and Q_t^D as the loan size and deposit size respectively. Aggregate quantities $\{Q_t^L, Q_t^D, K_{t+1}\}$ satisfy:*

$$Q_t^L = (1 - G(\overline{z_t}))(\lambda - 1)K_t \tag{2.8}$$

$$Q_t^D = G(\underline{z_t})K_t \tag{2.9}$$

$$\begin{aligned} q_t K_{t+1} = \beta\{&\int_{z_{min}}^{\underline{z}} \left[q_t(1 + r_t^d)\right] dG(z_t) + \int_{\underline{z_t}}^{\overline{z_t}} z_t dG(z_t) \\ &+ \int_{\overline{z_t}}^{z_{max}} \left[\lambda[z_t - q_t(1 + r_t^b)] + q_t(r_t^b + 1)\right] dG(z_t)\}K_t \end{aligned} \tag{2.10}$$

Equation (2.8) reveals that the aggregate loan demand is determined by three key factors: the fraction of borrowing entrepreneurs, the borrowing limit and entrepreneurs' initial capital holding. Similarly, the deposit supply is contingent upon the initial capital of lending entrepreneurs, as described by equation (2.9). Meanwhile, the law of motion for aggregate capital is encapsulated by equation (2.10). The future wealth of entrepreneurs, $q_t K_{t+1}$, depends on the saving rate β and the net return of entrepreneurs. Specifically, the three terms contained within the brackets on the right-hand side of equation (2.10) represent the return rates of depositors, borrowers, and autarky entrepreneurs, respectively. Notably, the constant saving rate across all entrepreneurs stems from the log utility functional form and the constant return to scale production function.

2.4.2 Bankers' Side

The optimal loan (deposit) rate is a function of the mark-up (mark-down) on banker i's marginal cost (benefit):

$$q_t(1+r_t^d) = \frac{\epsilon_t^d}{\epsilon_t^d + s_{it}^d} q_t \mu_{it} \tag{2.11}$$

$$q_t(1+r_t^b) = \frac{\epsilon_t^b}{\epsilon_t^b - s_{it}^b} (q_t \mu_{it} + \kappa q_t \chi_{it}) \tag{2.12}$$

where $q_t \mu_{it}$ is the multiplier on the balance sheet identity and $q_t \chi_{it}$ is the multiplier on the bank capital constraint. Equation (2.11) specifies that the deposit rate is determined solely by the marginal benefit of issuing deposits, which is the multiplier on the balance sheet identity. Meanwhile, the marginal cost of issuing loans is reflected in equation (2.12), as it tightens both the balance sheet identity and the capital constraint by κ. Moreover, the loan and deposit rates are influenced by mark-up and mark-down, which are functions of loan demand elasticity ϵ_t^b, deposit supply elasticity ϵ_t^d, and market shares of loans s_{it}^b and deposits s_{it}^d held by each banker. The Euler equation (2.13) is derived from the optimal condition for bank capital:

$$q_t = \beta q_{t+1}(\mu_{it+1} + \chi_{it+1}) \tag{2.13}$$

which equalizes the marginal benefit and cost of accumulating equity capital.

2.4.3 Steady State Equilibrium

In this section, I will define the symmetric equilibrium and subsequently focus on the steady state. To derive an analytical solution, I impose $\kappa = 0$, and assume a uniform distribution of productivity $U[z_{min}, z_{max}]$. It is important to note that all the results obtained in the decentralized equilibrium are contingent upon these specified assumptions. However, these functional assumptions do not fundamentally alter the principal findings presented in this chapter.

Definition 2.3 (Symmetric Equilibrium)**.** A *Symmetric Equilibrium* in the economy consists of a sequence of policy function of bankers' consumption, banker's equity capital holding $\{c_{it+1}^b, N_{it+1}\}_{t=0}^{\infty}$, a sequence of aggregate quantities for entrepreneurs $\{K_{t+1}, Q_t^D, Q_t^L\}_{t=0}^{\infty}$, a sequence of interest rates $\{r_t^b, r_t^d\}_{t=0}^{\infty}$, and a sequence of price $\{q_t\}_{t=0}^{\infty}$ such that:

1. Each entrepreneur maximizes life-time utility given loan rate, deposit rate and the price of capital;

2. Bankers maximize their life-time utility given (2.4), (2.5), (2.6) by competing for loans and deposits;

3. Bankers choose the same quantities for all assets and liabilities;

4. Market clearing condition for

   - loan market: $\sum_{i=1}^{M} Q_{it}^{L} = Q_t^L$;
   - deposit market: $\sum_{i=1}^{M} Q_{it}^{D} = Q_t^D$;
   - capital market: $\sum_{i=1}^{M} N_{it} + K_t = \overline{K}$.

Lemma 2.4. *Proportion of the autarky entrepreneurs is* $\frac{1}{M+1}$

The implication of Lemma 2.4 is that the presence of autarky entrepreneurs is contingent upon the level of competition in the banking sector. This is intuitively plausible since banks tend to charge a higher net margin in the presence of high bank concentration, thereby increasing the proportion of autarky entrepreneurs. This straightforward outcome enables me to concentrate on the conduct of borrowing entrepreneurs and lending entrepreneurs in the equilibrium.

Proposition 2.5. *There are two regions in the symmetric equilibrium: region 1 where the bank capital constraint is non-binding and region 2 where the bank capital constraint is binding.*

- *When* $\lambda > \overline{\lambda}(M)$, *equilibrium lies in region 1.*
- *The cutoff* $\overline{\lambda}(M)$ *is an increasing function of M.*
- *Define the bank total capital to asset ratio as* $\frac{N}{N+Q^D}$, *where* N *and* Q^D *is the equilibrium level of aggregate bank capital and deposit. In region 1, either higher bank concentration* ($\frac{1}{M}$) *or larger borrowing limit* (λ) *leads to an increase of bank capital to asset ratio.*

Proposition 2.5 suggests that the presence of financial constraints and imperfect competition in the banking sector affect agents' incentives to accumulate capital. To comprehend the mechanics behind Proposition 2.5, it is crucial to examine the primary sources of friction in the model, namely, the imperfect financial market and imperfect competition in the banking sector.

To this end, I consider the benchmark model, in which the financial market is perfect and the banking market is perfectly competitive, leading to the convergence of λ and M to infinity. In this scenario, capital allocation between bankers and entrepreneurs becomes indeterminate, as entrepreneurs' capital and bank capital become perfect substitutes. It can be observed that only entrepreneurs with the highest level of productivity engage in borrowing and production, thereby possessing complete control over resources during the production process. As a result, the returns of both entrepreneurs' and bankers' capital are dictated by the most productive entrepreneur, rendering the two forms of capital perfectly substitutable.

In the case where the financial market is perfect while the banking sector is monopolistically competitive, banks hold all the capital in equilibrium and direct their deposits and capital towards entrepreneurs with the highest level of productivity. The absence of heterogeneity among borrowing entrepreneurs enables banks to capture all the profits generated by loans, leading to the accumulation of equity capital by bankers until they possess all the capital, thereby leaving entrepreneurs with no capital. However, it is noteworthy that when $\lambda = \infty$, the presence of market power in the banking industry does not influence the optimal allocation of resources.

In contrast, in the presence of an imperfect financial market with a perfectly competitive banking sector, entrepreneurs hold all capital, as holding capital is non-optimal for bankers given the equilibrium condition $\beta(1 + r_d) = \beta(1 + r_b) < 1$. This extreme case, subject to a non-negative capital constraint ($\kappa = 0$), aligns with Moll (2014) where bankers are not modeled explicitly. In contrast, this chapter depicts bankers as financial intermediaries who do not accrue any profits.

Referring back to Proposition 2.5, it follows that the capital constraint is not binding if the borrowing limit exceeds $\overline{\lambda}(M)$. Moreover, the monotonicity of the cutoff $\overline{\lambda}(M)$ in M indicates that the capital constraint is not binding when the banking sector is highly concentrated. When the capital constraint does not bind, higher bank concentration and borrowing limits result in a higher bank capital to asset ratio. To understand the positive relationship between financial market perfection and bank capital, one should consider the proportion of borrowing entrepreneurs. As the borrowing limit increases, borrowing entrepreneurs can obtain more loans, reducing both the proportion of borrowers and the heterogeneity of borrowing entrepreneurs. Consequently, bankers can extract a higher return from the borrowing entrepreneurs, which encourages them to accumulate more capital. The primary mechanism driving the positive correlation between bank concentration and bank capital is that in a more concentrated banking sector, the deposit rate decreases, which reduces deposit supply. Banks can raise funds for investment through deposits or capital. The substitution effect between the two liabilities increases bank capital. In the next section, I will provide a quantitative explanation for why the concentration in the deposit market dominates even when there is also loan market concentration.

Recall that the model economy encompasses two main frictions, namely imperfect competition in the banking sector and imperfect financial market. It is of interest to examine how these factors distort the equilibrium allocation from the efficient outcome. Notably, the output takes the form:

$$Y = Z\overline{K} = (uE[z|\underline{z} \leq z \leq \overline{z}] + \lambda v E[z|z \geq \overline{z}])(\overline{K} - N), \tag{2.14}$$

where Y represents aggregate output, and Z denotes the average productivity of the economy. The first equality follows directly from the linear production function. In Equation (2.14), the proportion of autarky entrepreneurs is denoted by u and the proportion of borrowing entrepreneurs is denoted by v. The equation's second equality indicates that only entrepreneurs are capable of producing. The productivity of entrepreneurs is determined by five factors: the weighted average productivity of autarky entrepreneurs with a productivity level between $\underline{z}$ and $\overline{z}$, represented as $E[z|\underline{z} \leq z \leq \overline{z}]$, multiplied by their proportion u, plus the weighted average productivity of borrowing entrepreneurs with a productivity level greater than or equal to $\overline{z}$, represented as $E[z|z \geq \overline{z}]$, multiplied by their leverage ratio λ and their proportion v. At the beginning of each period, the family of entrepreneurs possesses $\overline{K} - N$ units of capital.

Proposition 2.6. *Suppose λ is finite. As the bank concentration $\frac{1}{M}$ rises, output falls.*

As in previous discussions, the implications of Proposition 2.6 are discussed with respect to the two primary market frictions. In a frictionless market, output should be $z_{max}\overline{K}$. However, in an imperfect financial market with a perfectly competitive banking sector, the decentralized equilibrium becomes inefficient, as not all capital is allocated to the most productive entrepreneurs. The introduction of imperfect competition in the banking sector further distorts efficiency in terms of output. Bankers with larger market power would charge a wider net margin. In Lemma 2.4, it is demonstrated that a larger net margin leads to a higher proportion of autarky entrepreneurs, who are characterized as the most inefficient producers, resulting in decreased output. This mechanism is referred to as the "*net margin channel*". Empirical research conducted by Joaquim et al. (2019) has examined this channel, indicating that a rise in bank competition and a reduction of spread in Brazil to global levels could yield an output increase of approximately 5%.

It is worth highlighting that in the scenario of a perfect financial market, where λ equals infinity, bank concentration does not cause a detrimental impact on output as bankers possess the entire capital. Consequently, in this situation, autarky entrepreneurs have no initial endowment, and bankers impose a positive net margin when M is less than infinity.

The question arises as to whether the "*net margin channel*" is the sole transmission mechanism through which bank concentration affects the efficient allocation. Proposition 2.5 illustrates a positive correlation between bank capital and bank concentration, which raises the possibility that this relationship may also have a bearing on aggregate output. In order to shed light on this issue, the central planner's problem will be analyzed in the subsequent section. This analysis will offer a deeper understanding of the interplay between bank capital, bank concentration, and the broader macroeconomic performance.

2.4.4 Optimal Capital Allocation in Production

Consider a central planner who maximizes the aggregate output of the economy. The central planner possesses the authority to allocate capital resources between the families of entrepreneurs and bankers. Subsequently, individual choices made by the entrepreneurs and bankers are expected to maximize their respective utilities. Specifically, the capital market is closed, and the responsibility of deciding the quantum of capital flowing to the entrepreneurs and bankers is delegated to the social planner.

To comprehensively analyze the optimization problem of the social planner, it is necessary to distinguish between two closely related concepts, namely "capital allocation in production" and "allocation between entrepreneurial initial capital and bankers' capital". The latter term pertains to the allocation of initial capital resources between entrepreneurs and bankers at the beginning of each period. This allocation is likely to impact the "capital allocation in production", which pertains to the allocation of capital among the entrepreneurs during the course of the period.

Suppose that the strategy adopted by the social planner is to establish $\frac{N}{K} = \kappa_0$. Output can be then represented as follows:

$$
\begin{aligned}
Y &= (uE[z|\underline{z} \leq z \leq \overline{z}] + \lambda a E[z|z \geq \overline{z}])(\overline{K} - N) \\
&= \overline{K}\{\frac{1}{M+1}\frac{1}{\kappa_0+1}E[z|\underline{z} \leq z \leq \overline{z}] + (1 - \frac{1}{M+1}\frac{1}{\kappa_0+1})E[z|z \geq \overline{z}]\},
\end{aligned}
$$

where the first equality is derived from the definition stipulated in Equation (2.14), while the second equality is given by Lemma 2.1 and social planner's choice. This equation illustrates that the average productivity of the economy can be expressed as the weighted average productivity of autarky entrepreneurs and borrowing entrepreneurs, with a weight of $\frac{1}{M+1}\frac{1}{\kappa_0+1}$ that depends on the bank concentration and social planner's choice.

Proposition 2.7 (Optimal Capital Allocation)**.** *Assume $G(z)$ follows $U[z_{min}, z_{max}]$. Denote κ_0^* as the optimal ratio of $\frac{N}{K}$. There exists an optimal capital allocation that satisfies:*

$$\kappa_0^* = Max\{\frac{\sqrt{\lambda - 1}}{M+1} - 1, 0\}$$

To attain a more thorough comprehension of the rationale underlying the optimal allocation, it is imperative to explore why optimality itself exists. A higher value of κ_0, denoting the level of bank capital, engenders a reduction in the initial capital holdings of entrepreneurial families, thereby resulting in a commensurate decrease in the amount of capital available to autarky entrepreneurs for production purposes. This is advantageous in terms of average productivity, as the weight assigned to autarky entrepreneurs will be correspondingly diminished. Additionally, the rise in bank capital leads to distortions in the initial capital endowments of borrowing entrepreneurs, which, in turn, curtails their borrowing capacity with binding borrowing constraints. A higher proportion of entrepreneurs resorting to borrowing represents an undesirable outcome since it brings about a reduction in the average productivity of both autarky entrepreneurs and borrowing entrepreneurs. Thus, the tradeoff ensures the existence of an optimal capital allocation.

Proposition 2.7 proffers valuable insights into the manner in which bank concentration and borrowing limits interact with optimal capital allocation. The optimal capital allocation is positively correlated with bank concentration, as suggested by Proposition 2.7. This is attributable to the advantage of higher bank capital, which reduces the amount of capital held by autarky entrepreneurs. An increase in bank concentration results in a higher proportion of autarky entrepreneurs, leading to a greater benefit when bank capital increases. Moreover, Proposition 2.7 implies that a higher borrowing limit results in a higher optimal bank capital. An increase in the borrowing limit has the effect of mitigating the distortion caused by higher bank capital on the average productivity of both autarky entrepreneurs and borrowers. Consequently, the optimal position for bank capital is increased.

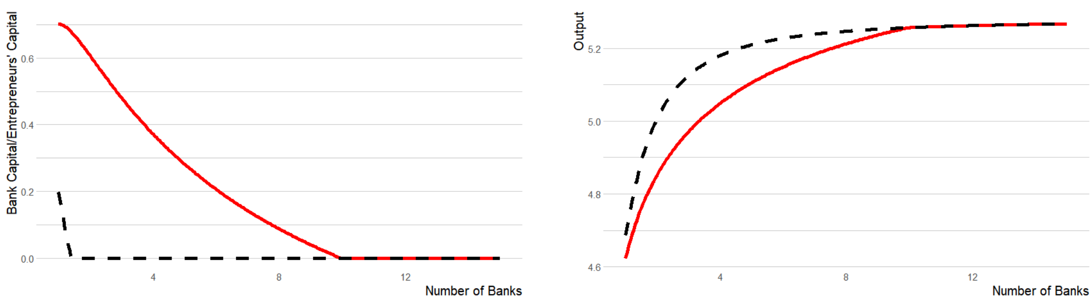
2.4.5 Bank Capital Channel

This section aims to investigate the impact of "the allocation between entrepreneurial initial capital and bank capital" on the "capital allocation in production". Specifically, it examines the extent to which the capital allocation in the decentralized equilibrium differs from that in the central planner's problem. Additionally, this study proposes a "bank capital channel" to gain insights into the relationship between bank concentration and misallocation.

Let the allocation between entrepreneurial initial capital and bank capital in the decentralized equilibrium be represented by $\frac{N^*}{K^*}$. Based on this, the following proposition can be derived in a straightforward manner.

Proposition 2.8. *Bankers are over-accumulating capital:* $\frac{N^*}{K^*} > \kappa_0^*$

Figure 2.3 portrays the implications of Proposition 2.8. The red line signifies the capital ratio and output in the decentralized equilibrium, while the same variable in the centralized equilibrium is depicted using a black dashed line. The decentralized equilibrium exhibits a consistently higher level of capital ratio output than the centralized equilibrium, with the intersection of the two lines on the left panel occurring when the natural capital constraint becomes binding. Therefore, a positive wedge arises between optimal output and output in the decentralized equilibrium when the capital ratios in the two scenarios differ.



Panel A: Capital Ratio | Panel B: Output (TFP)

Centralized Equilibrium v.s. Decentralized Equilibrium

Figure 2.3: The red line is decentralized capital ratio (output) between bankers and entrepreneurs, while the dashed black line is optimal one.

According to Proposition 2.8, excessive levels of bank capital can lead to inefficiencies in allocation. To elucidate this point, it is beneficial to examine the differences between the objectives of bankers and social planner and identify the pecuniary externalities. Central planner maximizes the output, which is expressed as:

$$Y = uKE[z|\underline{z} \leq z \leq \overline{z}] + vK\lambda E[z|z \geq \overline{z}] \tag{2.15}$$

The bank capital ratio in the decentralized equilibrium is established through the optimal decision of bankers, who strive to maximize their lifetime utility. In the steady state, bankers maximize the period consumption:

$$\begin{aligned} c_b &= qr^b(\lambda - 1)vK - qr^d(1 - v - u)K \\ &= vK(\lambda - 1)\frac{r^b}{1 + r^b}\overline{z} - (1 - v - u)K\frac{r^d}{1 + r^d}\underline{z} \end{aligned} \tag{2.16}$$

The initial row in equation (2.16) reveals that consumption is subject to the net return of loans and the costs associated with deposits, while the second row is obtained through the replacement of the loan and deposit rates with the productivity of marginal entrepreneurs. Given the substantial disparity in the objectives of the central planner and the bankers, determining the cause of bank capital overaccumulation may not be a straightforward process. Therefore, it would be useful to compare the factors present in both equations and assess how differences in each component contribute to distinct motives.

Both bankers and the social planner reap benefits from lending activities. Bankers earn $vK(\lambda - 1)\frac{r^b}{1+r^b}\overline{z}$ on loan lending, while the social planner values loans as a means of providing resources to more productive entrepreneurs, reflected in the second element in Equation (2.15). These two elements differ in three ways. Firstly, bankers place value on profits solely based on loan size, while the social planner values returns from both loans and the self-investment of borrowers, denoted by $\lambda - 1$ and λ, respectively. Secondly, the social planner is not subject to a capital cost when issuing loans, denoted by $\frac{r^b}{1+r^b}$, while the cost for bankers is 1. Lastly, the return on lending loans for bankers is based on the productivity of marginal entrepreneurs, who are indifferent between borrowing and staying autarky. In contrast, the social planner bases their returns on the average productivity of borrowers, represented by $\overline{z}$ and $E[z|z \geq \overline{z}]$, respectively. Accumulated bank capital leads to an increase in lending activities in both centralized and decentralized equilibria, but the social planner derives a higher return and incurs lower costs from lending activities relative to bankers. This inherent conflict between output and profit prompts the social planner to accumulate more bank capital than bankers, thus generate opposite implication to Proposition 2.8.

Bankers are not concerned with the behavior of the autarky entrepreneurs, which appears in the central planner's problem. (the first element in Equation (2.15)). Social planner recognizes that accumulating more bank capital entails allocating fewer resources to autarky entrepreneurs. This process follows the same mechanism as Proposition 2.7, where a higher bank capital corresponds to lower initial entrepreneurial capital and lower initial capital holdings for the autarky entrepreneurs. Hence, the social planner take this into account and accumulate more capital than bankers do.

Moreover, bankers incur costs on deposits, which are repaid to lenders, and this cost is not valued by the social planner. Bankers, however, could use bank capital to finance investment, which in turn reduces the cost of deposits. Consequently, bankers are motivated to accumulate more capital than intended by the social planner.

Recall that the deposit market concentration can result in an increase in bank capital, due to the substitution effect between bank capital and deposits. This effect may also be responsible for bankers holding excessive capital. The primary factor that drives these findings is believed to be the concentration in the deposit market, although the current model does not differentiate between the concentration in the deposit market and the loan market, both of which are subject to the influence of the number of bankers (M). To address this issue, an extended model will be presented in the following section, which allows for a separate variation of concentration in the deposit and loan markets, to quantitatively analyze the main findings.

2.5 Quantitative Analysis

This section of the study will begin by calibrating the parameters in the model. A comprehensive analysis of the possible policy implications will be presented based on the quantitative implication of the model. Following this, an extended model will be introduced with the aim of disentangling the bank concentration in both the deposit market and loan market.

2.5.1 Calibration

Parameters have been selected to match the key moments of the US economy between the years 2001 and 2020. Calibration of these parameters will also involve calibrating the distribution of productivity and the quality of financial institutions, represented by the limits of borrowing constraint (λ) for the US.

In the preceding sections, it was assumed that the distribution of productivity follows a uniform distribution, characterized by parameters z_{max} and z_{min}. Calibration of these two parameters will entail matching the first and second moments of the productivity distribution for US in the sample periods. As highlighted in Hsieh and Klenow (2009), the dispersion (standard deviation) of the logarithm of TFPQ[3] in the United States in 2005 is 0.84, and the difference between the 75th and 25th percentiles is 1.17. The probability distribution function of $log(z)$ is $\frac{e^z}{z_{max}-z_{min}}$ when the productivity z follows a uniform distribution[4]. Based on the distribution function, it is then feasible to establish z_{min} and z_{max}. Nevertheless, as indicated by the previous sections, the values of the two parameters are not of utmost importance since they do not impact the primary findings.

The model features two fundamental parameters, namely the parameter that regulates the quality of financial institutions denoted as λ, and the parameter governing the degree of bank concentration, represented by the inverse of the number of bankers in the market, denoted as $\frac{1}{M}$. By the definition of HHI, the relationship between the number of bankers in the model and bank concentration measure HHI is given by

$$HHI = \sum_{i=1}^{M} s_i^{d2} = \sum_{i=1}^{M} (\frac{1}{M})^2 = \frac{1}{M} \tag{2.17}$$

where the first equality follows the definition of HHI and the second equality follows that in the steady state of the symmetric equilibrium, all the bankers constitute $1/M$ market share in both deposit market and loan market. The average HHI in the US from 1994 to 2020 is calculated as the weighted average of branch-level HHI, using branch deposits for the weights, and amounts to 0.1342. Using Equation (2.17), I obtain $M = 7.45$. By matching the model's implied bank capital to asset ratio with that of the US in years between 2001 and 2017, I choose λ. A higher value of λ indicates a more efficient financial market in the economy. The model's bank capital to asset ratio is $1 - \frac{\sqrt{1+\lambda\frac{M2}{2M+1}}}{\lambda-1}$, while the average bank regulatory capital to risk-weighted assets for the US in years between 2001 and 2017, according to FRED, is 13.71%. Given M, we obtain an implied value of $\lambda = 6.74$.

[3] As reported in Hsieh and Klenow (2009), Total Factor Productivity Quality (TFPQ) is a measure of "physical productivity". The authors also introduce the concept of Total Factor Productivity Revenue (TFPR), which refers to "revenue productivity". In their paper, Hsieh and Klenow attempt to differentiate between these two measures, where the use of plant-specific deflator gives TFPQ, while the industry deflator provides TFPR. The TFPQ measure corresponds to the productivity captured in the baseline model used in Chapter 2.

[4] Assume there is a random variable X which follows a uniform distribution $U[a, b]$, and define $Y = log(X)$. The cumulative distribution function (cdf.) of X is $F_X(x) = Pr(X \leq x) = \frac{x-a}{b-a}$. Then the cdf. of Y is $F_Y(x) = Pr(Y \leq x) = Pr(log(X) \leq x) = Pr(X \leq e^x) = \frac{e^x-a}{b-a}$. The probability distribution function is therefore $\frac{e^x}{b-a}$.

The parameter κ is selected to satisfy the policy requirement prescribed by Basel III. The minimum Total Capital Ratio according to Basel III regulations is fixed at 8%. Moreover, the inclusion of the capital conservation buffer raises the required total capital amount for financial institutions to 10.5% of risk-weighted assets. As the model does not incorporate the risk exposure, κ is simply set at 0.08. It should be noted that the value of κ has no impact on the key results in the baseline model, but determines the regions in the equilibrium and the conditions under which the capital constraint is binding.

One period in my model corresponds to one year. Following Gali and Monacelli (2005) and Christiano et al. (2005), the discount factor β is calibrated at 0.96, which implies a riskless annual rate of about 4% in the steady state. Additionally, a depreciation rate of $\delta = 0.1$ is adopted to more realistically account for the capital's wear and tear, resulting in an annual depreciation rate of 10%. The baseline model requires a modification, whereby capital suppliers do not provide an exogenous amount of capital each period, but rather the aggregate capital remains constant at $\overline{K}$, normalized to 1. The calibration of all parameters is summarized in Table 2.3.

Parameters	Values	Target
β	0.96	Risk-free interest rate
δ	0.1	Annual rate of depreciation on capital
λ	6.74	Bank capital to asset ratio
M	7.45	Average HHI between 2000-2020
z_{max}	5.7	Hsieh and Klenow (2009)
z_{min}	≈ 0	Hsieh and Klenow (2009)
κ	0.08	Basel III regulations
$\overline{K}$	1	Normalized to 1

Table 2.3: Calibrated Parameter Values

2.5.2 Policy Implications

As has been demonstrated in preceding sections, bankers tend to accumulate an excessive amount of capital relative to the optimal level. The over-accumulation of bank capital may have negative implications for allocative efficiency, highlighting the need to consider appropriate policy measures to maintain the social welfare in a decentralized equilibrium. In this section, an examination is undertaken to explore the effectiveness of different policy measures in regulating the banking sector.

Deposit Rate Floor

Assume there is a deposit rate floor, which serves as a minimum limit for the deposit rate. Under this assumption, the equilibrium deposit rate at time t is given by:

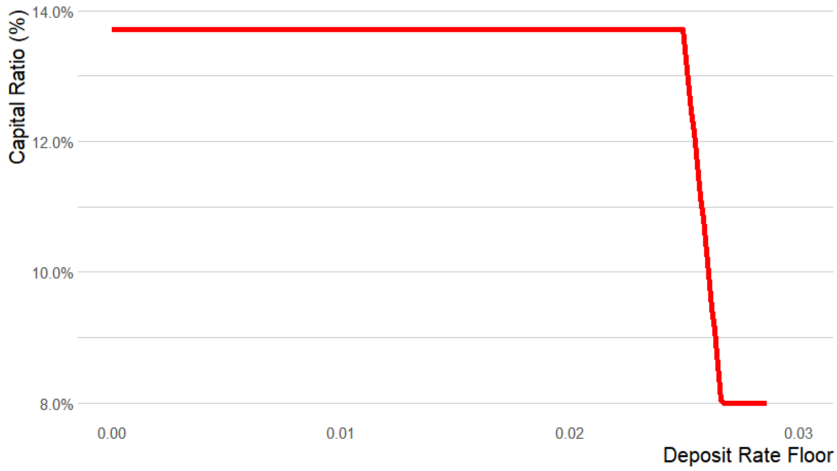
$$\tilde{r}_t^d = Max\{\overline{r}, r_t^d\} \tag{2.18}$$

where $1 + r_t^d = \frac{\epsilon_t^d}{\epsilon_t^d + s_{it}^d}\mu_{it}$ represents the equilibrium deposit rate in the absence of any restrictions, $\overline{r}$ denotes the minimum deposit rate allowed for bankers to set. The deposit rate floor becomes binding when the deposit rate r_t^d reaches the minimum deposit rate $\overline{r}$.

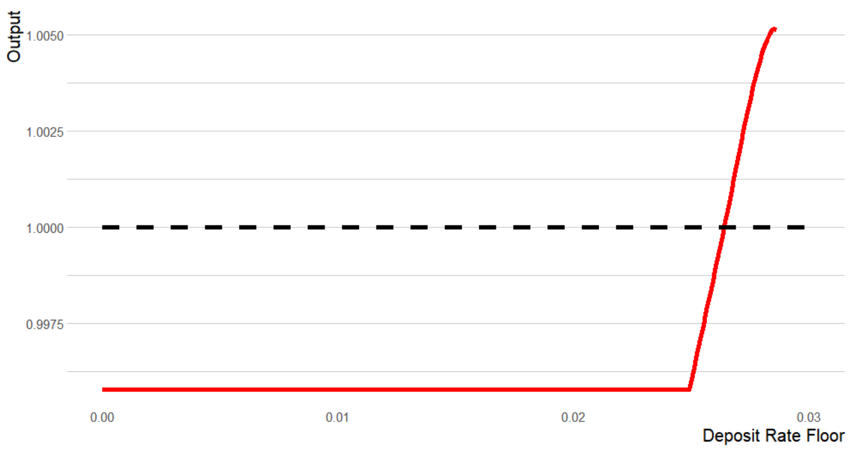
It is expected that intense competition among bankers would prevent them from charging excessively low deposit rates, hence avoiding the deposit rate floor from being reached. However, as the level of bank concentration increases, each banker's ability to charge a lower deposit rate increases, thus making it more likely for the deposit rate floor to become binding.

Using the calibrated parameters, Figure 2.4 shows the impact of the deposit rate floor on bank capital ratio (panel A) and output (panel B). The results reveal that when the deposit rate floor is low, the decentralized equilibrium is attained, and the capital ratio remains significantly higher than the minimum capital requirement, whereas the output is low. As the policy becomes more restrictive, deposit rate floors begin to take effect, causing a substantial decline in the capital ratio and an increase in output. The responses of output and capital ratio to the deposit rate floor are such that an increase in the deposit rate floor from 2.5% to 2.87% raises output by 1 percent and brings the capital ratio to the minimum requirement. The underlying intuition is straightforward: the deposit rate floor restricts the benefits of holding capital, leading to a decline in bankers' capital and an increase in the output.

In panel (b) of Figure 2.4, the optimal output level (black dashed line) is normalized to 1. This level is achieved when the social planner allocates a portion of capital to bankers to ensure that the bank capital to asset ratio meets the minimum capital requirement. It is worth noting that the optimal level of banking capital should ideally be zero. Nonetheless, in this instance, an effort has been made to make the centralized and decentralized equilibria comparable. Notably, it is observed that the deposit rate floor can result in an even higher output level than in the social planner's problem. This is due to the deposit rate floor's dual impact of forcing the capital ratio to an efficient level and simultaneously reducing the spread between the deposit rate and loan rate. The resulting decrease in the



Panel A: Capital ratio under different deposit rate floors



Panel B: Output (TFP) under different deposit rate floors

Effects of the Deposit Rate Floor

Figure 2.4: Capital ratio and output under different deposit rate floor are depicted with red lines. The black dashed line illustrates the output level in the central planner's problem, which is normalized to 1.

net margin implies a lower proportion of autarky entrepreneurs and, ultimately, a higher output level.

The deposit rate floor's ability to raise the output level is limited by its negative impact on the banks' return on intermediation. Continuously increasing the deposit rate floor would eventually result in negative returns for bankers, leading them to withdraw from the market and derive zero utility. Additionally, even when M is finite, an increase in the deposit rate will cause it to approach the loan rate. If the deposit rate floor is further raised, it would distort the loan size, leading to underutilization of redundant resources in the production process and ultimately an undesirable output level.

Transaction Cost of Bank Capital

An assumption commonly made in the literature is that equity capital is a more costly source of financing for bankers than deposits.[5] Suppose bankers are required to undertake transaction costs in order to accumulate capital, the budget constraint of the individual banker i would be modified as follows:

$$c_{it}^{b} + q_t N_{it+1} + C(N_{it+1}) \leq (1 + r_t^b) q_t Q_{it}^{L} - (1 + r_t^d) q_t Q_{it}^{D} \tag{2.19}$$

where $C(N_{it+1})$ is the cost must be incurred in the process of accumulating N_{it+1} amount of capital. A linear form $C(N) = cN$, where c is an exogenous constant, is assumed for the transaction cost. As a result, Equation (2.13) can be rewritten as follows:

$$q_t + c = \beta q_{t+1}(\mu_{it+1} + \chi_{it+1})$$

The introduction of transaction costs associated with bank capital may result in reduced motivation of capital accumulation for bankers. Figure 2.5 depicts the impact of transaction costs on the capital ratio and aggregate output (consumption) under various scenarios. As indicated in panel A of Figure 2.5, when the cost of holding capital for bankers escalates, the bank capital ratio declines. As the value of c increases from 0 to 0.053, the capital requirement becomes binding.

Panel B of Figure 2.5 illustrates the correlation between transaction costs and aggregate output. The panel reveals that aggregate output experiences a rise of 0.04% upon the

[5]The rationale behind the imposition of the assumption is that equity is more costly than debt. However, the theoretical basis for this assumption is lacking in the literature. The narrative that "equity is more profitable and costly" is challenged by scholars such as Miller (1995), Brealey (2006), and Admati et al. (2010).

capital requirement becoming binding. However, as the transaction cost increases further, output declines. This can be attributed to the fact that an increase in transaction cost leads to a higher loan rate, which causes a greater proportion of autarky entrepreneurs and decreases output. Additionally, this mechanism also explains why the introduction of transaction costs related to bank capital fails to attain the output level in the centralized equilibrium. Specifically, as depicted in panel B of Figure 2.5, the red line representing output consistently falls below the black dashed line representing optimal output.

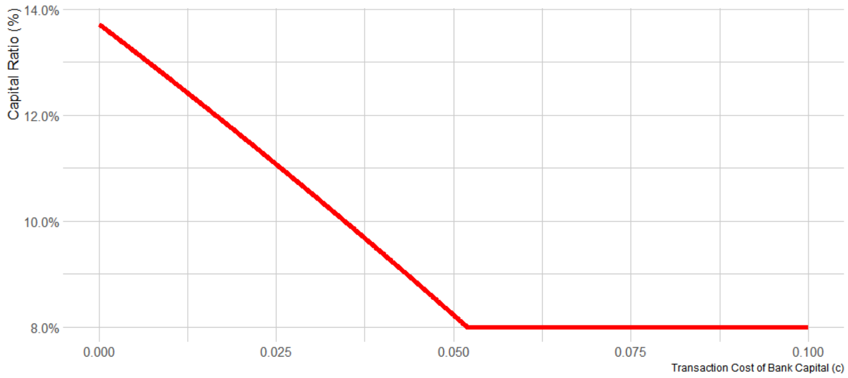
The effective output is also depicted in panel B of Figure 2.5, which is defined as aggregate output minus the resources that cannot be consumed or saved. Previously, aggregate output and effective output were indistinguishable in the absence of transaction costs on bank capital, as shown in panel B of Figure 2.5 where the red and green lines coincide for $c = 0$. However, as the transaction cost of bank capital increases, more resources are allocated to its accumulation, and the difference between aggregate and effective output expands. Although effective output follows a similar pattern as aggregate output, it might provide a more precise indication of welfare in this context. Both aggregate and effective output attain their maximum level when the bank capital ratio reaches the capital constraint.

Capital Requirement Ceiling

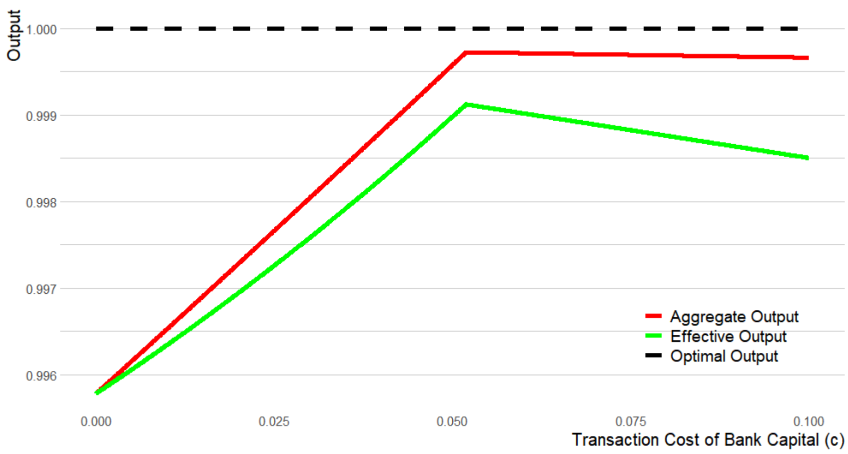
Following the financial crisis in 2008, policymakers implemented a minimum bank capital requirement to address the issue of risk exposure. High leverage ratios are known to incentivize banks to take risks, and the imposition of a bank capital requirement serves to mitigate these incentives by putting the bank's equity capital at risk. This chapter analyzes safe investments with different returns and, in the absence of risk considerations, examines the impact of changes in bank capital levels on the allocation of resources across different projects, providing insights into the role of bank capital in promoting allocative efficiency.

The preceding sections have demonstrated that bankers are accumulating an excessive amount of bank capital in comparison to the level observed in the centralized problem. The implications of these results suggest that implementing a capital requirement ceiling would aid in sustaining efficiency.

Proposition 2.9. *When there is no minimum capital requirement and a zero capital requirement ceiling, the capital allocation in the decentralized equilibrium is efficient when* $(M + 1)^2 + 1 \geq \lambda$.



Panel A: Capital Ratio under Different Transaction Costs of Bank Capital



Panel B: Output (TFP) under Different Transaction Costs of Bank Capital

Effects of Transaction Cost of Bank Capital

Figure 2.5: Capital ratio and output under different transaction cost floors are depicted with red lines. The black dashed line illustrates the output level in the central planner's problem, which is normalized to 1. Green line illustrates the effective output (output minus transaction cost)

The proof directly follows by the argument above. Both the capital ratio and output in the decentralized equilibrium is the same as that in the social planner's problem. When $(M+1)^2+1 \geq \lambda$, a zero capital ratio ceiling forces the bank capital ratio to zero, effectively replicating the optimal decision made by the social planner as implied by Proposition 2.7. Consequently, the allocation in the decentralized equilibrium precisely matches that of the central planner's problem.

In this section, I analyze three potential policies that could improve allocative efficiency: the deposit rate floor, the transaction cost of capital, and the capital requirement ceiling. The deposit rate floor is the most effective policy as it decreases the capital ratio and the proportion of autarky entrepreneurs. The introduction of the transaction cost of bank capital reduces the capital ratio but increases the fraction of autarky entrepreneurs, which lowers allocative efficiency. Notably, the decentralized equilibrium is identical to the centralized equilibrium when there is a capital requirement ceiling. Given these observations, policymakers may prefer the deposit rate floor to the capital requirement ceiling or the transaction cost of bank capital in their pursuit of improved allocative efficiency.

2.5.3 Disentangling Bank Deposit and Loan Market Concentration

The chapter establishes a positive correlation between bank concentration and bank capital. This relationship is premised on the ability of bankers to lower deposit rates in a less competitive banking industry, which leads to a decrease in deposit size. The substitutability of bank capital with deposits, in turn, drives up bank capital levels. Notably, the observed link between bank concentration and bank capital is contingent primarily on the concentration in the deposit market. However, in the baseline model, the number of bankers in the economy determines the concentration in both the deposit and loan markets. Consequently, as M varies, changes in the concentration of both markets occur simultaneously, which poses challenges in disentangling the impact of changes in deposit or loan market concentration alone on bank capital and allocative efficiency. This section aims to separate the effect of bank concentration in the deposit market and loan market for a clearer understanding of this relationship.

Consider the decisions faced by bankers in an economy with $M \geq 1$ bankers. I assume the effective deposit and loan market concentrations are no longer equal to $\frac{1}{M}$, but rather, $\frac{1}{M_d}$ and $\frac{1}{M_l}$, respectively. Here, $\frac{1}{M_d}$ refers to the effective deposit market concentration, and $\frac{1}{M_l}$ pertains to the effective loan market concentration. While both M_d and M_l may be dependent on the value of M, they need not be unit functions, as in the previous analysis where $M_d = M_l = M$. For instance, commercial banks in a particular geographic region may be more specialized in issuing deposits and have fewer operations in the loan market. Consequently, the deposit market concentration in such a location would be lower than the loan market concentration.

Given the aforementioned framework, the market clearing conditions for the deposit and loan markets are expressed as follows: $\sum_{i=1}^{M_d} Q_{it}^D = Q_t^D$ and $\sum_{i=1}^{M_l} Q_{it}^L = Q_t^L$. In the symmetric equilibrium, the optimal pricing for deposits and loans can be represented as:

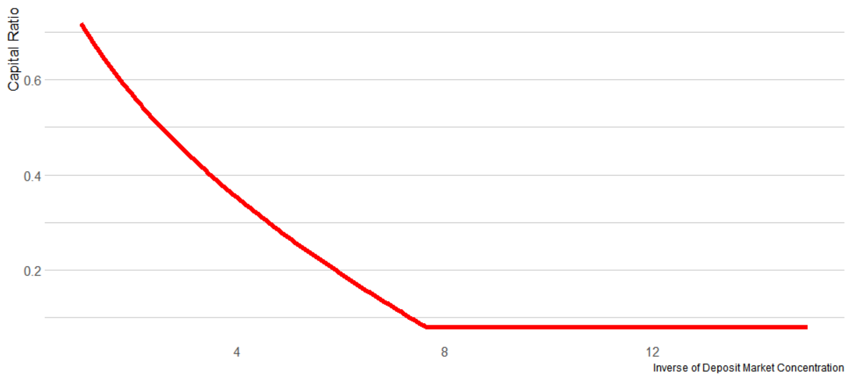
$$\delta + r_t^d = \frac{\epsilon_t^d}{\epsilon_t^d + 1/M_d}\mu_{it} \tag{2.20}$$

$$\delta + r_t^b = \frac{\epsilon_t^b}{\epsilon_t^b - 1/M_l}(\mu_{it} + \kappa\chi_{it}) \tag{2.21}$$

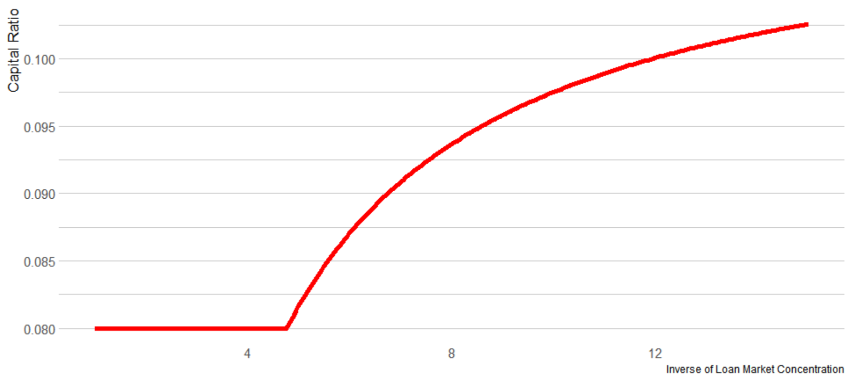
where on the left hand side of (2.20) and (2.21) shows the effective deposit rate and loan rate. The market shares held by individual bankers in the deposit and loan markets are determined by M_d and M_l.

Case 1 To investigate the impact of deposit market concentration on bank capital, I consider a scenario where loan market concentration remains fixed. Specifically, I set M_l to its calibrated parameter value of 7.45, and analyze the effects of varying M_d on bank capital. As illustrated in panel A of Figure 2.6, the results demonstrate that bank capital ratio increases as deposit market concentration rises. The observed relationship between bank capital and deposit market concentration in this scenario is precisely the same as that observed in the baseline model. Intuitively, the results suggest that when the deposit market is highly concentrated, bankers are inclined to set lower deposit rates. As deposit returns decline, fewer entrepreneurs are willing to accept deposit contracts, prompting bankers to accumulate more capital. Notably, when deposit market concentration (as measured by HHI) exceeds 0.25, the bank capital ratio may exceed 30%.

Case 2 To investigate the impact of loan market concentration on bank capital, I consider a scenario where deposit market concentration remains fixed. Specifically, I set M_d to a fixed value of 7.45 and examine how changes in M_l affect bank capital. As illustrated in panel B of Figure 2.6, the red line indicates a negative relationship between loan market concentration and bank capital ratio. This observed correlation is entirely opposite to that observed in Case 1. The results suggest that an increase in loan market concentration leads to a higher loan rate, resulting in smaller loan sizes for entrepreneurs. Due to the scarcity of investment opportunities, bankers accumulate less capital.



Panel A: Effect on Deposit Market Concentration on Bank Capital



Panel B: Effect on Loan Market Concentration on Bank Capital

Figure 2.6: Effects of Deposit (Loan) Market Concentration on Bank Capital

The two cases highlight the dominance of deposit market concentration as a driver of increases in bank capital. This conclusion is further supported by a comparison of the magnitudes of the capital ratio changes observed in the two cases, as depicted in the two panels of Figure 2.6. Specifically, the increase in deposit market concentration leads to a much larger rise in bank capital than the decrease in bank capital resulting from loan market concentration. Notably, in the baseline model, where $M = M_d = M_l$, I observe a positive correlation between bank concentration and bank capital ratio when the bank capital constraint is non-binding.

2.6 Conclusion

This chapter presents a dynamic model to explain why capital ratios surpass the minimum capital requirements. The model comprises two key elements: financially constrained entrepreneurs who are heterogeneous in productivity, and an imperfectly competitive banking sector. The analysis reveals that deposit market concentration plays a dominant role in driving up the bank capital ratio through a substitution effect between bank capital and deposits. Furthermore, the heterogeneity among entrepreneurs enables an investigation into how the interplay between bank concentration and bank capital influences capital allocation in production. The study indicates that banks hold excessive bank capital in terms of allocative efficiency. Based on these findings, the chapter provides several policy implications concerning bank capital and deposit rates.

The allocative efficiency in the model is expressed as a truncated weighted average of entrepreneurs' productivity, despite their rich heterogeneity. The bank capital ratio distorts TFP in two ways: (i) the capital allocation between autarky entrepreneurs and borrowing entrepreneurs, and (ii) the average productivity of marginal entrepreneurs. When both mechanisms are considered, the optimal bank capital for allocative efficiency is zero at the estimated parameters. This result does not contradict current policies, as it does not take into account banks' risk taking motives. Therefore, an extension of the framework that incorporates risky investments would be natural and significant. Such an extension would require banks and social planner to balance efficiency and stability, which could provide a more comprehensive understanding of the optimal bank capital. This issue was previously discussed in Chapter 1.

2.7 Appendix: Proofs

Proof for Lemma 2.1

Proof. All the entrepreneurs have a log preference over the current consumption. Specifically in the model, the entrepreneur maximizes its expected discounted utility of consumption subject to the budget constraint:

$$V(a_t, z_t) = \max_{\{c_t, k_t\}} \{log(c_t) + \beta V(a_{t+1}, z_{t+1})\}$$

$$s.t.\ c_t + q_t a_{t+1} \leq \begin{cases} z_t k_t - (r_t^b + 1)(k_t - a_t)q_t & k_t \geq a_t \\ z_t k_t + +(r_t^d + 1)(a_t - k_t)q_t & k_t \leq a_t \end{cases}$$

If $k_t \leq a_t$, denote the profit for lending entrepreneurs as $\Pi_t = z_t k_t - (r_t^d + 1)q_t k_t$. Π_t is positive if and only if $z_t \geq q_t(1 + r_t^d) \equiv \underline{z_t}$. $\underline{z_t}$ is an increasing function of r_t^d. Entrepreneur always produces a_t if the productivity is above the threshold.

Denote the net margin $s_t = r_t^b - r_t^d$. If $k_t \geq a_t$, denote the profit for the borrowing entrepreneurs as $\Pi_t' = z_t k_t - (r_t^b + 1)q_t k_t + q_t s_t a_t$. If $z_t \geq q_t(1 + r_t^b) \equiv \overline{z_t}$, entrepreneurs will produce and borrow up to the borrowing limit. $\overline{z_t}$ is an increasing function of r_t^b. □

Proof for Lemma 2.2

Proof. Equations (2.8) and (2.9) are directly obtained from Lemma 2.1, given that borrowing entrepreneurs borrow up to the borrowing limits and lending entrepreneurs deposit all their capital in the bank.

For lending entrepreneurs, the budget constraint becomes:

$$c_t + q_t a_{t+1} \leq (r_t^d + 1)q_t a_t$$

For borrowing entrepreneurs, the budget constraint becomes:

$$c_t + q_t a_{t+1} \leq \lambda(z_t - (r_t^b + 1)q_t)a_t + (r_t^b + 1)q_t a_t$$

For autarky entrepreneurs, the budget constraint becomes:

$$c_t + q_t a_{t+1} \leq z_t a_t$$

Because of the constant return to scale of the production function and log utility functional form, the saving rate is β. Therefore, the savings of the three types of entrepreneurs are: $q_t a_{t+1} = \beta[(r_t^d + 1)q_t a_t]$, $q_t a_{t+1} = \beta[\lambda(z_t - (r_t^b + 1)q_t)a_t + (r_t^b + 1)q_t a_t]$, and $q_t a_{t+1} = \beta z_t a_t$. Thus I obtain

$$q_t K_{t+1} = \beta\{\int_{z_{min}}^{\underline{z_t}} q_t(1 + r_t^d)dG(z_t) + \int_{\overline{z_t}}^{z_{max}} \lambda[(z_t - q_t(1 + r_t^b)] + q_t(r_t^b + 1)dG(z_t) + \int_{\underline{z_t}}^{\overline{z_t}} z_t dG(z_t)\}K_t$$

which is an equivalent formula of Equation (2.10). □

Proof for Lemma 2.4

Proof. The bellman equation for the banker i is:

$$V(N_{it}) = \max_{\{c_{it}^b, Q_{it}^L, Q_{it}^D\}} \{c_{it+1}^b + \beta V(N_{it+1})\}$$

subject to the balance sheet identity (2.4), the budget constraint (2.5), and the minimum capital requirement (2.6).

Under the assumption of uniform distribution of productivity, I obtain the first order condition with respect to deposits, loans and capital:

$$r_t^b + 1 = \frac{(\mu_{it} + \kappa\chi_{it})M\overline{z_t}}{(M + 1)\overline{z_t} - z_{max}} \tag{2.22}$$

$$r_t^d + 1 = \frac{\mu_{it} M \underline{z_t}}{(M + 1)\underline{z_t} - z_{min}} \tag{2.23}$$

$$q_t = \beta q_{t+1}(\mu_{it+1} + \chi_{it+1}) \tag{2.24}$$

$$\chi_{it+1}(N_{it+1} - \kappa Q_{it+1}^L) = 0 \tag{2.25}$$

where $q_t\mu_{it}$ is the multiplier of the bank's balance sheet identity. $q_t\chi_{it+1}$ is the multiplier of the capital constraint. Equation (2.25) is the complementary and slackness condition for the minimum capital requirement.

With the assumption $\kappa = 0$, the combination of Equations (2.22) and (2.23) with Lemma 2.1 leads to the conclusion that $\frac{\overline{z_t} - \underline{z_t}}{z_{max} - z_{min}} = \frac{1}{M+1}$, which implies that the fraction of autarky entrepreneurs is $\frac{1}{M+1}$. □

Proof for Proposition 2.5

Proof. In the context of the symmetric equilibrium, the bank capital constraint may or may not be binding in the steady state. To determine the solution in both cases, I use the Guess and Verify method.

Case 1. Let me first consider the scenario in which the bank capital constraint is binding. In this case, bankers fund loans solely with deposits. Denote the proportion of autarky entrepreneurs and borrowing entrepreneurs as u and v, respectively[6]. From Lemma 2.4, we have $u = \frac{1}{M+1}$, while Equation (2.4) implies that $v = \frac{M}{\lambda(M+1)}$.

Substituting the formula of u and v into the law of motion for aggregate capital (Equation (2.10)), I obtain

$$\frac{1}{\beta} - 1 - r^d = (r^d+1)\frac{\overline{z}-\underline{z}}{\underline{z}}v + \lambda(r^d+1)\frac{z_{max}-\overline{z}}{2\underline{z}}v + (r^d+1)\frac{\overline{z}-\underline{z}}{2\underline{z}}u$$

$$\begin{aligned}\Rightarrow \frac{1}{\beta}\underline{z} &= (r^d+1)[\underline{z} + (\frac{u}{2}+v)u + \frac{1}{2}\lambda v^2](z_{max}-z_{min})) \\ &= (r^d+1)(\underline{z} + [\frac{1}{2}\lambda(\frac{M}{\lambda(M+1)})^2 + \frac{1}{M+1}(\frac{1}{2(M+1)} + \frac{M}{\lambda(M+1)})](z_{max}-z_{min})]\end{aligned}$$

Therefore, I obtain

$$r^d + 1 = \frac{2\frac{1}{\beta}\lambda(M+1)^2\underline{z}}{[(2M^2+2M+1)\lambda - M^2]z_{max} + [(2M+1)\lambda + M^2]z_{min}} \tag{2.26}$$

$$r^b + \delta = (r^d+\delta)\frac{\overline{z}}{\underline{z}} \tag{2.27}$$

where $\overline{z} = z_{max} - \frac{M}{\lambda(M+1)}(z_{max}-z_{min})$ and $\underline{z} = z_{min} + \frac{(\lambda-1)M}{\lambda(M+1)}(z_{max}-z_{min})$.

Case 2. Now suppose that the bank capital constraint is non-binding. By Equation (2.25), $\chi = 0$. Plugging this into Equation (2.24), I obtain $\mu = \frac{1}{\beta}$. Then the deposit rate and loan rate become:

$$r^d + 1 = \frac{\frac{1}{\beta}M\underline{z}}{(M+1)\underline{z} - z_{min}}$$

$$r^b + 1 = \frac{\frac{1}{\beta}M\overline{z}}{(M+1)\overline{z} - z_{max}}$$

The only unknowns are $\underline{z}$ and $\overline{z}$. To obtain this, I substitute the above 2 equations into

[6]To simplify the notation, I eliminate all time indices since this proposition pertains to the steady state.

the law of motion of aggregate capital:

$$\frac{M}{M+1} - v = \frac{M}{M+1}v + \frac{M\lambda}{2}v^2 + \frac{M}{2(M+1)^2}$$

which takes the form of $av^2 + bv + c = 0$ with $a > 0$, $b > 0$ and $c < 0$. So there must be a positive root and negative root, and the positive one equals:

$$v = \frac{-(2M+1) + \sqrt{4M^2 + 4M + 1 + (2M^3 + M^2)\lambda}}{(M+1)M\lambda} \tag{2.28}$$

The formula of u and v will pin down $\overline{z}$ and $\underline{z}$

The subsequent step involves verifying and determining the conditions under which the equilibrium lies in distinct regions, referred to as region 2 (Case 1) and region 1 (Case 2). The critical factor that ascertains whether the capital constraint is binding is whether $\mu < \frac{1}{\beta}$ or $\mu = \frac{1}{\beta}$. Region 1 corresponds to a binding capital constraint, where $\mu < \frac{1}{\beta}$. By combining Equation (2.22) and (2.26), the following expression is obtained:

$$\lambda > \frac{M^2 + 4M + 2}{2M + 1} \equiv \lambda(M) \tag{2.29}$$

It is apparent that $\lambda(M)$ is a monotonically increasing function of M.

The final stage of the proof is to establish that when the capital constraint is non-binding, there exists a positive correlation between bank concentration and the bank capital to asset ratio. In particular, in region 1, the bank capital to asset ratio is determined as follows:

$$\begin{aligned} \frac{N}{N+D} = \frac{\lambda v - \frac{M}{M+1}}{(\lambda-1)v} &= \frac{\lambda}{\lambda-1}(1 - \frac{M^2}{-(2M+1) + \sqrt{4M^2 + 4M + 1 + (2M^3 + M^2)\lambda}}) \\ &= \frac{\lambda}{\lambda-1}(1 - \frac{\sqrt{1 + \lambda\frac{M^2}{2M+1}} + 1}{\lambda}) \end{aligned} \tag{2.30}$$

where the first two equalities follows by the equilibrium conditions in region 1, and the last equality is a straightforward transformation of the formula. It is evident that the bank capital to asset ratio is positively related to λ and negatively associated with M. □

Proof for Proposition 2.6

Proof. The net margin, which is define as difference between loan rate and deposit rate, is:

$$r_b - r_d = \begin{cases} \frac{\frac{M}{M+1}\frac{1}{\beta}(z_{max}-z_{min})}{(M-(M+1)v)z_{max}+(M+1)vz_{min}},\ where\ v = \frac{-(2M+1)+\sqrt{(2M+1)^2+(2M^3+M^2)\lambda}}{M(M+1)\lambda} & \text{In Region 1} \\ \frac{2\lambda(M+1)\frac{1}{\beta}(z_{max}-z_{min})}{[(2M^2+2M+1)\lambda-M^2]z_{max}+[(2M+1)\lambda+M^2]z_{min}} & \text{In Region 2} \end{cases}$$

The monotone relationship between the net margins and the bank concentration is straightforward in Region 2, so I will focus on the proof in the Region 1. In Region 1,

$$\begin{aligned} \frac{\partial(r^b - r^d)}{\partial M} &= -\frac{\frac{1}{\beta} - 1 + \delta)(z_{max} - z_{min})}{(M + 1 - \frac{(M+1)^2}{M}v(z_{max} - z_{min}))^2} \\ &\quad * [z_{max} - \frac{M^2 - 1}{M^2}v(z_{max} - z_{min}) - \frac{(M+1)^2}{M}\frac{\partial v}{\partial M}(z_{max} - z_{min})] \end{aligned}$$

Denote $z_{max} - z_{min} = \Delta z$ and plug v into the above equation, the second term in the bracket becomes:

$$\frac{M^2 - 1}{M^2}v\Delta z = \frac{(M-1)(2M+1)}{M((2M+1) + \sqrt{(2M+1)^2 + (2M^3 + M^2)\lambda}}\Delta z \leq \frac{M-1}{2M}\Delta z$$

The third term in the bracket becomes:

$$\begin{aligned} \frac{(M+1)^2}{M}\frac{\partial v}{\partial M}\Delta z &\leq \frac{2M^2 + 2M + 1 - \frac{\lambda+15}{\sqrt{3\lambda+9}}}{\lambda M^3}\Delta z \leq \frac{2M^2 + 2M - 3}{\lambda M^3}\Delta z \\ &\leq \frac{M+1}{2M}\frac{2(2M+1)(2M^2+2M-3)}{(M^2+4M+2)(M+1)M^2}\Delta z \leq \frac{M+1}{2M}\frac{6}{14}\Delta z \leq \frac{M+1}{2M}\Delta z \end{aligned}$$

where the first inequality follows that $\frac{(M^4+M^3-M)\lambda+4M^3+6M^2+4M+1}{\sqrt{(2M^3+M^2)\lambda+(2M+1)^2}}$ is increasing in M, the second inequality follows that the minimum of $\frac{\lambda+15}{\sqrt{3\lambda+9}}$ is realized at $\lambda = 9$, the third inequality follows that $\lambda > \frac{M^2+4M+2}{2M+1}$, and the last two inequalities follow that the minimum is realized at $M = 1$. So $\frac{\partial(r^b-r^d)}{\partial M} < 0$, that is, the net margin is an increasing function of bank concentration in Region 1. Since the net margin is a continuous function, it is an increasing function of bank concentration in both of the regions.

Output takes the form

$$\begin{aligned} Y &= \frac{1}{M+1}KE[z|\underline{z} \leq z \leq \overline{z}] + \lambda vKE[z|z \geq \overline{z}] \\ &= \frac{\overline{K}}{2}(2z_{max} - (\frac{\frac{1}{M+1} + v}{\lambda v(M+1) + 1} + v)(z_{max} - z_{min})) \\ &= \overline{K}(z_{max} - \frac{M - (M+1)v}{\lambda M(M+1)v + M}\Delta z) \end{aligned}$$

The derivative of Y with respect to M in Region 1 is:

$$\frac{\partial Y}{\partial M} = \Delta z\overline{K}\frac{(\lambda M^2-1)v+\lambda(M+1)^2v^2+M(M+1)(1+\lambda M)\frac{\partial v}{\partial M}}{(\lambda M(M+1)v+M)^2}$$

where $\frac{\partial v}{\partial M} = \frac{\lambda\{2M^2+2M+1-\frac{(M^4+M^3-M)\lambda+4M^3+6M^2+4M+1}{\sqrt{(2M^3+M^2)\lambda+(2M+1)^2}}\}}{(\lambda M(M+1))^2} > -\frac{M^3+M^2-1}{M(M+1)^2\sqrt{(2M^3+M^2)\lambda+(2M+1)^2}}$. Therefore

$$\frac{\partial Y}{\partial M} \geq C[\frac{\lambda(2M+1)M^3-(1+\lambda M)(M^3+M^2-1)}{(M+1)\sqrt{(2M^3+M^2)\lambda+(2M+1)^2}}+(M^2+M-1)v] > 0$$

when $M \geq 1$, where $C = \frac{\overline{K}\Delta z}{(\lambda M(M+1)v+M)^2}$. Output is therefore an increasing function of M in region 1. The positive correlation between M and Y is straightforward in Region 2. By continuity of Y, output is a decreasing function of the bank concentration. □

Proof for Proposition 2.7

Proof. Suppose that the central planner implements the capital allocation between the entrepreneurs and bankers by $N = \kappa_0 K$. Then $N = \frac{\kappa_0}{1+\kappa_0}\overline{K}$ and $K = \frac{1}{1+\kappa_0}\overline{K}$. Therefore, the output is represented as:

$$Y = \underbrace{u}_{\text{fraction of autarky entrepreneurs}} KE[z|\underline{z} \leq z \leq \overline{z}] + \lambda \underbrace{v}_{\text{fraction of borrowing entrepreneurs}} KE[z|z \geq \overline{z}]$$
$$= \frac{1}{2}\overline{K}\{\frac{1}{(1+M)(1+\kappa_0)}\underline{z} + (1-\frac{1}{(1+M)(1+\kappa_0)})z_{max} + \overline{z}\}$$

where $\overline{z} = z_{max} - (\frac{M}{\lambda(M+1)} + \frac{\kappa_0}{\lambda})(z_{max-z_{min}})$, and $\underline{z} = z_{min} + (\frac{(\lambda-1)M}{\lambda(M+1)} - \frac{\kappa_0}{\lambda})(z_{max-z_{min}})$. Plugging the formula of $\overline{z}$ and $\underline{z}$ into the equation of output, I solve out the first order condition with respect to κ_0 and obtain:

$$\kappa_0^* = \frac{\sqrt{\lambda-1}}{M+1} - 1$$

Since κ_0^* should be non-negative, $\kappa_0^* = Max\{\frac{\sqrt{\lambda-1}}{M+1} - 1, 0\}$. □

Proof for Proposition 2.8

Proof. When the capital constraint is not binding, the optimal capital ratio between the bank capital and entrepreneurs' capital is

$$\begin{aligned}\frac{N^*}{K^*} &= \lambda v - \frac{M}{M+1} \\ &= \frac{-(2M+1) + \sqrt{(2M+1)^2 + (2M^3+M^2)\lambda}}{M(M+1)} - \frac{M}{M+1} \\ &= \frac{\sqrt{(2M+1)^2 + (2M^3+M^2)\lambda}}{M(M+1)} - \frac{M+1}{M}\end{aligned} \tag{2.31}$$

where the first equality in the aforementioned equation is derived from the balance sheet identity of bankers, while the second equality is a result of the optimal condition for the proportion of autarky entrepreneurs. The final equality is obtained through straightforward algebraic manipulations. I will prove this proposition by Guess and Verify. Suppose that $\frac{N^*}{K^*} > \kappa_0^*$, then:

$$\begin{aligned}&\frac{\sqrt{(2M+1)^2 + (2M^3+M^2)\lambda}}{M(M+1)} - \frac{M+1}{M} > \frac{\sqrt{\lambda-1}}{M+1} - 1 \\ \Rightarrow &\sqrt{(2M+1)^2 + (2M^3+M^2)\lambda} > M\sqrt{\lambda-1} + (M+1)\end{aligned} \tag{2.32}$$

Equation (2.32) becomes $2M + 1 + \lambda M^2 > (M+1)\sqrt{\lambda - 1}$, which always holds because $2M^2(2M+1) > (M+1)^2$. □

Bibliography

Admati, A. R., DeMarzo, P. M., Hellwig, M., and Pfleiderer, P. (2010). Fallacies, irrelevant facts, and myths in the discussion of capital regulation: Why bank equity is not expensive. Technical report, Preprints of the Max Planck Institute for Research on Collective Goods.

Alfon, I., Argimon, I., and Bascuñana-Ambrós, P. (2004). *What determines how much capital is held by UK banks and building societies?* Financial Services Authority London.

Allen, F., Carletti, E., and Marquez, R. (2011). Credit market competition and capital regulation. *The Review of Financial Studies*, 24(4):983–1018.

Angeletos, G.-M. (2007). Uninsured idiosyncratic investment risk and aggregate saving. *Review of Economic dynamics*, 10(1):1–30.

Bartelsman, E., Haltiwanger, J., and Scarpetta, S. (2013). Cross-country differences in productivity: The role of allocation and selection. *American economic review*, 103(1):305–34.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Black, S. E. and Strahan, P. E. (2002). Entrepreneurship and bank credit availability. *The Journal of Finance*, 57(6):2807–2833.

Blum, J. and Hellwig, M. (1995). The macroeconomic implications of capital adequacy requirements for banks. *European economic review*, 39(3-4):739–749.

Bolton, P. and Freixas, X. (2006). Corporate finance and the monetary transmission mechanism. *The Review of Financial Studies*, 19(3):829–870.

Brealey, R. (2006). Basel ii: The route ahead or cul-de-sac? *Journal of Applied Corporate Finance*, 18(4):34–43.

Bresnahan, T. F. (1989). Chapter 17 empirical studies of industries with market power. volume 2 of *Handbook of Industrial Organization*, pages 1011–1057. Elsevier.

Brunnermeier, M. K. and Koby, Y. (2018). The reversal interest rate. Technical report, National Bureau of Economic Research.

Cavalcanti, T. V., Kaboski, J. P., Martins, B. S., and Santos, C. (2021). Dispersion in financing costs and development. Technical report, National Bureau of Economic Research.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, 113(1):1–45.

Corbae, D., D'Erasmo, P., et al. (2021). Capital buffers in a quantitative model of banking industry dynamics. *Econometrica*, 8.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The Rise of Market Power and the Macroeconomic Implications*. *The Quarterly Journal of Economics*, 135(2):561–644.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6):2437–71.

Diez, M. F., Leigh, M. D., and Tambunlertchai, S. (2018). *Global market power and its macroeconomic implications*. International Monetary Fund.

Drechsler, I., Savov, A., and Schnabl, P. (2017). The Deposits Channel of Monetary Policy*. *The Quarterly Journal of Economics*, 132(4):1819–1876.

Flannery, M. J. and Rangan, K. P. (2008). What caused the bank capital build-up of the 1990s? *Review of finance*, 12(2):391–429.

Gali, J. and Monacelli, T. (2005). Monetary policy and exchange rate volatility in a small open economy. *The Review of Economic Studies*, 72(3):707–734.

Gindling, T. H. and Newhouse, D. (2014). Self-employment in the developing world. *World development*, 56:313–331.

Gu, J. (2021). Financial intermediation and occupational choice. *Journal of Economic Dynamics and Control*, 133:104238.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448.

Jayaratne, J. and Strahan, P. E. (1996). The Finance-Growth Nexus: Evidence from Bank Branch Deregulation*. *The Quarterly Journal of Economics*, 111(3):639–670.

Joaquim, G., van Doornik, B. F. N., Ornelas, J. R., et al. (2019). *Bank competition, cost of credit and economic activity: evidence from Brazil*. Banco Central do Brasil.

Kiyotaki, N. (1998). Credit and business cycles. *The Japanese Economic Review*, 49(1):18–35.

Kiyotaki, N. and Moore, J. (2019). Liquidity, business cycles, and monetary policy. *Journal of Political Economy*, 127(6):2926–2966.

Li, J. (2019). Imperfect banking competition and financial stability. Technical report, Working Paper.

Miller, M. H. (1995). Do the m & m propositions apply to banks? *Journal of Banking & Finance*, 19(3-4):483–489.

Moll, B. (2014). Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, 104(10):3186–3221.

Repullo, R. (2004). Capital requirements, market power, and risk-taking in banking. *Journal of financial Intermediation*, 13(2):156–182.

Scharfstein, D. and Sunderam, A. (2016). Market power in mortgage lending and the transmission of monetary policy. *Unpublished working paper. Harvard University*, 2.

Ulate, M. (2021). Going negative at the zero lower bound: The effects of negative nominal interest rates. *American Economic Review*, 111(1):1–40.

Van den Heuvel, S. J. (2008). The welfare cost of bank capital requirements. *Journal of Monetary Economics*, 55(2):298–320.

Wang, Y., Whited, T. M., Wu, Y., and Xiao, K. (2020). Bank market power and monetary policy transmission: Evidence from a structural estimation. Working Paper 27258, National Bureau of Economic Research.

Chapter 3

Bernanke and Gertler Meets Kiyotaki and Moore

3.1 Introduction

Chapters 1 and 2 examine the long term effects of the banking sector on macroeconomic variables, with Chapter 1 emphasizing stability and Chapter 2 focusing on efficiency. While these aspects are crucial for comprehending the broader implications of the financial market, it is equally significant to comprehend the impact of the financial market on the dynamics and transition of the economy. Chapter 3 addresses this topic.

The prolonged period of sluggish growth that has persisted since the financial crisis of 2008 (Reinhart and Rogoff (2009); Hall (2015); Blecker (2016)) has resulted in a situation where both output growth and unemployment rates have remained at low levels. This circumstance has engendered a vigorous debate in the field of macroeconomics, which has been focused on examining how frictions in the financial sector impact the real economy (Cúrdia and Woodford (2016); Brunnermeier et al. (2012)). Specifically, Kiyotaki and Moore (1997) study how collateral constraints act as an amplifier in exacerbating financial difficulties. Similarly, Bernanke and Gertler (1989) demonstrate the countercyclical nature of agency costs, which also exacerbates the economic dynamics. This chapter reinvestigates the amplification mechanism by constructing a model that incorporates both collateral constraints and screening costs.

The model considers a general equilibrium economy featuring two agents, namely banks and firms. Both agents are assumed to be risk-neutral and possess the ability to use capital in the production process. However, they diverge with respect to three key dimensions. Firstly, firms are endowed with superior productivity relative to banks, with the latter being conceptualized as unproductive firms. Secondly, firms confront financial constraints that preclude unfettered access to loans from banks unless they are fully collateralized by real estate holdings. Lastly, firms exhibit a lower level of patience compared to banks.

The primary innovation of the model pertains to the differentiation of debt into two distinct categories: screened debt and unscreened debt. Banking institutions that opt to issue screened debt are expected to bear an associated screening cost. Correspondingly, firms are optimally positioned to determine the ratio between screened and unscreened capital, which can be utilized for borrowing purposes. While screened capital is assumed to be more readily pledgeable, firms are required to incur an additional transformation cost when amassing screened capital.

The model reveals that the presence of collateral constraints and screening costs creates a feedback mechanism that reinforces the economic dynamics. In particular, negative shocks have a cascading effect on firms' investment and net worth, which in turn leads to a decline in the demand for capital and further reductions in the net worth of firms. This leads to a decline in the price of capital, thereby amplifying the initial impulse responses. The persistence and amplification of the effect due to the feedback mechanism is consistent with the so-called "Kiyotaki and Moore effect".

In addition, the presence of screening costs and differences in the pledgeability of capital induce firms to strategically shift from unscreened to screened capital during periods of recession. This reallocation leads to an upward shift in the average screening cost, exacerbating the distortion in firms' investment by causing a disproportionate decline in the demand for unscreened capital. This mechanism is referred to as the "Bernanke and Gertler effect", which operates through a countercyclical average screening cost. The effect persists due to the collateral constraint.

The baseline model, which incorporates collateral constraints and screening costs, yields impulse responses that are relatively weaker compared to those predicted by the Kiyotaki and Moore (1997) model. Specifically, as the screening cost increases, the magnitude of the impulse responses diminishes. This can be attributed to the decline in the ratio between screened and unscreened capital in the steady state, resulting from higher screening costs. Consequently, the average collateral pledgeability decreases, attenuating the original Kiyotaki and Moore effect. Quantitatively, the combined effect of the attenuated Kiyotaki and Moore effect and the additional Bernanke and Gertler effect is smaller than the original

Kiyotaki and Moore effect alone.

The model's results align with empirical observations on the impact of information structures on both production and credit allocation during economic upswings and downturns. The model's predictions indicate that in the event of a positive productivity shock, the ratio between screened capital and unscreened capital decreases, whereas in the case of a negative productivity shock, screened capital is more desirable. This phenomenon has been observed in the real economy, where booms are frequently accompanied by a depletion of information (Dell'Ariccia and Marquez (2006); Bo Becker (2018)).

The imposition of screening costs on debt issuance has resulted in a protracted path to the steady state in the economy, compared with Kiyotaki and Moore (1997). This theoretical framework offers insights into the prolonged economic downturn following the 2008 financial crisis. The decision to issue screened debt is characterized by a tradeoff between the stabilization of the economy and the prolongation of recessions.

Related Literature

This model is motivated by the observation of the prolonged and amplified recession that followed the most recent economic crisis (Cerra and Saxena (2008); Ball (2014)). Furthermore, the link between information production and economic fluctuations has been established in the literature (Asea and Blomberg (1998); Gourinchas et al. (2001); Duprey (2016)). In particular, previous studies have shown that information tends to decay during credit booms, which could have significant implications for the stability and resilience of the financial system. Therefore, this chapter aims to further examine the role of screening and collateral constraint in the amplification and persistence of economic cycles, and to shed light on the underlying mechanisms of the observed recessionary behavior.

This academic work pertains to analytical models that examine the propagation and magnification of exogenous economic shocks over time. The seminal contribution of Kiyotaki and Moore (1997) highlights a key mechanism - the collateral constraint channel - which has sparked a vast body of macroeconomic literature focusing on collateral constraints. In their study, negative temporary shocks cause a reduction in the net worth of productive firms and their demand for capital, leading to a decline in capital prices. Concurrently, the reduction in productive firms' net worth and their demand for capital is exacerbated by the fall in capital prices. Bernanke and Gertler (1989), however, emphasize the role of agency costs, finding that reductions in collateral during a recession increase the agency costs of borrowing, which in turn depresses the demand for investment. The presence of these financial factors tends to amplify fluctuations in real output. Additionally, Bernanke et al. (1999) develop a quantitative model that demonstrates how credit market frictions can provide an explanation for typical business cycle fluctuations. They refer to this phenomenon as the "financial accelerator". The current model integrates both collateral constraints and screening costs to analyze how these two mechanisms amplify economic shocks.

This study is not the first theoretical investigation into the relationship between information production and economic cycles, as evidenced by previous works such as van Nieuwerburgh and Veldkamp (2006), Gorton and Ordoñez (2014), and Asriyan et al. (2021). In particular, Gorton and Ordoñez (2014) demonstrate that financial fragility builds up gradually as information about real estate deteriorates. A crisis occurs when a shock suddenly incentivizes agents to produce information, resulting in a decline in output. Asriyan et al. (2021) show that during credit booms fueled by high collateral values, economic activity expands while the economy's stock of information on existing projects diminishes. They examine the differences between production booms and credit booms. This study contributes to the literature by demonstrating that even a productivity shock can generate an amplified and persistent effect on the dynamics of the economy through information frictions. The amount of information, as indicated by the ratio of screened capital to unscreened capital in the model, in both steady-state and transitional periods affects the amplification of the impulse responses of the economy.

This work is also related to existing literature on the amplification mechanism. Cao and Nie (2017) suggest that market incompleteness, rather than collateral constraints, significantly contributes to the amplified and asymmetric responses of the economy to exogenous shocks. Mendoza (2010) compares the equilibrium under a collateral constraint to the equilibrium under an exogenous borrowing constraint limit and discovers that imposing an exogenous borrowing limit instead of a collateral constraint considerably weakens the amplification effect on asset prices. Contrary to Kiyotaki and Moore (1997), Cordoba and Ripoll (2004) find that collateral constraints typically result in small output amplification. This chapter contributes to the literature by simply expanding the model proposed by Kiyotaki and Moore (1997). I demonstrate that the amplification mechanism may not be significant because information costs are too high during recovery periods from recession.

A significant body of literature has explored the topic of lending standards during economic booms. Dell"Ariccia et al. (2008) have shown that the likelihood of booms is positively associated with a decrease in lending standards, and the extent of the decline depends mainly on the number of incumbent banks in the region. Similarly, Petriconi (2015) has demonstrated that credit screening is often inadequate during a rapid economic expansion. To contribute to the existing literature, this study reveals that in boom periods, the proportion of screened to unscreened capital declines, implying that capital becomes less pledgeable as collateral and consequently lending standards are tightened. This phenomenon is a significant contributor to the inherent instability and fragility of the financial system, and explains why booms inevitably end in severe recessions.

The remaining sections of this chapter are structured as follows. Section 3.2 provides an overview of the model's framework. In Section 3.3, we establish the equilibrium conditions of the model and examine the properties of its steady state. Section 3.4 investigates the dynamics of the economy and links them to the Kiyotaki and Moore effect, as well as the Bernanke and Gertler effect. Furthermore, Section 3.5 introduces extensions to the model. Finally, Section 3.6 summarizes the chapter.

3.2 Model Environment

Consider a discrete-time economy with an infinite horizon, where time is indexed by $t = 0, 1, 2, \cdots$. The economy features two types of agents: a continuum of infinitely-lived firms and m infinitely-lived banks. There are two types of goods in the economy: durable assets and nondurable assets. Nondurable assets represent consumption or dividend payouts that cannot be stored and are realized at the end of each period. Durable assets, on the other hand, are interpreted as real-estate values or capital, possess a fixed supply of $\bar{K}$, and do not depreciate for the sake of simplification. Both banks and firms engage in production and consumption, albeit with differing production technologies. There is no aggregate uncertainty in the economy.

The discount factors for firms and banks are β and $\tilde{\beta}$, respectively, both of which fall strictly within the range of 0 and 1. I make the following assumption (Kiyotaki and Moore (1997)).

Assumption 3.1. $\beta < \tilde{\beta}$

I will elaborate on this assumption further in later sections, wherein I will demonstrate that it guarantees that firms in equilibrium will not seek to defer consumption due to their relatively greater degree of impatience.

3.2.1 Banks

There are m banks who are characterized as homogeneous entities, each of which is risk-neutral and features an identical production technology that demonstrates decreasing returns to scale. There is a perfectly competitive bond market, where banks are not credit-constrained and function as debt issuers in the equilibrium. A distinctive feature of the model is the existence of two types of debt, namely, screened debt (denoted as $\tilde{b}^s_t$) and unscreened debt (denoted as $\tilde{b}^u_t$). When banks issue screened debt, they incur a screening cost represented as $c(\tilde{b}^s_t)$, where $c(\cdot)$ is assumed to be an increasing and convex function. Each banker maximizes its expected discounted sum of consumption, subject to the budget constraint

$$\sum_{t=0}^{\infty} \tilde{\beta}^t \tilde{c}_t$$

$$s.t.\ \tilde{c}_t + q_t\tilde{k}_t + \tilde{b}^s_t + c(\tilde{b}^s_t) + \tilde{b}^u_t \leq G(\tilde{k}_{t-1}) + q_t\tilde{k}_{t-1} + R^s_t\tilde{b}^s_{t-1} + R^u_t\tilde{b}^u_{t-1} \tag{3.1}$$

where $G(\cdot)$ denotes the production function of banks, satisfying the properties $G' > 0$ and $G'' < 0$. At period t, the bank is endowed with capital $\tilde{k}_{t-1}$ and promised debt repayments of $R^s_t\tilde{b}^s_{t-1} + R^u_t\tilde{b}^u_{t-1}$. After production in period t, the bank may allocate resources to consumption $\tilde{c}_t$, new debt issuance $\tilde{b}^s_t + c(\tilde{b}^s_t) + \tilde{b}^u_t$, and capital accumulation $\tilde{k}_t$ at a price of q_t. In the following section, it will be demonstrated that the imposition of screening costs will result in a discernible credit spread between debt instruments that have been subjected to screening, and those that have not.

3.2.2 Firms

Firms are considered to be homogeneous and risk-neutral, possessing an initial capital stock denoted by k_{t-1} at time t. The production function of the firms is specified as $F(k_t) = (a + c)k_{t-1}$, where ak_{t-1} represents pledgeable income and ck_{t-1} denotes a non-pledgeable component that cannot be traded in the market. The inclusion of the non-pledgeable component is motivated by the desire to prevent firms from continually postponing consumption due to their risk neutrality[1]. Firms' consumption is thus constrained by

$$c_t \geq ck_{t-1} \tag{3.2}$$

[1] It is equivalent to assume that firms have an i.i.d. probability of surviving until the next period, as discussed in Gertler and Kiyotaki (2015), where they consume upon exit.

The bond market is active, which can be attributed to the variances in production efficiency that distinguish firms from banks. However, the firms' borrowing capabilities in the market are restricted, such that all debt obligations must be collateralized against their durable assets, i.e., capital. Specifically, firms have the option of selecting between (un)screened capital, denoted by k_t^s (k_t^u), which can be pledged against (un)screened debt, represented by k_t^s (k_t^u), respectively. Screened capital is characterized by a higher degree of pledgeability compared to its unscreened counterpart. However, firms must incur a cost parameterized by κ for each unit of screened capital accumulated. As such, each firm is subjected to two collateral constraints, where the maximum amount of their debt repayment is determined by a fraction of their future collateral value.

$$R_{t+1}^s b_t^s \leq \theta_s q_{t+1} k_t^s \tag{3.3}$$

$$R_{t+1}^u b_t^u \leq \theta_u q_{t+1} k_t^u \tag{3.4}$$

where $k_t^s + k_t^u = k_t$.

Assumption 3.2. $\theta_s \geq \theta_u$

Assumption 3.2 guarantees that screened debt is more pledgeable in nature. Each firm maximizes its expected discounted sum of consumption, subject to the budget constraint

$$\sum_{t=0}^{\infty} \beta^t c_t$$

$$s.t.\ c_t + q_t(k_t^s + k_t^u) + \kappa k_t^s + R_t^s b_{t-1}^s + R_t^u b_{t-1}^u \leq (a+c)(k_{t-1}^s + k_{t-1}^u) + q_t(k_{t-1}^s + k_{t-1}^u) + b_t^s + b_t^u \tag{3.5}$$

and Equation 3.2, Equation 3.3, and Equation 3.4. As implied by Equation 3.5, firms engage in borrowing activities up to the limit of $b_t^s + b_t^u$, which are then used towards their production endeavors. The resulting resources are subsequently allocated to consumption c_t, capital accumulation $(q_t + \kappa)k_t^s + q_t k_t^u$, and repayment of the relevant debt contracts $R_t^s b_{t-1}^s + R_t^u b_{t-1}^u$.

3.3 Equilibrium Characterization

This section presents the model equilibrium, where I focus on the perfect foresight competitive equilibrium throughout the chapter. I will first discuss the steady state characterization.

3.3.1 Firms' Side

To characterize the equilibrium, I first consider firms' decisions. Equation 3.2 is equivalent to

$$c_t \geq c(k^s_{t-1}+k^s_{t-1}) \Leftrightarrow (a+q_t)(k^s_{t-1}+k^u_{t-1})+b^s_t+b^u_t-q_t(k^s_t+k^u_t)-\kappa k^s_t-R^s_t b^s_{t-1}-R^u_t b^u_{t-1} \geq 0, \tag{3.6}$$

because the budget constraint always binds. I refer Equation 3.6 as the non-pledgeablity condition. Lagrangian multipliers associated with Equation 3.3, Equation 3.4, and Equation 3.6 are denoted as μ_s, μ_u, and μ_c, respectively. The optimal conditions for b^s_t and b^u_t are then given by

$$1+\mu_c-\mu_s R^s_{t+1} = \beta R^s_{t+1}(1+\mu_c) \tag{3.7}$$

$$1+\mu_c-\mu_u R^u_{t+1} = \beta R^u_{t+1}(1+\mu_c) \tag{3.8}$$

Equation 3.7 and Equation 3.8 demonstrate that the marginal benefit of both types of debt accrues a return of $1+\mu_c$ in terms of consumption. The return may exceed unity, as borrowing further relaxes the non-pledgeability condition by μ_c (income effect). The marginal cost of (un)screened debt is that it tightens the collateral constraint by μ_s (μ_u) and raises the repayment in the next period. Simultaneously, the optimal conditions for k^s_t and k^u_t

$$-(1+\mu_c)(q_t+\kappa)+\mu_s\theta_s q_{t+1}+\beta(c+(1+\mu_c)(a+q_{t+1})) = 0 \tag{3.9}$$

$$-(1+\mu_c)q_t+\mu_u\theta_u q_{t+1}+\beta(c+(1+\mu_c)(a+q_{t+1})) = 0 \tag{3.10}$$

Equations 3.9 and 3.10 entail that marginal costs associated with acquiring screened and unscreened capital are $(1+\mu_c)(q_t+\kappa)$ and $(1+\mu_c)q_t$, respectively. The cost may surpass the prevailing capital price since procuring capital tightens the non-pledgeability condition by μ_c. The marginal benefit of (un)screened capital is that it relaxes the collateral constraint by μ_s (μ_u) and boosts production in the next period.

3.3.2 Banks' Side

To model banks' optimal choice, I assume that the function $c(\cdot)$ takes a quadratic form, where $c(\tilde{b}^s_t) = \frac{\phi}{2}\tilde{b}^{s\,2}_t$. ϕ here represents a positive parameter that reflects banks' efficiency in terms of screening technology. When $\phi = 0$, banks do not incur any costs in ascertaining firms' preference between k^s_t and k^u_t. Based on the functional form of screening cost, I obtain

$$1 = \tilde{\beta} R^u_{t+1} \tag{3.11}$$

$$1 + \phi \tilde{b}^s_t = \tilde{\beta} R^s_{t+1} \tag{3.12}$$

$$q_t - \tilde{\beta} q_{t+1} = \tilde{\beta} G'(\tilde{k}_t) \tag{3.13}$$

where Equation 3.11 and Equation 3.12 capture the conditions under which banks optimize their choice between screened and unscreened debt. In addition, Equation 3.13 represents the point at which the present value of the marginal product of capital ($\tilde{\beta} G'(\tilde{k}_t)$) is equal to the opportunity cost, or user cost, of holding capital $q_t - \tilde{\beta} q_{t+1}$.

Lemma 3.3. *Define credit spread d_{t+1} as $R^s_{t+1} - R^u_{t+1}$. d_{t+1} is an increasing function of screening efficiency ϕ.*

Lemma 3.3 demonstrates that a higher screening cost is associated with a higher credit spread. Although seemingly self-evident, this finding holds substantial significance in providing insight into how the dynamics of the economy differ based on the nature of the credit spread's response. I will show this later in the quantitative analysis.

3.3.3 Steady State Equilibrium

Definition 3.4. A *competitive equilibrium* consists of a sequence of firms' feasible allocation $\{c_t, k^s_t, k^u_t, b^s_t, b^u_t\}^{\infty}_{t=0}$, a sequence of banks' feasible allocation $\{\tilde{c}_t, \tilde{k}^s_t, \tilde{k}^u_t, \tilde{b}^s_t, \tilde{b}^u_t\}^{\infty}_{t=0}$ and a sequence of non-negative prices $\{R^s_{t+1}, R^u_{t+1}, q_t\}^{\infty}_{t=0}$, such that

- Given prices, the allocations maximize firms' and banks' life-time utility.
- Market clearing conditions for
  - Screened debt market: $b^s_t = \tilde{b}^s_t$
  - Unscreened debt market: $b^u_t = \tilde{b}^u_t$
  - Capital market: $k^s_t + k^u_t + m\tilde{k}_t = \overline{K}$

Proposition 3.5. *Both collateral constraints bind.*

Proof. The proof employs the Guess and Verify method and considers positive values of μ_s and μ_u. Using Equation 3.7 and Equation 3.8, it follows that

$$\frac{\mu_s}{\mu_u} = \frac{\frac{(1+\mu_c)(1-\beta R^s_{t+1})}{R^s_{t+1}}}{\frac{(1+\mu_c)(1-\beta R^u_{t+1})}{R^u_{t+1}}} = \frac{1 - \beta R^s_{t+1}}{1 - \beta R^u_{t+1}} \frac{R^u_{t+1}}{R^s_{t+1}} \tag{3.14}$$

Similarly, applying Equation 3.9 and Equation 3.10 leads to

$$\frac{1-\beta R^{s}_{t+1}}{R^{s}_{t+1}}\theta_s q_{t+1} - \frac{1-\beta R^{u}_{t+1}}{R^{u}_{t+1}}\theta_u q_{t+1} = \kappa \tag{3.15}$$

By combining Equation 3.14 and Equation 3.15, I obtain

$$R^{s}_{t+1} < \frac{1}{\frac{\theta_u}{\theta_s}\frac{1}{R^{u}_{t+1}} + (1-\frac{\theta_u}{\theta_s})\beta} \tag{3.16}$$

Moreover, under the assumptions of Equation 3.11 and Assumption 3.1, it follows that $R^{u}_{t+1} = \frac{1}{\tilde{\beta}} < \frac{1}{\beta}$. This implies that μ_u is positive. By substituting this result into Equation 3.16, it is shown that μ_s is also positive. Hence, the multipliers on the two collateral constraints are positive, indicating that the constraints are binding. □

Proposition 3.5 implies that firms will invariably seek to obtain the maximum possible borrowing amount. Furthermore, similar to Kiyotaki and Moore (1997), the non-pledgeability condition also binds in the steady state under the following assumption.

Assumption 3.6. $c > \frac{1}{\beta}\{\frac{1-\theta_u(\tilde{\beta}-\beta)-\beta}{(1-\tilde{\beta})\theta_u}(a+\kappa) - \beta a\}$

Assumption 3.6 is the sufficient,but not necessary condition for a binding non-pledgeability condition. This assumption is mild because β is close to 1. When $\theta_u = 1$ and $\kappa = 0$, the condition is the same as Kiyotaki and Moore (1997). When c is large enough, firms' saving rate is determined by share of non-pledgeable income. Therefore, $c_t = ck_{t-1}$.

The binding non-pledgeability condition and collateral constraints imply that:

$$k_t = \frac{1}{q_t + \kappa\frac{k^s_t}{k_t} - \theta_s\frac{k^s_t}{k_t}\frac{q_{t+1}}{R^s_{t+1}} - \theta_u\frac{k^u_t}{k_t}\frac{q_{t+1}}{R^u_{t+1}}}[(a+q_t)k_{t-1} - R^s_t b^s_{t-1} - R^u_t b^u_{t-1}] \tag{3.17}$$

The term $(a+q_t)k_{t-1} - R^s_t b^s_{t-1} - R^u_t b^u_{t-1}$ represents the net worth of firms at the start of date t. Equation 3.17 denotes that firms use all their net worth to finance capital expenditure. The user cost of the capital, which is the difference between the present value of the capital and the amount firms can borrow with each unit of capital, is given by $q_t + \kappa\frac{k^s_t}{k_t} - \theta_s\frac{k^s_t}{k_t}\frac{q_{t+1}}{R^s_{t+1}} - \theta_u\frac{k^u_t}{k_t}\frac{q_{t+1}}{R^u_{t+1}}$.

Proposition 3.7. *When $\theta_u = \theta_s = 1$, the model equilibrium is exactly the same as Kiyotaki and Moore (1997). All debt will be unscreened debt.*

The proof is straightforward. When $\theta_s = \theta_u$, the benefits for screened and unscreened debt are identical. Consequently, firms allocate all their capital to k^u_t during period t due to its lower cost. As a result, $b_t^s = k_t^s = 0$, and $R_s^* = R_u^* = \frac{1}{\tilde{\beta}} \equiv R^*$. In this case, the equilibrium conditions imply a unique steady state denoted by (q^*, k^*, b_u^*, R^*).

$$\frac{R^* - 1}{R^*} q^* = a \tag{3.18}$$

$$\frac{1}{R^*} G'(\frac{\overline{K} - k^*}{m}) = a \tag{3.19}$$

$$b_u^* = \frac{a}{R^* - 1} k^* \tag{3.20}$$

In the steady state, the income that can be pledged as collateral precisely covers the debt interest (Equation 3.20), while the user cost of capital equals the technology level (Equation 3.18). It is important to note that due to the collateral constraints, a first-best allocation of capital is not attainable in the economy. The efficient allocation is achieved when the marginal product of firms equals that of banks, i.e., $a + c = G'(\frac{\overline{K} - k^*}{m}) = aR^*$. Given the binding non-pledgeability condition, it is straightforward to deduce that in the steady state, $aR^* < a + c$, which implies that firms are expected to borrow more while being constrained by the friction.

3.4 Dynamics

In this section, I calibrate key parameters and present the dynamics of the economy in response to an unanticipated productivity shock. Assuming that the economy is in the steady state at time $t - 1$, an unforeseen shock in productivity impacts both firms and banks. I discuss how the impulse responses differ in the presence of collateral constraints and screening cost.

3.4.1 Calibration

One period in my model corresponds to one quarter. Following Gali and Monacelli (2005) and Christiano et al. (2005), the discount factor $\tilde{\beta}$ for banks is calibrated at 0.99, which implies a riskless annual rate of about 4% in the steady state. Firms are assumed to be less patient so that $\beta = 0.96$. Furthermore, banks' production is assumed to be a Cobb-Douglas function with inelastic labor supply, that is, $G(\tilde{k}_t) = \tilde{k}_t^{\alpha}$, with the capital ratio equal to 1/3, resulting in the specification of α as 0.33.

Calibrating the parameters that correspond to information friction is pivotal. θ_s is normalized to 1, as it ensures that the model's economic dynamics when both ϕ and κ are equal to zero are commensurate in magnitude to those of Kiyotaki and Moore (1997). Furthermore, the parameter θ_u is calibrated in a manner that aligns with the screened to unscreened debt ratio, with a specific value of 0.14. The screening cost is postulated to be 0.0001, while the screened capital adjustment cost is set at 0.05.

The aggregate capital supply $\overline{K}$, and productivity on pledgeable income a are normalized to 1. The parameter c has no effect on the dynamics of the economy (except for output) as long as it satisfies assumption 3.6. In this regard, I set $c = 2$, which results in a saving rate of 0.33% for firms. The calibration details of all the parameters are presented in Table 3.1.

Parameters	Values	Description
$\tilde{\beta}$	0.99	Risk-free interest rate*
β	0.96	Firms' lower patience
α	0.33	Capital ratio*
θ_s	1	Normalized to 1
θ_u	0.95	Screened/Unscreened Debt Ratio 0.14
a	1	Normalized to 1
c	2	saving rate of firm 33%
ϕ	0.0001	screening cost
κ	0.05	screened capital adjustment cost
$\overline{K}$	1	Normalized to 1

Calibrated Parameter Values

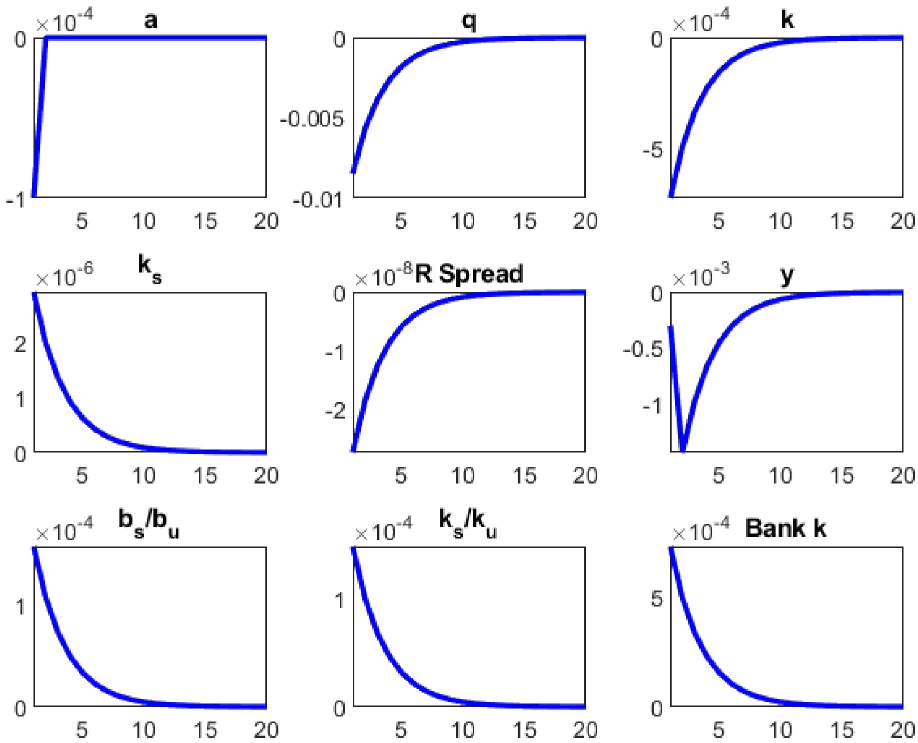
Table 3.1: * indicates that the parameter is calculated to match moments from data

3.4.2 Impulse Responses in the Baseline Model

In the baseline model, there are mainly two frictions: collateral constraint and screening cost, the effect of which are emphasized by Kiyotaki and Moore (1997) and Bernanke and Gertler (1989), respectively. Assuming a scenario where both banks and firms experience a reduction in productivity by 0.01%, the responses of endogenous variables to the aforementioned shock are observed in Figure 3.1.

Kiyotaki and Moore Effect

In the context of reduced productivity, credit constrained firms experience a decline in their net worth. Consequently, firms are compelled to reduce their total investment expenditures, namely capital accumulation k_t. This reduction causes a ripple effect, as firms earn less revenue in the subsequent period, leading to further reductions in their net worth.



Impulse Responses in the Baseline Model

Figure 3.1: The figure illustrates the dynamic response of endogenous variables, namely the capital price, firms' capital, firms' screened capital, credit spread, output, screened and unscreened debt ratio, screened and unscreened capital ratio, as well as banks' capital, to a negative productivity shock.

The collateral constraints continue to limit investment, creating persistent effects that reduce the constrained firms' demand for capital not only in period t but also in subsequent periods $t+1$, $t+2$, $\cdots$.

Following the capital market clearing condition, it is crucial for banks to increase their demand for capital in periods t, $t+1$, $t+2$, and so forth. To achieve this, they must reduce the opportunity cost of holding capital in these periods, given by $\{q_s - \tilde{\beta} q_{s+1}\}_{s=t}^{\infty}$. This anticipated decline in user costs in periods t, $t+1$, $t+2$, $\cdots$ is reflected by a fall in the capital price in period t because

$$q_t = \sum_{s=t}^{\infty} \tilde{\beta}^{s-t} (q_s - \tilde{\beta} q_{s+1}) \tag{3.21}$$

The aforementioned process further diminishes the net worth of firms during period t. Persistence and amplification reinforce each other. This explains the dynamics of q, k, and $\tilde{k}$ in Figure 3.1.

Bernanke and Gertler Effect

The graph presented in Figure 3.1 demonstrates a countercyclical trend in the average screening cost, denoted as b_s/b_u. This pattern can be attributed to the fact that a decrease in productivity results in a reduction in firms' net worth and capital prices, ultimately limiting their ability to pledge capital to borrow funds. Based on Assumption 3.2 and Equation 3.17, firms respond optimally to the decrease in productivity by adjusting their borrowing strategies, shifting towards a greater reliance on screened capital and lowering the user cost of capital.

Due to increased screening of capital and consequent higher average screening costs, banks experience lower net worth. To ensure that the capital market clears, banks further decrease the user cost of capital, which in turn exacerbates the decline in price of capital at period t. Consequently, the demand for capital by firms is further suppressed, with a particular more than one-to-one decrease in unscreened capital in response to an increase in screened capital. The amplification of economic dynamics can be attributed to the countercyclical behavior of the ratio between the levels of screened and unscreened borrowing, denoted as b_s/b_u and k_s/k_u.

Additionally, the introduction of positive κ results in a decrease in credit spreads during a recession, as the relative cost of screened capital for firms ($\frac{q+\kappa}{q}$) increases. These dynamics account for the fluctuations in k_s, b_s/b_u, k_s/k_u, and credit spreads over time.

The impulse response function provides insight into the relationship between the quality of capital, specifically its pledgeability, and the credit cycle. During economic downturns, there is an increase in banks' average screening costs, leading to a larger pool of pledgeable capital and improved information. This outcome is a consequence of firms' rational decision-making between screened and unscreened capital. Furthermore, the amplified effects during a recession are a result of increased expenditures on information acquisition during the recovery period.

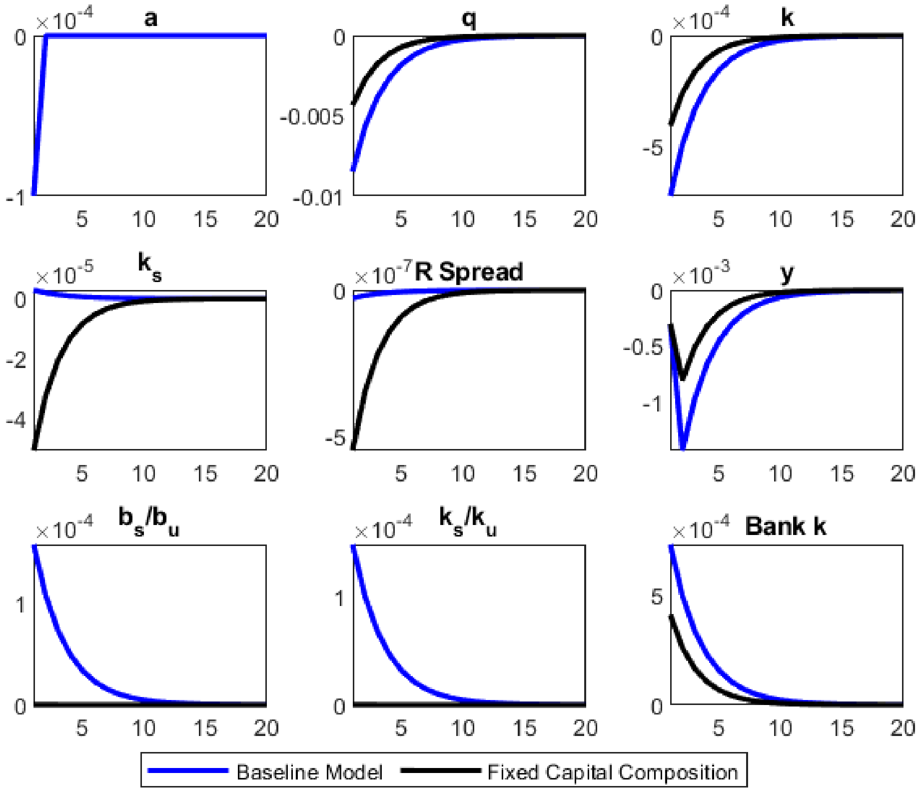
3.4.3 Fixed Capital Composition

This subsection presents the scenario that banks are unable to adjust between screened and unscreened capital, even when κ is small. Instead, the ratio between screened and unscreened capital is assumed to be fixed at the steady-state level in the baseline model, as given by:

$$\frac{k_t^s}{k_t^u} = \frac{k_s^*}{k_u^*} \tag{3.22}$$

A comparison between the impulse responses of the baseline model and the model with fixed capital composition is presented in Figure 3.2. The results indicate that allowing firms to adjust their capital decision leads to significantly stronger impulse responses. This comparison provides greater clarity regarding how an increase in average screening cost during a recession, due to firms' optimal selection between screened and unscreened capital, can amplify economic dynamics. Specifically, the higher average screening cost curbs firms' demand for capital, thereby further reducing the price of capital. This has a persistent effect through the collateral constraints.

As illustrated in Figure 3.2, the Bernanke and Gertler effect augments the impulse responses beyond those generated by the Kiyotaki and Moore effect. Notably, the persistence of the Bernanke and Gertler effect is also attributable to the presence of collateral constraints.



Impulse Responses in the Model with Fixed Capital Composition

Figure 3.2: The figure illustrates the dynamic response of endogenous variables, namely the capital price, firms' capital, firms' screened capital, credit spread, output, screened and unscreened debt ratio, screened and unscreened capital ratio, as well as banks' capital, to a negative productivity shock.

3.5 Discussions

In this section, I discuss how the magnitude of screening cost affects the dynamics of the economy. Moreover, a comparison between the baseline model and Kiyotaki and Moore (1997) is conducted.

3.5.1 Effects of Screening Cost

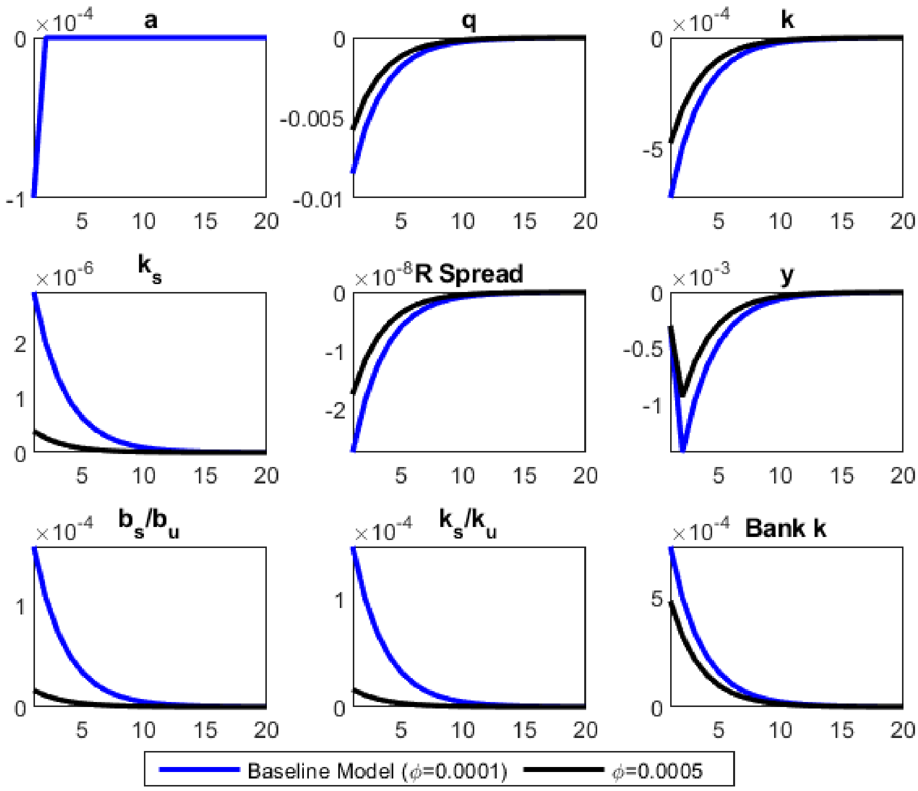
The key ingredient of the baseline model is that it incorporates two types of debt contracts. The extent of debt heterogeneity in the model is contingent on the coefficient ϕ, with larger values indicating higher transaction costs for banks when modifying screened debt. Notably, when ϕ is equal to zero, screened debt becomes identical to unscreened debt ($\kappa = 0$), and the model's dynamics replicate those of Kiyotaki and Moore (1997).

As presented in Figure 3.3, a comparison is made between the economic dynamics under two different values of the screening cost parameter, namely $\phi = 0.0001$ (the baseline model) and $\phi = 0.0005$. The results reveal that a lower screening cost leads to stronger impulse responses on output and price of capital. Conversely, when the screening cost is higher, there is an increase in credit spread, which, in turn, raises the cost of screened capital. A rise in screening costs results in a decrease in the steady-state proportion of screened to unscreened capital. Combining Equation 3.3 and 3.4, we obtain that

$$R^s_{t+1} b^s_t + R^u_{t+1} b^u_t \leq (\theta_s \frac{k^s_t}{k_t} + \theta_u \frac{k^u_t}{k_t}) q_{t+1} k_t. \tag{3.23}$$

Following the assumption that $\theta_s > \theta_u$, firms' capital is less pledgeable on average. This effect serves to attenuate the Kiyotaki and Moore effect, which operates through collateral constraints. Hence, the incorporation of screening costs introduces a new amplification effect while simultaneously mitigating the original effect. The quantitative magnitude of the mitigation in the Kiyotaki and Moore effect surpasses that of the Bernanke and Gertler effect.

It is worth noting that a higher screening cost implies a lower level of information in the steady state, as well as during the recovery phase from recession, which ultimately leads to a weaker impulse response. In the extreme scenario of $\kappa = 0$ and $\phi = 0$, the amplification effect is observed to be the strongest, which is shown in the next section.



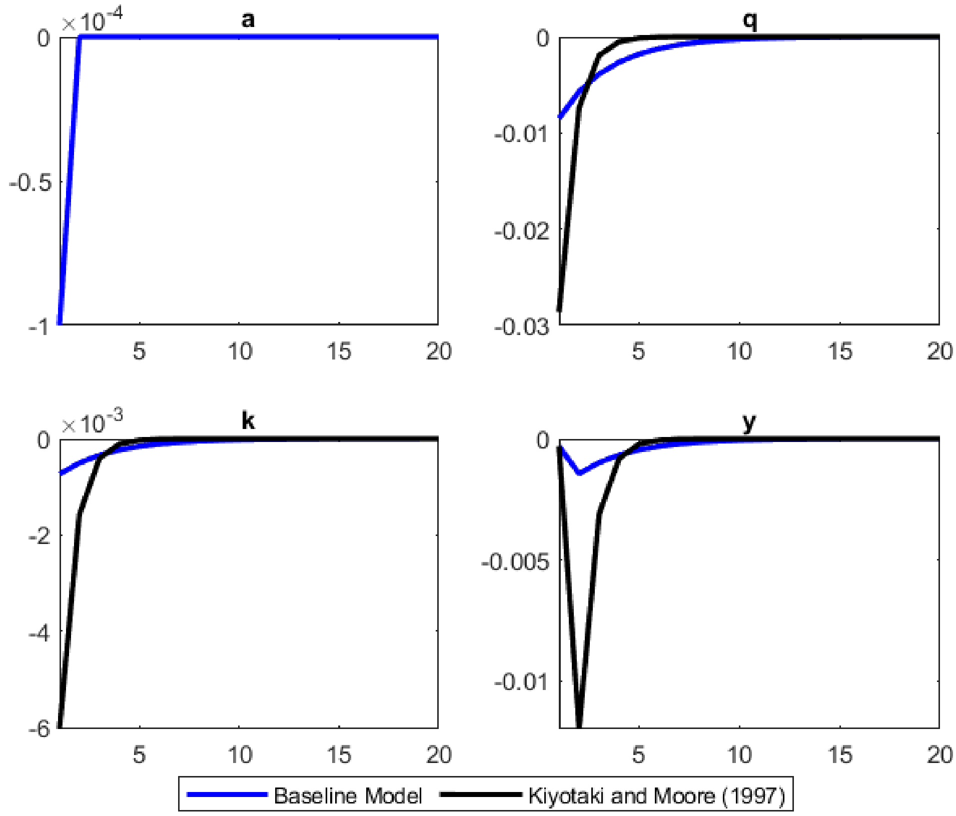
Impulse Responses in the Model with Different Screening Cost

Figure 3.3: The figure illustrates the dynamic response of endogenous variables, namely the capital price, firms' capital, firms' screened capital, credit spread, output, screened and unscreened debt ratio, screened and unscreened capital ratio, as well as banks' capital, to a negative productivity shock.

3.5.2 Comparison to Kiyotaki and Moore (1997)

In this section, I compare the impulse responses of the baseline model with that of Kiyotaki and Moore (1997). The results, illustrated in Figure 3.4, indicate that the impulse responses in the baseline model are significantly weaker than those in Kiyotaki and Moore (1997). This finding can be attributed to the introduction of positive screening cost, which mitigates the original Kiyotaki and Moore effect. The steady state exhibits more unscreened capital (debt), which in turn reduces the pledgeability of capital on average. In quantitative terms, the cumulative impact of the attenuated Kiyotaki and Moore effect, together with the additional Bernanke and Gertler effect, is smaller in magnitude compared to the original Kiyotaki and Moore effect on its own.

Figure 3.4 provides compelling evidence that the impulse response exhibits greater persistence compared to Kiyotaki and Moore (1997). This can be attributed to the fact that lower screening costs lead to an increase in overall pledgeability, thereby enabling shocks to transmit more seamlessly.



Impulse Responses in the Baseline Model and Kiyotaki and Moore (1997)

Figure 3.4: The figure illustrates the dynamic response of endogenous variables, namely the capital price, firms' capital, output, to a negative productivity shock.

3.6 Conclusion

This chapter represents a straightforward extension of the theoretical framework established by Kiyotaki and Moore (1997), whereby I introduce two types of debt and capital: screened and unscreened ones. In the model, firms optimally select the proportion of screened capital versus unscreened capital, both in steady state and during transitional periods. This decision-making process has important implications for the overall pledgeability of the economy. The primary findings of the chapter lies in demonstrating that unanticipated and temporary shocks are amplified and prolonged by virtue of collateral constraints, known as the Kiyotaki and Moore effect, and through countercyclical average screening costs, commonly referred to as the Bernanke and Gertler effect. However, it is worth noting that the degree of amplification is alleviated, while persistence is enhanced, as screening costs increase.

The model provides valuable insights into the dynamics of information and credit cycles. Moreover, it is worth noting that the baseline model can be extended to investigate the impact of monetary policy on the credit cycle by considering the effects of collateral constraints and screening costs. This avenue of inquiry has the potential to yield significant contributions to our understanding of the interplay between monetary policy, credit markets, and macroeconomic performance.

Bibliography

Asea, P. K. and Blomberg, B. (1998). Lending cycles. *Journal of Econometrics*, 83(1):89 – 128.

Asriyan, V., Laeven, L., and Martín, A. (2021). Collateral Booms and Information Depletion. *The Review of Economic Studies*, 89(2):517–555.

Ball, L. (2014). Long-term damage from the great recession in oecd countries. *European Journal of Economics and Economic Policies: Intervention*, 11(2):149–160.

Bernanke, B. and Gertler, M. (1989). Agency costs, net worth, and business fluctuations. *The American Economic Review*, 79(1):14–31.

Bernanke, B. S., Gertler, M., and Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. *Handbook of macroeconomics*, 1:1341–1393.

Blecker, R. (2016). The us economy since the crisis: slow recovery and secular stagnation. *European Journal of Economics and Economic Policies: Intervention*, 13(2):203–214.

Bo Becker, Marieke Bos, K. R. (2018). Bad times, good credit. *Swedish House of Finance Research Paper.*

Brunnermeier, M. K., Eisenbach, T. M., and Sannikov, Y. (2012). Macroeconomics with financial frictions: A survey. *Working Paper.*

Cao, D. and Nie, G. (2017). Amplification and asymmetric effects without collateral constraints. *American Economic Journal: Macroeconomics*, 9(3):222–66.

Cerra, V. and Saxena, S. C. (2008). Growth dynamics: The myth of economic recovery. *American Economic Review*, 98(1):439–457.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, 113(1):1–45.

Cordoba, J.-C. and Ripoll, M. (2004). Credit cycles redux. *International Economic Review*, 45(4):1011–1046.

Cúrdia, V. and Woodford, M. (2016). Credit frictions and optimal monetary policy. *Journal of Monetary Economics*, 84:30 – 65.

Dell"Ariccia, G., Igan, D., and Laeven, L. (2008). Credit booms and lending standards: Evidence from the subprime mortgage market. *C.E.P.R. Discussion Papers, CEPR Discussion Papers*, 44.

Dell'Ariccia, G. and Marquez, R. (2006). Lending booms and lending standards. *The Journal of Finance*, 61(5):2511–2546.

Duprey, T. (2016). Bank screening heterogeneity. Bank of Canada Staff Working Paper 2016-56, Bank of Canada, Ottawa.

Gali, J. and Monacelli, T. (2005). Monetary policy and exchange rate volatility in a small open economy. *The Review of Economic Studies*, 72(3):707–734.

Gertler, M. and Kiyotaki, N. (2015). Banking, liquidity, and bank runs in an infinite horizon economy. *American Economic Review*, 105(7):2011–43.

Gorton, G. and Ordoñez, G. (2014). Collateral crises. *American Economic Review*, 104(2):343–78.

Gourinchas, P.-O., Valdes, R., and Landerretche, O. (2001). Lending booms: Latin america and the world. *Economía Journal*, 0(Spring 20):47–100.

Hall, R. E. (2015). Quantifying the lasting harm to the us economy from the financial crisis. *NBER Macroeconomics Annual*, 29(1):71–128.

Kiyotaki, N. and Moore, J. (1997). Credit cycles. *Journal of Political Economy*, 195(2):211–248.

Mendoza, E. G. (2010). Sudden stops, financial crises, and leverage. *American Economic Review*, 100(5):1941–66.

Nieuwerburgh, S. V. and Veldkamp, L. (2006). Learning asymmetries in real business cycles. *Journal of Monetary Economics*, 53(4):753–772.

Petriconi, S. (2015). Bank competition, information choice and inefficient lending booms. *Working Paper*.

Reinhart, C. M. and Rogoff, K. S. (2009). The aftermath of financial crises. *American Economic Review: Papers Proceedings*, 99(2):466–472.