Breaking Down the Relationship Between Bank Concentration and Risks: The Role of Bank Capital

Yu Yi*

March 31, 2023

Click Here for the Latest Version

Abstract

How does bank capital affect the relationship between bank concentration and risk taking? I develop a tractable dynamic model with 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 entrepreneurs' 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 the loan rate. The two mechanisms also jointly establish a non-monotonic relationship between bank concentration and allocative efficiency. Two pieces of micro-level evidence in U.S. support the model predictions: first, the relationship between bank concentration and loan rate is non-monotonic; second, the effect of bank concentration on loan rate is positive only when the bank capital ratio is low. I discuss how efficiency and stability can be enhanced simultaneously.

Keywords: Bank Concentration, Bank Capital, Risk Shifting, Risk Taking, Net Margin

1 Introduction

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 policy-makers 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 both efficiency and stability. Conversely, proponents of a negative concentration-risk relationship emphasize the trade-off between efficiency and stability, highlighting concerns about a heavily concentrated banking sector and the need for alternative policy instruments to enhance efficiency while mitigating economic risk.

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

---

[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.

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.

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. Our model's

2

predictions are supported by empirical evidence, which reveals a negative correlation between bank concentration and loan rate at Herfindahl-Hirschman Index (HHI) levels of approximately 0.7.

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 both efficiency and stability when bankers accumulate excess capital above the required minimum.

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 risk taking. 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 safe when the bank capital constraint is binding. In this scenario, the reduction in bank concentration is expected to increase efficiency without significantly affecting stability when the bank capital ratio is close to the minimum requirement. However, when the capital ratio exceeds the minimum requirement, enhancing both efficiency and stability 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 commonly held view 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 improve stability 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 and D'Erasmo (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 and D'Erasmo (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 paper is organized as follows. In section 2, I lay out the model environment. Section 3 characterizes the symmetric model equilibrium and discusses the implication

of the *risk shifting mechanism* and *net margin mechanism*. Section 4 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 5, I present micro-data evidence on the relationship between bank concentration and loan rate, which supports *risk shifting mechanism* in the model. Additionally, policy implications are presented. Finally, Section 6 concludes the paper. The appendix contains the proofs and robustness checks.

2   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. During each period, bankers intermediate resources among a continuum of ex-ante heterogeneous entrepreneurs, while capital suppliers provide capital to both bankers and entrepreneurs.

2.1   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 $G(z)$ that 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 $\alpha > 1$ *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 opt to lend. It is assumed that borrowers are unable to make a commitment, while lenders are unable to

enforce their promises. To enable bankers to operate as financial intermediaries, bankers are endowed with the ability to commit and enforce. Entrepreneurs may receive external financing from the bankers, and are required to repay their debt at a fixed loan rate, denoted by r_t^b, upon successful completion of their projects. Additionally, entrepreneurs may deposit their resources in banks and earn a fixed deposit rate, denoted by r_t^d. Upon production and trading in the loan and deposit markets, entrepreneurs will give birth to offspring. Entrepreneurs of this generation will consume a certain percentage (s) of their net returns and invest the remainder in capital. This capital will then be equally distributed among the next generation of entrepreneurs. It is noteworthy that, unlike Moll (2014), there is no heterogeneity of wealth among entrepreneurs of the same generation. While this homogeneity of wealth is not a necessary assumption, it ensures that every entrepreneur receives a non-zero endowment, even in cases where their parents leave no inheritance.

Additionally, entrepreneurs face a borrowing constraint

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

Finite λ implies an imperfect financial market, reflecting the notion that entrepreneurs are constrained by their initial endowment when borrowing. The parameter λ measures the effectiveness of the financial market, with a value of 1 indicating a complete shutdown of the credit market, leading to all entrepreneurs being financially inactive. Conversely, as λ approaches infinity, the financial market becomes perfect. I denote $\theta_t = \frac{k_t}{a_t}$, where θ_t represents the entrepreneurs' actual leverage ratio. Accordingly, entrepreneurs' decisions can be characterized using θ_t and p.

2.2 Bankers

The primary assumption governing the banking sector is imperfect competition. To characterize this, I assume there are $1, 2, \cdots, M$ long-lived bankers in the economy, each of whom competes for the quantity of loans Q_{it}^L and deposits Q_{it}^D à la Cournot[2]. When $M = 1$, the economy is composed of a single monopoly bank, whereas as M approaches infinity, the banking sector tends towards perfect competition. 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

---

[2]Imperfect bank competition is modeled in accordance with Van Hoose et al. (2010) using Cournot competition. It is a simple yet effective approach to exploring the banking sector between the extremes of perfectly competitive banking ($M = \infty$) and monopoly banking ($M = 1$). Under Cournot competition, the extreme scenarios are analogous to those obtained through the application of Bertrand competition (i.e., monopoly banking with $M = 1$ and perfectly competitive banking with $M > 1$). Nevertheless, additional frictions may be necessary to create an intermediate market structure with price competition.

payouts

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

Bankers act as financial intermediaries by facilitating the borrowing and lending activities among entrepreneurs. Using equity capital and deposits, bankers provide loan contracts that can be either safe or risky, with the proportion of risky loans being represented by v_{rt}. The accumulation of bank equity capital is achieved solely through retained earnings[3]. Balance sheet identity of banker i then follows

$$Q_{it}^L = Q_{it}^D + N_{it} \tag{2}$$

The balance sheet items at the beginning of period t are summarized in Table 1. I assume that each banker is able to fully diversify the idiosyncratic risk and analyze the equilibrium in regions where no bankers default on deposits. At the end of period t, bankers' dividend payouts and retained earnings are financed by the returns from their activities in the loan and deposit markets. The intratemporal decision is reduced to a standard consumption and saving problem, where banker i must adhere to a budget constraint.

$$c_i^b + q_t N_{it+1} \leq (1 + r_t^b)q_t(1 - v_{rt})Q_{it}^L + p(1 + r_t^b)q_t v_{rt}Q_{it}^L - (1 + r_t^d)q_t Q_{it}^D \tag{3}$$

The right-hand side terms in equation (3) represent the income obtained by banker i from issuing safe and risky loan contracts, minus the repayment to depositors, which is used to finance the left-hand side variables— consumption and the accumulation of bank capital. The price of capital is denoted as q_t. To simplify equation (3), I define

$$p_t^e = (1 - v_{rt}) \cdot 1 + v_{rt} \cdot p, \tag{4}$$

which captures the expected probability of loan repayment, with the weights reflecting the respective proportion of safe and risky loans. Equation (3) then becomes

$$c_i^b + q_t N_{it+1} \leq q_t\{(1 + r_t^b)p_t^e Q_{it}^L - (1 + r_t^d)Q_{it}^D\}. \tag{3'}$$

Asymmetric information between entrepreneurs and bankers is a prominent feature in the financial market. Bankers lack information regarding the types of entrepreneurs, including their productivity and project choices, leading to the implementation of an identical loan rate that applies to the entire population of borrowing entrepreneurs. Nevertheless, the minimum capital

---

[3]The possibility of equity issuance by new investors does not affect the underlying mechanisms of the model.

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: Banker's Balance Sheet

requirement constrains the maximum amount of loans that bankers can issue.

$$N_{it} \geq \kappa Q_{it}^L \tag{5}$$

where κ measures the flexibility of the minimum capital requirement. This requirement mandates that a fraction κ of bank loans must be funded by capital. The Basel Committee on Banking Supervision implemented the initial framework for the minimum capital requirement in the late 1990s to regulate market risk. This constraint was imposed to ensure that banks maintained adequate regulatory capital to absorb financial losses. In this context, the capital constraint is represented by a simplified minimum requirement over capital to asset ratio. The appendix demonstrates that integrating the capital to risk-weighted asset ratio does not modify the principal mechanism of the model.

2.3 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. Capital suppliers are assumed to lack storage technology so they rationally choose to be hand-to-mouth.

3 Equilibrium Characterization

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

3.1 Entrepreneurs' Side

This section first derives the equilibrium conditions under which certain entrepreneurs choose to invest in gambling projects. These entrepreneurs seek to maximize their expected consumption and are motivated to borrow funds to enjoy the benefits of limited liability. Alternatively, they may invest in the prudent project that promises a higher expected return. Given

the linearity of the production function, borrowing entrepreneurs who gamble will always borrow up to the borrowing limit. Consequently, the incentive compatibility condition requires that gambling borrowers obtain a higher expected return than if they were to self-finance[4]:

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

With the probability of p, borrowing entrepreneurs who gamble receive a positive net return from their investment. Due to their higher leverage ratio compared to self-financing entrepreneurs, there exists a lower bound on productivity, which is denoted by z_2:

$$z \geq \frac{(\lambda - 1)p}{\lambda\alpha p - 1}q(1 + r_b) \equiv z_2 \tag{6}$$

Moreover, the benefits of borrowing entrepreneurs who invest in gambling projects should exceed those of borrowing entrepreneurs who invest in the prudent project:

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

Investing in the prudent project may be desirable due to its higher expected return ($\alpha p < 1$). Conversely, gambling may be preferred since borrowers who choose to gamble face a probability of repayment less than 1. Due to the inefficiency of the gambling project, the incentive compatibility condition yields an upper bound on productivity, which is denoted by z_3:

$$z \leq \frac{(\lambda - 1)(1 - p)}{\lambda(1 - \alpha p)}q(1 + r_b) \equiv z_3 \tag{7}$$

When both incentive compatibility conditions are satisfied, borrowing entrepreneurs will invest in the gambling project. The two conditions hold simultaneously in the equilibrium if and only if $z_2 < z_3$, which leads to the following assumption:

Assumption 2 $\frac{(\lambda-1)p}{\lambda\alpha p-1} < 1$.

Assumption 2 posits that entrepreneurs with productivity between z_2 and z_3 choose to invest in the gambling project. This assumption is dependent on three parameters, namely α, p, and λ. The likelihood of the inequality in Assumption 2 being upheld increases when the parameters α, λ, or p exhibit larger values. Intuitively, the benefits of gambling are more pronounced when there is a higher excess return (α) or a higher success probability (p). When asset pledgeability (λ) is higher, the heterogeneity among borrowing entrepreneurs decreases, enabling bankers to extract greater profits from borrowers and inducing them to gamble.

---

[4]For simplification purposes, the time index is neglected in this discussion.

Given the deposit and loan rate, Proposition 1 fully characterizes the financial decisions of entrepreneurs.

Proposition 1 *There are three productivity cutoffs z_1, z_2 and $\overline{z_3}$, such that*

- *The capital demand for individual entrepreneur is:*

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

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

Moreover, $z_1 = q(r^d + 1)$, $\overline{z_3} = Min\{z_3, z_{max}\}$.

The cutoff property relies heavily on the constant return to scale of the production function. According to Proposition 1, the optimal capital demand decision is at the corners. Specifically, entrepreneurs with productivity levels lower than z_1 will demand zero capital, whereas those with productivity levels greater than z_2 will demand the maximum amount allowed by their borrowing constraint. Entrepreneurs with productivity levels between z_1 and z_2 will demand an amount equal to their initial wealth. Capital demand distinguishes two types of marginal entrepreneurs. For those with productivity level z_1, the return on each additional unit of capital $\frac{z}{q}$ is equal to the opportunity cost of not depositing the capital in the bank, $r^d + 1$. Conversely, entrepreneurs with productivity level z_2 are indifferent between using external finance to gamble and using their own capital to produce. According to Assumption 2, $z_2 < q(1 + r^b)$, indicating that borrowing entrepreneurs with productivity level z_2 will not invest in the prudent project.

Entrepreneurs whose productivity levels exceed z_2 optimally choose between the prudent and gambling projects. While investing in the prudent project is desirable due to its higher expected return, it does not provide limited liability protection. Proposition 1 suggests that entrepreneurs with productivity levels above $\overline{z_3}$ will invest in the prudent project, while those with productivity levels in the range $z_2 < z < \overline{z_3}$ will gamble. Given that bankers are unable to observe entrepreneurs' productivity types and project choices, they set the same loan rate for all borrowing entrepreneurs. Entrepreneurs whose productivity levels exceed $\overline{z_3}$ receive a higher proportion of the net return from the loan contract and will invest in the project with a higher expected return. Conversely, entrepreneurs with productivity levels in the range $z_2 < z < \overline{z_3}$ will gamble, as they may not gain significant benefits from the loan issuance but rather benefit from limited liability. An extreme case is that those with productivity $q(1 + r^b)$ obtain no return
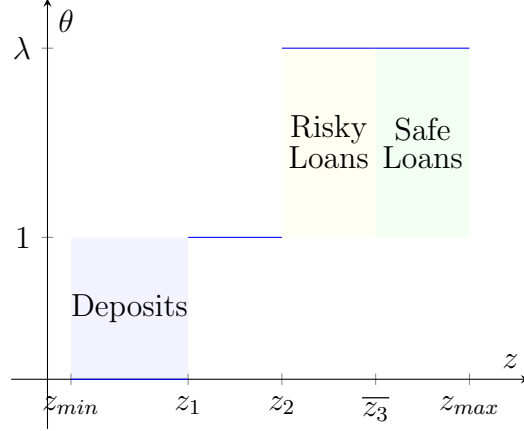
Figure 1: Entrepreneurs' Leverage and Project Choice

from the loan contract if they invest in the product project but receive a positive return if they gamble and the project succeeds.

Therefore, entrepreneurs with productivity levels below z_1 are termed as lending entrepreneurs, those with productivity levels above z_2 are referred to as borrowing entrepreneurs, and those with productivity levels in between are known as autarky entrepreneurs. The productivity of lending entrepreneurs is so low that it is not profitable for them to produce and, therefore, deposit all their endowment in banks. On the other hand, borrowing entrepreneurs, with their high productivity levels, are willing to borrow up to the borrowing limit. The presence of imperfect competition in the banking sector leads to the emergence of a third type of entrepreneur. Typically, banks charge a positive net margin between their loan and deposit rates, which discourages some entrepreneurs from either borrowing or lending. Such entrepreneurs with productivity levels between z_1 and z_2 are denoted as autarky entrepreneurs. Since autarky entrepreneurs use their own funds to produce, 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. Loan contracts could either be risky or safe, depending on the borrower's investment preferences. In the extreme scenario where $\overline{z_3} > z_{max}$, all loans are risky. In the next section, I will discuss how the equilibrium behaves when there are only risky loans, as well as when both risky and safe loans are present. It is noteworthy that entrepreneurs of each generation are endowed with identical initial wealth, leading to the demand for aggregate entrepreneurial capital being equivalent to the individual demand.

Lemma 1 *Denote Q_t^L and Q_t^D as the loan size and deposit size respectively. Aggregate quan-*

tities $\{Q_t^L, Q_t^D, a_{t+1}\}$ *satisfy:*

$$Q_t^L = (1 - G(\overline{z_t}))(\lambda - 1)a_t \tag{8}$$

$$Q_t^D = G(\underline{z_t})a_t \tag{9}$$

$$
q_t a_{t+1} = s\{\int_{z_{min}}^{z_{1t}} q_t(1 + r_t^d)dG(z_t) + \int_{\overline{z_{3t}}}^{z_{max}} \lambda[(z_t - q_t(1 + r_t^b)] + q_t(r_t^b + 1))dG(z_t) \\
+ \int_{z_{1t}}^{z_{2t}} z_t dG(z_t) + p\int_{z_{2t}}^{\overline{z_{3t}}} \alpha\lambda z_t - (\lambda - 1)q_t(r_t^b + 1)dG(z_t)\}a_t
\tag{10}
$$

The equation (8) reveals that the aggregate loan demand is determined by three factors: the percentage of borrowing entrepreneurs, the magnitude of borrowing per entrepreneur, and entrepreneurs' initial capital holdings. Correspondingly, the deposit supply is determined by the total initial capital of the lending entrepreneurs, as depicted by equation (9). Furthermore, the equation (10) captures the law of motion for the aggregate entrepreneurial capital demand, wherein the net wealth of the entrepreneurs in the next generation $q_t a_{t+1}$ relies on their saving rate s and net return. In precise terms, the net return is a weighted average of net return realized by depositors, borrowing entrepreneurs who invest in the prudent project, autarky entrepreneurs, and borrowing entrepreneurs who invest in the gambling project, respectively.

3.2 Bankers' Side

Banker i's optimization over quantities yields the optimal loan and deposit rates as a function of the mark-up (-down) on banker i's marginal cost (benefits)

$$1 + r^d + Q_i^D \frac{\partial r^d}{\partial Q_i^D} = \mu_i \tag{11}$$

$$p_e[(1 + r^b) + Q_i^L \frac{\partial r^b}{\partial Q_i^L}] + (1 + r^b)Q_i^L \frac{\partial p_e}{\partial r^b}\frac{\partial r^b}{\partial Q_i^L} = \mu_i + \kappa\chi_i \tag{12}$$

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 (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 (12), the marginal cost of issuing loans is a tightening of both the balance sheet identity and bank capital constraint, scaled by κ. In contrast, the marginal benefit of issuing loans is determined by the elasticity of loan demand and the response of the expected probability of loan repayment

to the loan size.

Proposition 2 (*Risk Shifting Mechanism*) *Assume $\frac{zg(z)}{1-G(z)}$ is increasing,*

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

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

According to Proposition 2, in partial equilibrium, an increase in the loan rate results in a higher proportion of risky loans and a lower expected probability of loan repayment. Specifically, when the loan rate rises, borrowing entrepreneurs experience a decrease in their net return from the loan contract. Consequently, they are motivated to engage in risky investments, resulting in a higher proportion of risky loans. The underlying *risk shifting mechanism* closely resembles that outlined in Boyd and De Nicolo (2005), albeit the model in this paper offers a micro-foundation based on the entrepreneurial perspective, instead of relying solely on the functional assumption of project returns. However, when $\overline{z_3} = z_{max}$, all loans become risky, $v_r = 1$, $p_e = p$, and the *risk shifting mechanism* is shut down. As I will demonstrate in the following section, the relationship between bank concentration and risks is not solely determined by the *risk shifting mechanism*.

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. Equation (11) and (12) then become:

$$1 + r^d = \frac{\epsilon^d}{\epsilon^d + s_i^d}\mu_i \tag{13}$$

$$p_e(1 + r^b) = \frac{\epsilon^b}{\epsilon^b - s_i^b[1 + \frac{\partial log p_e}{\partial(1+r^b)}]}(\mu_i + \kappa\chi_i) \tag{14}$$

The aforementioned equations reveal that the optimal loan (deposit) rate represents a mark-up (-down) over the marginal cost (benefit) of issuing the loan (deposit). The markup or markdown on the marginal cost or benefit is determined by the banker's share of the loan or deposit market. In a perfectly competitive banking sector, where s_i^b and s_i^d converge to zero, there will be no mark-up (-down). However, the *risk shifting mechanism* introduces a new term to Equation (14), which decreases the markup. This is because bankers recognize that setting a high loan rate encourages entrepreneurs to engage in gambling behavior. As the bank concentration increases, the incentive for bankers to raise loan rates might be diminished.

14

The optimal condition for bank capital yields

$$q_t = \beta q_{t+1}(\mu_{it+1} + \chi_{it+1}), \tag{15}$$

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

3.3 Steady State Equilibrium

In this section, I will turn to the general equilibrium, with a focus on the symmetric equilibrium throughout the paper.

Definition 1 (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 $\{a_{t+1}, Q_t^D, Q_t^L\}_{t=0}^{\infty}$, 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:*

- *Entrepreneurs and bankers maximize expected life-time utility given prices and interest rates;*

- *Bankers choose the same quantities for all assets and liabilities;*

- *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: $\sum_{i=1}^{M} N_{it} + a_t = \overline{K}$.*

The notion of symmetry alludes to the absence of heterogeneity among bankers in the equilibrium. To begin with, it would be insightful to explore how the symmetric equilibrium in a perfectly competitive banking sector deviates from the benchmark scenario, where no risk taking behaviors are involved, i.e., when $\alpha = p = 1$.

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

$$1 + r^b = \frac{1 + r^d}{p} \tag{16}$$

$$z_2 = qp(1 + r^b)\frac{\lambda - 1}{\lambda \alpha p - 1} > q(1 + r^d) = z_1 \tag{17}$$

In the absence of risk-taking, the equilibrium in the model with a perfectly competitive banking sector is similar to that of Moll (2014) without labor. In his paper, there is a single cutoff that determines who will be the creditors and lenders, with no positive margin. However, with a slight deviation ($\alpha p \lessgtr 1$) from the benchmark scenario, bankers introduce a positive wedge between the loan rate and the deposit rate, which is interpreted as a risk premium. Furthermore, due to the inefficiency of the gambling project ($\alpha p < 1$), there exists a fraction of autarky entrepreneurs. As a result, the risk taking motive is undesirable, not only because the gambling project is inefficient, but also because it leads to the inefficient allocation of resources towards unproductive producers.

The subsequent analysis concerns the effects of bank concentration on risks. In the model, the degree of entrepreneurial risk taking is measured by the level of capital investment in the gambling project, known as risky capital, denoted by rc_t at period t. The equilibrium level of risky capital is given by:

$$rc = v_r[\overline{K} - (1 - v_a)(\overline{K} - N)] \tag{18}$$

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

$$\frac{\partial rc}{\partial M} = (1 - v_a)\overline{K}\frac{\partial v_r}{\partial M} - v_r\overline{K}\frac{\partial v_a}{\partial M} + v_a v_r\frac{\partial N}{\partial M} \tag{19}$$

The initial component of Equation (19) concerns the effect of bank concentration on the share of risky loans. According to Proposition 2, bank concentration affects the loan rate, thereby influencing the proportion of risky loans via the *risk shifting mechanism*. Nevertheless, the impact of bank concentration on loan rates is still widely uncertain in the model.

The sign of the second element in Equation (19) depends on the impact of bank concentration on the proportion of autarky entrepreneurs. When the banking sector becomes more concentrated, the net margin widens, leading to an increase in the fraction of autarky entrepreneurs. I refer to this as the *"net margin mechanism"*. The validity of this channel has been established in Yi (2022) under the assumption of a uniform distribution of productivity and the absence of a risk taking motive. This paper will demonstrate the *net margin mechanism* quantitatively in the following section. Autarky entrepreneurs, who finance their projects with their own funds, tend to invest in safe ventures. Therefore, the *net margin mechanism* generates a negative correlation between bank concentration and risks.

The third component in equation (19) hinges on the sensitivity of bank capital to variations in bank concentration. The subsequent section of this paper will provide clarification that this term holds a positive value, but it is quantitatively insignificant.

Therefore, the relationship between bank concentration and risk depends on the impact of

bank concentration on the loan rate, as well as the magnitude of the *risk shifting mechanism* and the *net margin mechanism*. Specifically, if a highly concentrated banking sector leads to a higher loan rate, bank concentration would increase risks through the *risk shifting mechanism*, while would reduce risks through the *net margin mechanism*. The overall effect is ambiguous and relies on the magnitude of these two mechanisms. Conversely, if higher bank concentration results in a lower loan rate, both mechanisms would lead to a negative correlation between bank concentration and risks.

Role of Bank Capital. Given the limited information available to bankers about entrepreneurs' productivity and investment preferences, they are limited to two instruments in the loan market: bank capital and loan quantity (loan rate). In a highly concentrated banking sector, bankers are disincentivized from raising loan rates too high due to the *risk shifting mechanism*. Instead, bankers may opt to increase the probability of loan repayment by accumulating more capital to lower the loan rate. As such, the amount of bank capital plays a crucial role in determining the impact of bank concentration on loan rates and risks through the *risk shifting mechanism*. In the following section, a quantitative examination of the relationship between bank concentration and risk will be presented. The analysis demonstrates that this relationship depends on whether the bank capital constraint is binding.

4   Quantitative Analysis

This section first calibrates the parameters in the model, followed by a quantitative analysis of how bank concentration affects risks through the *risk shifting mechanism* and *net margin mechanism*. Specifically, I will examine two scenarios: 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 will also enable me to investigate the impact of bank concentration on allocative efficiency.

4.1   Calibration

I choose parameters to match several key moments of the U.S. economy in the period between 1994 and 2020. The primary focus of this calibration is on three key factors: the distribution of productivity, the level of bank concentration, and the quality of U.S. financial institutions as measured by the asset pledgeability parameter λ.

Bank concentration ($\frac{1}{M}$) is measured using the average HHI of the U.S. banking sector over

the years between 1994 and 2020. The HHI is defined as follows

$$HHI = \sum_{i=1}^{M} s_i^{d2} = \sum_{i=1}^{M} (\frac{1}{M})^2 = \frac{1}{M} \tag{20}$$

where the second equality follows that in the steady state of the symmetric equilibrium, each banker represents a market share of $1/M$ in the deposit market. To measure the bank concentration, I adopt the approach of Drechsler et al. (2017), where HHI is computed as a weighted average of branch-level HHI using branch deposits as weights. Based on Equation (20), M is estimated to be approximately 7.45.[5]

In the preceding sections, no specific distribution of productivity has been emphasized. Nonetheless, the assumption that firm productivity follows a Pareto distribution has been widely adopted, largely due to Melitz (2003). In this paper, the Pareto distribution is replaced by the bounded Pareto distribution, in line with the assumption in Proposition 2. The bounded Pareto distribution is characterized by the shape parameter γ, the maximal value z_{max}, and the minimal value z_{min}, with z_{min} normalized to 1. I calibrate z_{max} and γ to match the dispersion of productivity and markups for U.S. in the sample years. As illustrated in Hsieh and Klenow (2009), the discrepancy between the 75th and 25th percentiles of TFPR amounts to 0.53.[6] The cumulative density distribution function of $log(z)$ is $\frac{z_{min}^{-\gamma} - e^{-\gamma z}}{z_{min}^{-\gamma} - z_{max}^{-\gamma}}$ when the productivity z follows a bounded Pareto distribution[7]. Therefore, z_{max} is approximately 3. Additionally, the shape parameter γ is set to be 1.5, which is chosen to maintain the markup at around 20%, as suggested by Liu and Wang (2014).[8]

Based on the value of M, the asset pledgeability parameter λ is chosen to match the bank capital to asset ratio in the U.S. between 2001 and 2017. Specifically, a higher value of λ corresponds to a more efficient financial market and results in a higher bank capital to asset ratio. According to FRED, the average bank regulatory capital to risk-weighted assets for the U.S. during the aforementioned period is 13.71%. Given the previously calibrated value of M,

---

[5]Note that M must be an integer in the model economy. However, to avoid compromising the precise calibration of bank concentration, the approximation of the number of banks to 7 or 8 is avoided. In comparative statics, however, Ms are set to be integers.

[6]Hsieh and Klenow (2009) distinguish between TFPQ and TFPR, where TFPQ is calculated using the plant-specific price deflator and TFPR is computed using the industry price deflator. However, due to the normalization of the price of the consumption good, TFPQ and TFPR are essentially equivalent in this paper.

[7]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 $F_X(x) = Pr(X \leq x) = \frac{L^{-\gamma} - x^{-\gamma}}{L^{-\gamma} - H^{-\gamma}}$. Therefore, the $c.d.f.$ of Y is $F_Y(x) = Pr(Y \leq x) = Pr(log(X) \leq x) = Pr(X \leq e^x) = \frac{L^{-\gamma} - e^{-\gamma x}}{L^{-\gamma} - H^{-\gamma}}$. Correspondingly, the probability distribution function of Y is $\frac{\gamma e^{-\gamma x}}{L^{-\gamma} - H^{-\gamma}}$.

[8]It should be noted that I introduce an imperfect competition in the banking sector, which results in an increase in the markup. As a result, the required value of γ is not as high as in Liu and Wang (2014).

λ is approximately 15.

In accordance with Basel III, the parameter κ is used to generate the implied policy requirement. III mandates a minimum Total Capital Ratio of 8%. Additionally, the capital conservation buffer requires financial institutions to hold at least 10.5% of risk-weighted assets in capital. Since the benchmark model does not include the risk-based capital constraint, the value of κ is simply set to 0.08. However, the appendix will demonstrate that incorporating risk-based capital constraints does not substantially alter the primary mechanisms in the model.

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 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

Table 2: Calibrated Parameter Values

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

4.2 Equilibrium with only Risky Loans

One potential scenario is that all entrepreneurs exhibit a preference for gambling projects, resulting in all loans being risky. Such an extreme case is characterized by

Corollary 2 *Assume $\alpha p \lessapprox 1$. All loans are risky in the equilibrium.*

The proof can be derived from Equation (7), where z_3 converges to infinity as the value of αp approaches 1. Notably, compared to the prudent project, a gambling project involves higher costs due to its lower expected return. When $\alpha p \lessapprox 1$, the difference in expected returns between gambling and prudent projects diminishes, prompting all borrowing entrepreneurs to opt for gambling projects.

With $p = 0.9$ and $\alpha = 1.05$, Figure 2 presents the comparative statics with respect to bank concentration. As depicted in Panel (b), all loans are risky. When the banking sector is highly concentrated, Panel (a) demonstrates that the bank capital to asset ratio greatly exceeds the minimum capital requirement. However, when bank concentration is low, the minimum bank capital requirement becomes binding. According to empirical evidence presented in Yi (2022), there exists a positive correlation between bank concentration and bank capital, and the U.S. banking sector exhibits an excess accumulation of bank capital ratio relative to the minimum capital requirements. The underlying intuition is straightforward: a more concentrated banking sector implies a lower deposit rate and deposit supply. Since deposits and equity capital are both liabilities on the bank's balance sheet, a more concentrated banking sector results in an increase in the bank capital ratio through the substitution effect between the two.

Panel (c) of Figure 2 presents evidence of how reduced bank concentration leads to an increase in the allocation of capital to gambling projects. The relationship between bank concentration and risks is significantly negative regardless of whether the capital constraint is binding. With all loans being risky, the *risk shifting mechanism* becomes irrelevant, leaving only the *"net margin mechanism"* to shape the relationship between bank concentration and risk taking behavior. A more concentrated banking sector, as illustrated in panels (e) and (f) of Figure 2, leads to a higher spread charged by bankers, resulting in an increased percentage of autarky entrepreneurs. These autarky entrepreneurs who use their initial endowments to engage in production activities will choose the prudent project. Consequently, bank concentration and risk exhibit a significantly negative correlation.

The allocation distortion induced by bank concentration can also be explained by the *"net margin mechanism"*, as depicted in panel (d) of Figure 2. As bank concentration increases, both the net margin and the proportion of autarky entrepreneurs rise. Inefficient autarky entrepreneurs receive more capital allocation through the extensive margin in a more concentrated banking sector. Additionally, Yi (2022) demonstrates that the non-binding capital constraint might also lead to capital misallocation.

4.3 Equilibrium with Both Safe and Risky Loans

In a more general scenario of the equilibrium, loans may be either risky or safe. Under this condition, in addition to the *net margin mechanism*, the *risk shifting mechanism* described in Proposition 2 reshapes the relationship between bank concentration and risks. To allow for the coexistence of gambling and prudent projects in the steady state, I set $\alpha = 1.05$ and $p = 0.7$ in this section.[9] The impact of bank concentration on the loan rate is depicted in Figure 3.
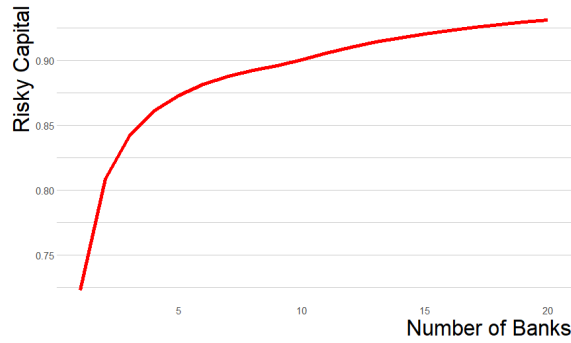

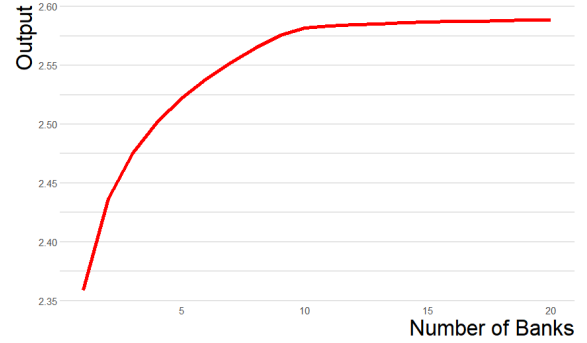[9]The calibration is performed under this parameter setting.
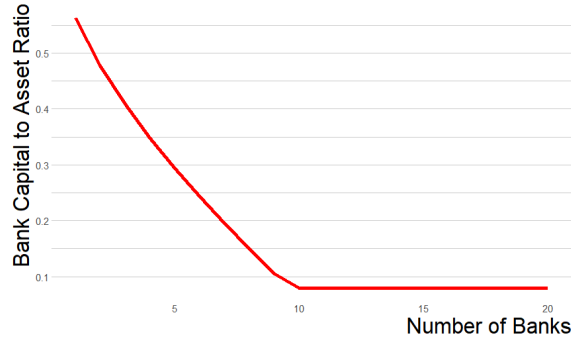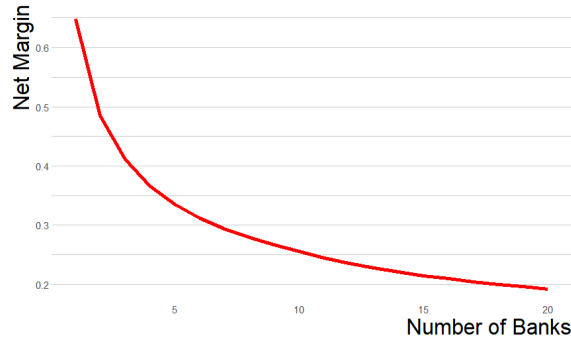
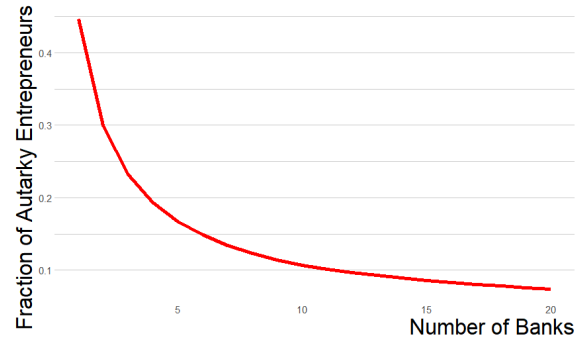(a) Bank Capital to Asset Ratio
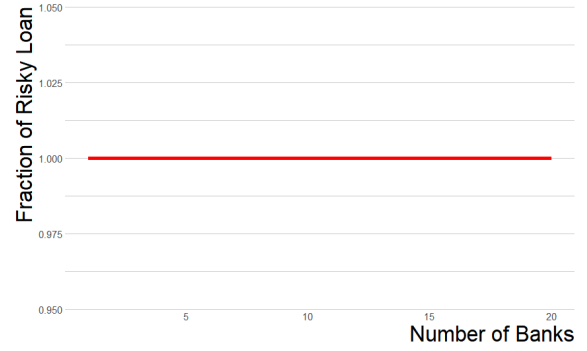
(b) Fraction of Risky Loan

(c) Risky Capital

(d) Output

(e) Net Margin

(f) Fraction of Autarky Entrepreneurs

Figure 2: Comparative Statics of Model with only Risky Loan

Notes: 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.
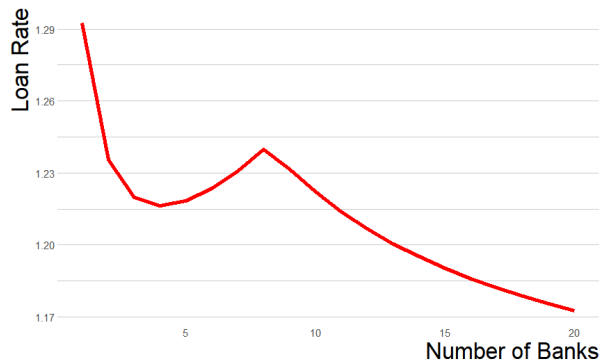
Figure 3: Bank Concentration and Loan Rate

Upon considering the *risk shifting mechanism*, Figure 3 illustrates that the relationship between bank concentration and loan rate is non-monotonic. The loan rate may increase in response to profit maximization when bank concentration is high. Specifically, as the banking sector becomes more concentrated, the elasticity of loan demand decreases. However, the *risk shifting mechanism* implies that bankers charging a higher loan rate will attract more gambling entrepreneurs and a lower expected probability of loan repayment. In response, bankers internalize the best response of the entrepreneurs and will not set a rate that is too high. Figure 3 demonstrates a surprising negative correlation between bank concentration and loan rate when there are approximately 4 to 8 banks in the economy.

In contrast, in a highly concentrated banking sector, there exists a positive correlation between bank concentration and the loan rate. This is due to the fact that, in the general equilibrium, bank concentration reduces output, thereby decreasing capital demand. Consequently, the price of capital (q) falls, which drives up the loan rate again.

The point at which banks begin to accumulate excess capital is represented by the kink in Figure 3, as demonstrated in panel (a) of Figure 4. Bankers in a highly concentrated banking sector are incentivized to maintain a capital ratio exceeding the minimum requirement. This motivation is further enhanced when considering the *risk shifting mechanism*. Bankers with more capital may issue more loans, thereby reducing the loan rate and the proportion of risky loans.

Panel (b) of Figure 4 depicts the relationship between bank concentration and the fraction of risky loans. The results reveal that when the bank capital constraint is binding, a rise in bank concentration leads to an increase in the proportion of risky loans, while when the bank capital constraint is non-binding, the fraction of risky loans declines with increased bank concentration. Specifically, when the bank capital constraint binds, bankers with greater market power tend to charge higher loan rate, which results in a higher fraction of risky loans due to the *risk shifting*
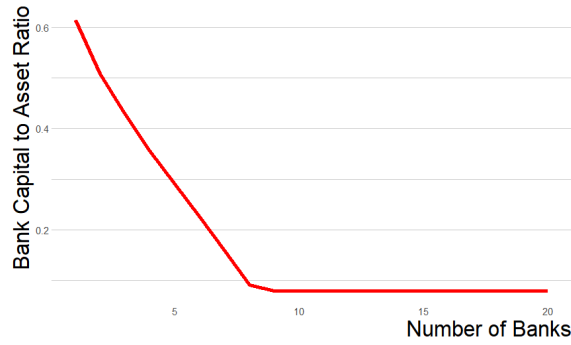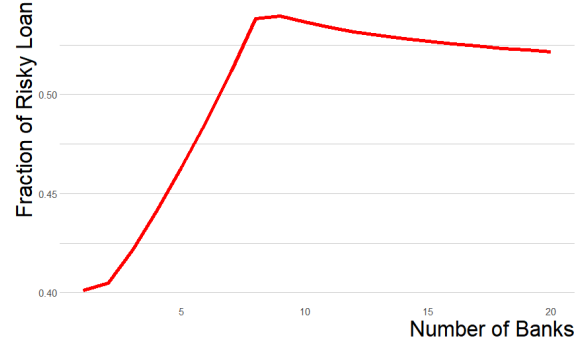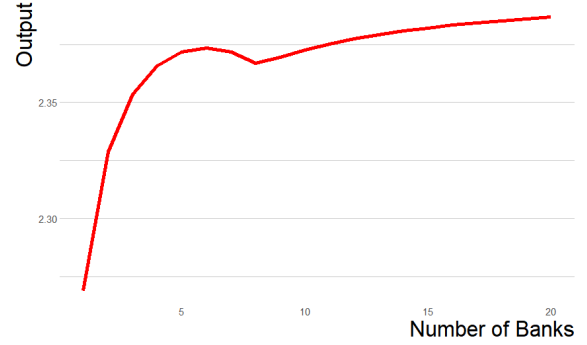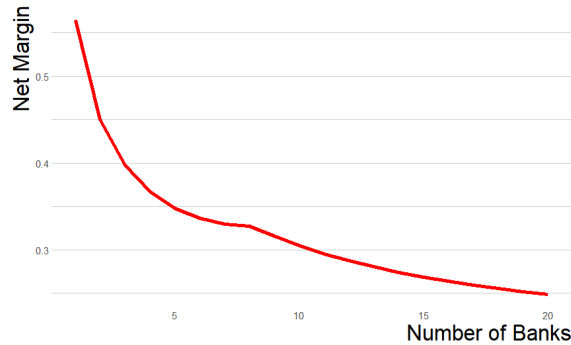
22

(a) Bank Capital to Asset Ratio
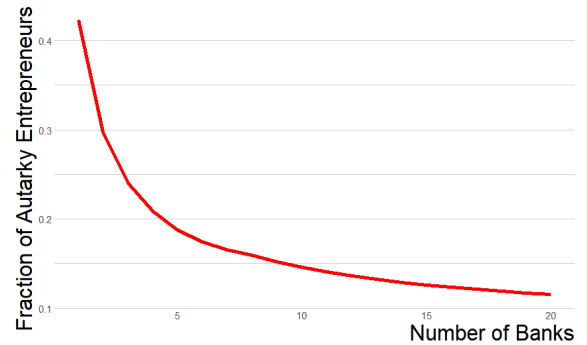
(b) Fraction of Risky Loan
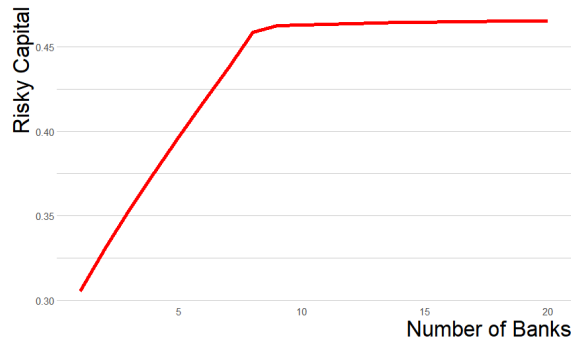
(c) Risky Capital

(d) Output

(e) Net Margin

(f) Fraction of Autarky Entrepreneurs

Figure 4: Comparative Statics of Model with Both Safe and Risky Loan

Notes: 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.

mechanism. In contrast, when bankers maintain excess capital above the required minimum, higher bank concentration leads to a lower loan rate and a smaller proportion of risky loans.

It is noteworthy to acknowledge that the relationship between the proportion of risky loans and the loan rate in the general equilibrium is non-monotonic. Specifically, when bank concentration reaches extremely high levels, the equilibrium effect through the price of capital leads to an increase in the loan rate. However, in terms of consumption, the return rate of loans $q(1+r^b)$ is positively correlated with the fraction of risky loans v_r in the steady state. Further details regarding the relationship between $q(1+r^b)$ and the number of bankers M are provided in Figure D.1 in the appendix. In fact, increasing the price of capital also results in a higher proportion of risky loans due to the elevated external funding costs incurred by entrepreneurs.

4.3.1 Bank Concentration and Risk Taking

The relationship between bank concentration and entrepreneurs' risk taking is illustrated in panel (c) of Figure 4. When the bank capital constraint is non-binding, the level of risky capital is negatively correlated with bank concentration. However, when the constraint is binding, the correlation between the two variables is relatively weak.

In the presence of a binding bank capital constraint, an increase in bank concentration leads to higher loan rates. Although the *risk shifting mechanism* suggests that bank concentration and risky capital should be positively correlated, this relationship is not evident in panel (c) of Figure 4. This can be attributed to the presence of the second element in Equation (19), i.e., the *net margin mechanism*, which exerts a compensating effect in the opposite direction. Specifically, as bank concentration increases, the proportion of autarky entrepreneurs, who invest only in prudent projects, is also expected to be higher. As a result, the level of risky capital decreases, counteracting the positive relationship between bank concentration and risky capital. Based on calibrated parameters to match U.S. moments, the magnitudes of the two mechanism are so similar that the correlation between bank concentration and risks is almost negligible. However, when the capital ratio exceeds the minimum requirement, both mechanisms lead to a negative correlation between bank concentration and risk taking.

4.3.2 Bank Concentration and Output

Panel (d) of Figure 4 illustrates the impact of bank concentration on output. Interestingly, a non-monotonic relationship is observed between these two variables, taking into account both the *net margin mechanism* and the *risk shifting mechanism.*

The *net margin mechanism* suggests that lower bank concentration leads to higher output. Specifically, when bank concentration is high, bankers charge a wider spread between the loan

and deposit rates, thereby increasing the allocation of resources to autarky entrepreneurs, who tend to be the least efficient producers. Empirical evidence by Joaquim et al. (2019) indicates that a reduction in lending spreads to the world average would increase Brazilian output by 5%. Additionally, higher bank concentration leads to a higher bank capital ratio in the equilibrium region where the bank capital constraint is non-binding, which reduces the proportion of risky loans through the *risk shifting mechanism*. Under Assumption 1, where gambling projects have a lower expected payoff, the negative impact of bank concentration on output is mitigated. In fact, the negative relationship between bank concentration and output is reversed when the number of banks ranges from approximately 6 to 8, which represents a local optimum. Given that the number of banks M is calibrated at 7.45 in the preceding section, this finding may have significant quantitative implications.

5   Discussions

In this section, corroborative evidence is presented based on the model predictions. In line with the model, the observations suggest a non-monotonic relationship between bank concentration and loan rates in the U.S. Furthermore, the model characterization provides scope for regulatory measures aimed at enhancing both efficiency and stability.

5.1   Supporting Evidence

In this section, novel empirical evidence is presented based on U.S. data to examine the non-monotonic relationship between bank concentration and the loan rate. Although the findings do not directly address the impact of bank concentration on entrepreneurs' risk taking behaviors, they shed light on the workings of the *risk shifting mechanism* through the loan rate.

5.1.1   Data Description

The analysis draws on three distinct data sources: (i) Summary of Deposits from the Federal Deposit Insurance Corporation (FDIC), (ii) bank balance sheet information from U.S. Call Reports provided by the Federal Reserve Bank of Chicago, and (iii) branch-level rate data from RateWatch. The primary characteristics of each dataset are discussed in detail in this section. **Deposit Quantity.** The dataset on deposit quantities is obtained from the 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.[10]

RateWatch. RateWatch data covers monthly loan rates at the branch level. The sample period spans from 1994 to 2021. The analysis focuses on one of the most frequently observed loan types, namely auto loans with a 72-month maturity.[11] Using this strategy, I am able to mitigate potential issues associated with observed and unobserved heterogeneity among different loan products. The analysis specifically considers the branches that actively participate in setting the loan rate.

Following Drechsler et al. (2017), I use HHI to measure bank concentration. Specifically, I begin by constructing HHI at the county-year level, which is calculated as the sum of the squared deposit market share of each branch by county for each year (as shown in Equation (20)). To obtain a bank-level HHI, I compute the weighted average HHI of all branches belonging to the same bank institution, with branch deposit sizes as the weights.

5.1.2 Bank Concentration and Loan Rate Revisited

As predicted by the theoretical model, the relationship between bank concentration and the loan rate is non-monotonic. In this section, I aim to verify this prediction empirically by employing a fixed effect regression model to analyze the effect of branch-level HHI on loan rates. Specifically, I start by estimating the following regression

$$LoanRate_{kt} = \sum_{i=1}^{10} \beta_i HHI_{c(k)t} * \mathbb{1}(HHI_{c(k)t} \in (\frac{i-1}{10}, \frac{i}{10}]) + \alpha_{j(k)} + \alpha_t + \alpha_{s(k)t} + \epsilon_{jt} \qquad (21)$$

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 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 $\alpha_{s(k)t}$ into the regression is grounded on Rice and Strahan (2010), where they construct a deregulation index at the state level.[12] I cluster the standard error at bank

---

[10]In the Appendix, a local polynomial smoothing technique will be employed to demonstrate the non-monotonic correlation between bank concentration and loan rate at the bank level.

[11]According to the model, the borrowers represent entrepreneurs who use external resources for production. This paper considers auto loans as they constitute the most commonly observed loan type in the dataset. Additionally, households are obviously not the only borrowers, some firms also purchase automobiles to promote their businesses. Appendix, I will conduct further analyses using business loans as a robustness check.

[12]Interstate branching was not permitted in the U.S. until the Riegle-Neal Act was passed in 1994. In order to mitigate risks associated with financial institutions, the Dodd-Frank Act was enacted in 2010. However,

level. The main coefficients of interest in the regression are β_i, where i=1, 2, $\cdots$, 10. These coefficients capture the heterogeneous effect of bank concentration on loan rate within deciles of bank concentration. Specifically, a positive β_1 indicates a positive correlation between HHI and loan rate when HHI falls within $(0, 0.1]$. Conversely, a negative β_i suggests a negative correlation between HHI and loan rate when HHI falls within the $i - 1$th to ith decile. As per the model's prediction, β_i is expected to take on both positive and negative values.

I control for bank fixed effect and time fixed effect in Figure 5. The results show that the coefficients are positive and significant when HHI is in the range of $(0, 0.6]$. This indicates a positive correlation between bank concentration and loan rate in this range. In contrast, the coefficient becomes negative and significant at the 1% significance level as HHI increases. Specifically, β_7 is -0.27. This means that as HHI increases from 0.6 to 0.7, the loan rate decreases by 0.027%. This negative correlation is in accordance with the model prediction and can be attributed to the *risk shifting mechanism* and non-binding capital constraint: when the loan rates rises, entrepreneurs are more likely to engage in gambling projects. Banks internalize the entrepreneurs' decisions and dislike a high loan rate, even in a highly concentrated banking sector. Instead, they accumulate excess capital above the minimum capital requirement. In addition, the positive and significant values of β_9 and β_{10} are a manifestation of the general equilibrium effect through the price of capital. Overall, The non-monotonic relationship between HHI and loan rate is consistent with the model equilibrium. Further specifications are presented in Table 6 in the Appendix.

5.1.3 Interaction between Bank Concentration and Bank Capital

In accordance with the model predictions and the results presented in Table 6, the relationship between bank concentration and loan rate is not monotonic. The non-monotonicity hinges on the binding nature of the bank capital constraint. Specifically, when the bank capital constraint is binding, higher bank concentration is associated with higher loan rates. However, when banks hold excess capital above the minimum requirement, the relationship between bank concentration and loan rate becomes ambiguous (U-shaped) due to the *risk-shifting mechanism* and the general equilibrium effect. To validate the model predictions, I estimate the following regression:

$$LoanRate_{kt} = \beta_1 HHI_{c(k)t} + \beta_2 HHI_{c(k)t} * Low\ Capital_{jt} + \alpha_{j(k)} + \alpha_t + \alpha_{s(k)t} + \epsilon_{jt} \qquad (22)$$

states are authorized to utilize the four key provisions contained in the IBBEA to restrict or increase the cost of out-of-state entry. Based on this, Rice and Strahan (2010) created a bank deregulation index, which ranges from 0 to 4, with 4 indicating states with the most stringent entry requirements for out-of-state banks. To address concerns related to varying deregulation policies across states, the state-time fixed effect is added.
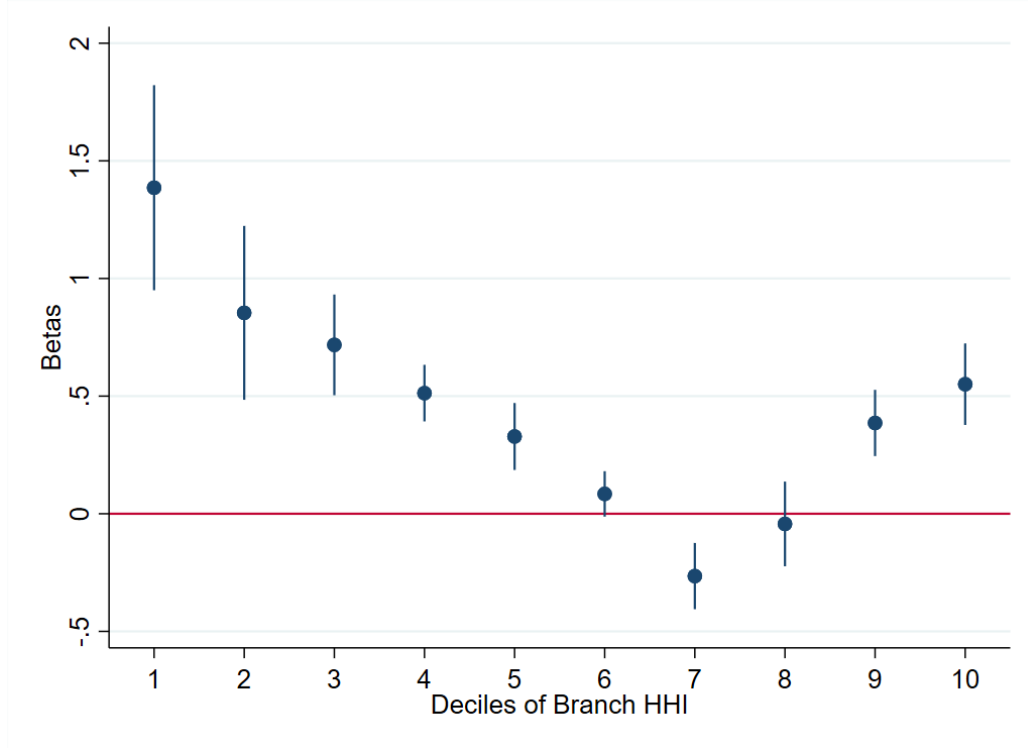
Figure 5: Branch-level HHI and Loan Rate (Auto Loan)

Notes: Figure 5 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 21). The figure shows pointwise estimates and the 5 % 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 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 captures the heterogeneous effect of bank concentration on loan rate across different groups categorized by their capital ratios. I cluster the standard error at bank level.

The results are reported in Table 3 and various fixed effects are controlled for in each column. The findings indicate that the impact of bank concentration on loan rate is not significant when the bank capital ratio is high, which is consistent across different specifications. In contrast, the coefficient of the interaction term between bank concentration and low bank capital ratio is positive and statistically significant in columns 2 and 3. Specifically, the estimate of β_2 is approximately 0.279 and 0.308 in columns 2 and 3, respectively. These results suggest that the effect of bank concentration on loan rate varies significantly across groups with different bank capital ratios. Moreover, $\beta_1 + \beta_2$ represents the total effect of bank concentration on loan rate when the bank capital ratio is low, which is positive and significant at a 5% level.

According to table 3, the effect of bank concentration on loan rate is positive when the bank capital ratios are low, while it becomes ambiguous when the capital ratios are high. The observations with high capital ratios reflect the behavior of bankers who accumulate excess capital above the minimum requirement in the model. Therefore, the positive and significant estimate of $\beta_1 + \beta_2$ in Table 3 is attributed to the decline in the elasticity of loan demand with the rise in bank concentration, only when the bank capital constraint is binding. In contrast, when banks accumulate excess capital above the minimum requirement, the *risk shifting mechanism* and the general equilibrium effect lead to a U-shaped correlation between bank concentration and loan rate, rendering β_1 insignificant.

5.2 Policy Implications

Drawing on the model equilibrium characterization, it becomes clear that the bank capital constraint plays a significant role in determining the relationship between bank concentration, risks, and allocative efficiency. In this section, I discuss the policy implications of improving stability and efficiency simultaneously.

A pertinent question is whether the removal of barriers to bank competition would be adequate to enhance both efficiency and stability. When the bank capital constraint is binding, decreasing bank concentration would improve efficiency, and have a negligible impact on risks as the *risk shifting mechanism* and the *net margin mechanism* operate in opposite directions.

Variables	(1) OLS	(2) OLS	(3) OLS
Branch-HHI	-0.0114 (0.104)	0.0738 (0.111)	0.0737 (0.108)
Branch-HHI*Low Capital	0.242 (0.152)	0.279* (0.168)	0.308* (0.170)
Constant	5.42*** (0.0180)	5.41*** (0.0162)	5.41*** (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

Table 3: Bank Concentration and Loan Rate in Low/High-Capital-Ratio Group (Auto Loan)

Notes: Table 3 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.

However, when banks accumulate a capital ratio exceeding the minimum requirement, a trade-off between efficiency and stability arises. To simultaneously enhance efficiency and stability, it would be advisable to decrease bank concentration and increase the minimum capital requirement. A higher level of bank capital would not only expand the region where the bank capital constraint is binding and risk is insensitive to bank concentration, but would also diminish the level of risk taken by entrepreneurs by lowering the loan rate.

In accordance with Figure 4, the reduction of bank concentration may lead to lower efficiency and higher risks when the number of banks falls within the range of approximately 6 to 8. Therefore, reducing bank concentration in the short run may not always be beneficial, as a local rather than global optimum is reached at an approximate number of 7 banks. As a consequence, policymakers should be confident in their decision to reduce obstacles to bank competition, even if a short-term loss of welfare is observed. The argument is relevant given that the calibrated number of banks in the U.S. using HHI is 7.45.

5.3 Exogenous Variation of Bank Concentration

In the baseline model, the level of bank concentration is captured by an exogenous variable, denoted as M. Based on the equilibrium characterization, the relationship between bank con-
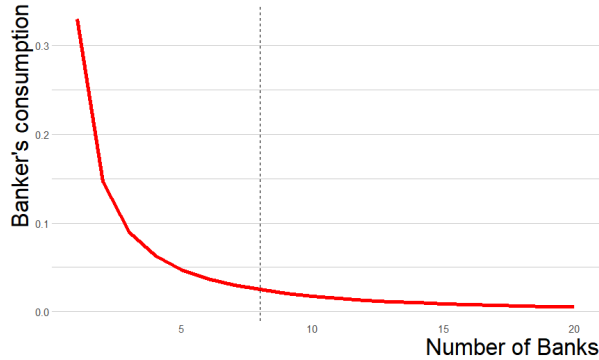
Figure 6: Number of Bankers and Bankers' Consumption

centration $(\frac{1}{M})$ and risk taking (rc) depends on whether the bank capital constraint is binding. However, in reality, the level of bank concentration is determined endogenously by other market conditions such as switching costs, entry costs, M&A episodes, and other factors.

This section is concerned with endogenizing the number of bankers M by allowing the entry cost to vary. Specifically, when bankers choose to enter the banking industry, they are required to pay a fixed amount of τ. In the steady state of a symmetric equilibrium, the free entry condition can be expressed as

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

which equates the lifetime utility derived from consumption with the entry cost. To gain insights into how entry costs affect bank concentration, it is informative to examine the relationship between bankers' consumption and the implied number of bankers. As depicted in Figure 6, there exists a negative correlation between bankers' consumption and the number of bankers. Therefore, as entry costs increase, the number of bankers participating in the banking industry decreases. This, in turn, leads to a highly concentrated banking sector

Given the monotonic relationship between the consumption of bankers and the number of bankers, the extension of the model to incorporate an endogenous M is entirely analogous to the baseline model. Thus, in the baseline model, the number of bankers M can be regarded as an exogenous variation of bank concentration accompanied by varying entry costs.

6   Conclusion

In this paper, I develop a tractable dynamic model to investigate how bank capital affects the relationship between bank concentration and risk taking. The model suggests that when the banking sector is highly concentrated, accumulating excess bank capital not only enables

31

banks to maximize their profits, but also mitigates the impact of the *risk shifting mechanism*. The relationship between bank concentration and risk taking is found to be kinked due to the interplay between the *risk shifting mechanism* and the *net margin mechanism*, and depends on whether the minimum capital requirement is binding. Specifically, the model indicates 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. The findings of this paper raise concerns about future empirical studies that examine the effects of bank concentration on risks without considering bank capital levels.

The purpose of this paper is to explore idiosyncratic risk, which is often believed to be associated with financial stability. Nevertheless, it would be advantageous and imperative to incorporate aggregate risk into the model, as it would allow for a more comprehensive investigation of financial crises, financial distress, etc. To keep the model tractable, all terms in the model are expressed in real terms. This model, characterized by rich heterogeneity, would be useful for studying monetary policy with the inclusion of price rigidity. I leave all these extensions for future research.

References

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. and D'Erasmo, P. (2021). Capital buffers in a quantitative model of banking industry dynamics. *Econometrica*, 89(6):2975–3023.

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. *International, 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, 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*.

Appendices

A Proofs

Proof of Proposition 1. 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 denotes 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*

$$\Pi_t = z_t k_t + q_t(r_t^d + 1)(a_t - k_t) = [z_t - q_t(r_t^d + 1)]k_t + q_t(r_t^d + 1)a_t \tag{23}$$

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:*

$$\Pi_t = z_t \theta a_t - q_t(r_t^b + 1)(\theta - 1)a_t = [z_t - q_t(r_t^b + 1)]\theta a_t + q_t(r_t^b + 1)a_t \tag{24}$$

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:*

$$\Pi_t = p\{\alpha z_t \theta a_t - q_t(r_t^b + 1)(\theta - 1)a_t\} = p\{[\alpha z_t - q_t(r_t^b + 1)]\theta a_t + q_t(r_t^b + 1)a_t\} \tag{25}$$

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 (6) and (7) that:

$$\frac{(\lambda - 1)p}{\lambda \alpha p - 1}q(1 + r^b) = z_2 < z < z_3 = \frac{(\lambda - 1)(1 - p)}{\lambda(1 - \alpha p)}q(1 + r^b) \tag{26}$$

Further, $\frac{(\lambda-1)p}{\lambda\alpha p-1} > \frac{1}{\alpha}$ following Assumption 1. Therefore, under the condition implied by Equation (26), entrepreneurs will borrow up to the borrowing limits λ and invest in the gambling project.

By Assumption 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. Equations (8) and (9) are directly obtained from Proposition 1, 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:

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

by simply aggregating all the entrepreneurs of different productivities. ∎

Proof of Proposition 2. 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 (2), the budget constraint (3) and the minimum capital requirement (5). The Lagrangian function for banker i becomes:

$$L_{it} = q_t\{(1+r_t^b)p_t^e Q_{it}^L - (1+r_t^d)Q_{it}^D\} - q_t N_{it+1} + \mu_{it}(Q_{it}^D + N_{it} - Q_{it}^L) + \chi_{it}(N_{it} - \kappa Q_{it}^L) \quad (27)$$

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 (11) and (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:

$$\frac{\partial v_r}{\partial r^b} = \frac{[g(z_3)a_3 - g(z_2)a_2](1 - G(z_2)) + (G(z_3) - G(z_2))g(z_2)a_2}{(1 - G(z_2))^2} \tag{28}$$

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 (28) becomes:

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

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 $\frac{\partial p_e}{\partial r^b} \leq 0$.

∎

B   Robustness Checks with Other Loan Type

In this section, I will conduct regressions similar to those in Equation 21 and Equation 22 using secured business loans from the RateWatch dataset. The number of observations for secured business loans is 17,282, which is significantly smaller than the number of observations for auto loans. Given the limited dataset size, I estimate the following regression:

$$LoanRate_{kt} = \sum_{i=1}^{5} \beta_i HHI_{c(k)t} * \mathbb{1}(HHI_{c(k)t} \in (\frac{i-1}{5}, \frac{i}{5}]) + \alpha_{j(k)} + \alpha_t + \alpha_{s(k)t} + \epsilon_{jt}, \tag{29}$$

where the regression divides the entire sample into five equal parts and includes interaction terms between HHI and quintile indicators. Table 4 presents the results, showing that bank concentration has a positive and statistically significant impact on loan rates when branch-level HHI falls into the second or fifth quintile. Conversely, there is no significant association between bank concentration and the loan rate in other quintiles.

Based on the model predictions, the effect of bank concentration on 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 the price of capital. Due to the *risk shifting mechanism* in the model, however, the correlation between bank concentration and loan rate should be negative when the level of bank concentration is in between. The reason for not obtaining negative estimates in Table 4 might be that the dataset contains too much noise. There is a significant

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

Table 4: Bank Concentration and Loan Rate (Business Loan)

Notes: Table 4 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.

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 loan rate, I run the following regression:

$$LoanRate_{kt} = \beta_1 HHI_{c(k)t} * High\ Capital_{jt} + \beta_2 HHI_{c(k)t} * Low\ Capital_{jt} + \alpha_{j(k)} + \alpha_t + \alpha_{s(k)t} + \epsilon_{jt},$$
(30)

which is similar to Equation 22. Nevertheless, β_2 in Equation 30 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 5. It is consistent with the model predictions and the results presented in Table 5.

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

Table 5: Bank Concentration and Loan Rate in Low/High-Capital-Ratio Group (Business Loan)

Notes: Table 5 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.

C   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 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 variables in Figure C.1. 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 C.2, these four loan types account 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 high, there is a region where the correlation between bank concentration and loan rate is negative. The model predictions and branch-level evidence is consistent with this non-monotonicity. This could be attributed to the *risk shifting mechanism* and the non-binding capital constraint.

D   Additional Tables and Figures

Variables	(1) OLS	(2) OLS	(3) OLS
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0, 0.1]$)	1.386*** (0.222)	0.573*** (0.153)	0.569*** (0.178)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.1, 0.2]$)	0.854*** (0.188)	0.494*** (0.114)	0.491*** (0.129)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.2, 0.3]$)	0.718*** (0.109)	0.448*** (0.0705)	0.469*** (0.0886)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.3, 0.4]$)	0.513*** (0.0613)	0.299*** (0.0900)	0.279*** (0.0665)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.4, 0.5]$)	0.328*** (0.0726)	0.231*** (0.0867)	0.200** (0.0848)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.5, 0.6]$)	0.0843* (0.0493)	0.0171 (0.0519)	0.00593 (0.0543)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.6, 0.7]$)	-0.265*** (0.0718)	-0.249*** (0.0647)	-0.228*** (0.0563)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.7, 0.8]$)	-0.0433 (0.0918)	-0.0207 (0.106)	-0.0768 (0.112)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.8, 0.9]$)	0.386*** (.0717)	0.291*** (0.0563)	0.438*** (0.0561)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.9, 1]$)	0.551*** (0.088)	0.435*** (0.0610)	0.440*** (0.121)
Constant	4.75*** (0.0138)	4.80*** (0.00755)	4.80*** (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

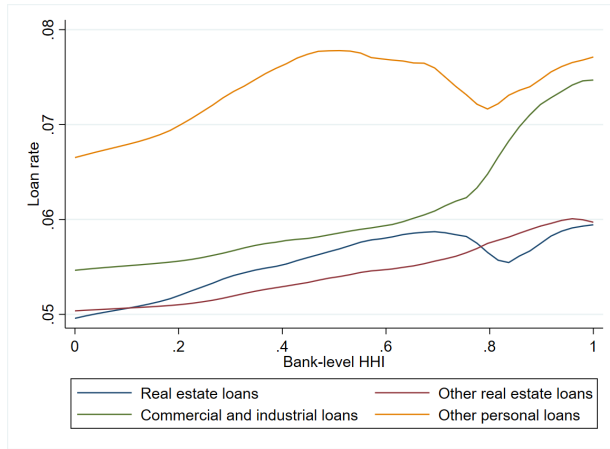Table 6: Bank Concentration and Loan Rate (Auto Loan)

Notes: Table 6 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 the 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.
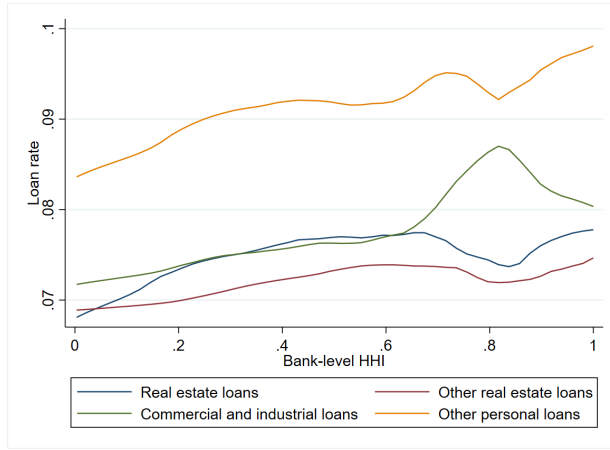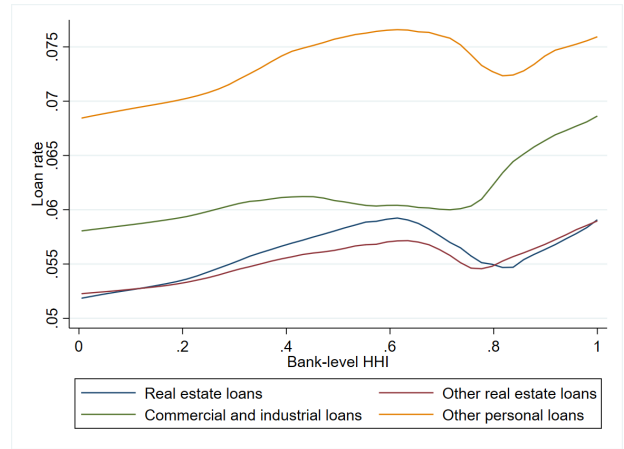
(a) Year 2008
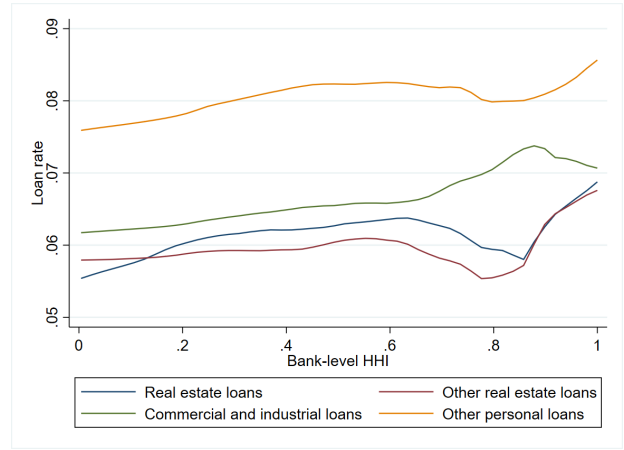
(b) Year 2012

(c) Year 2016

(d) Year 2020

Figure C.1: Bank Concentration and Loan Rate at Bank-level

Notes: 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.
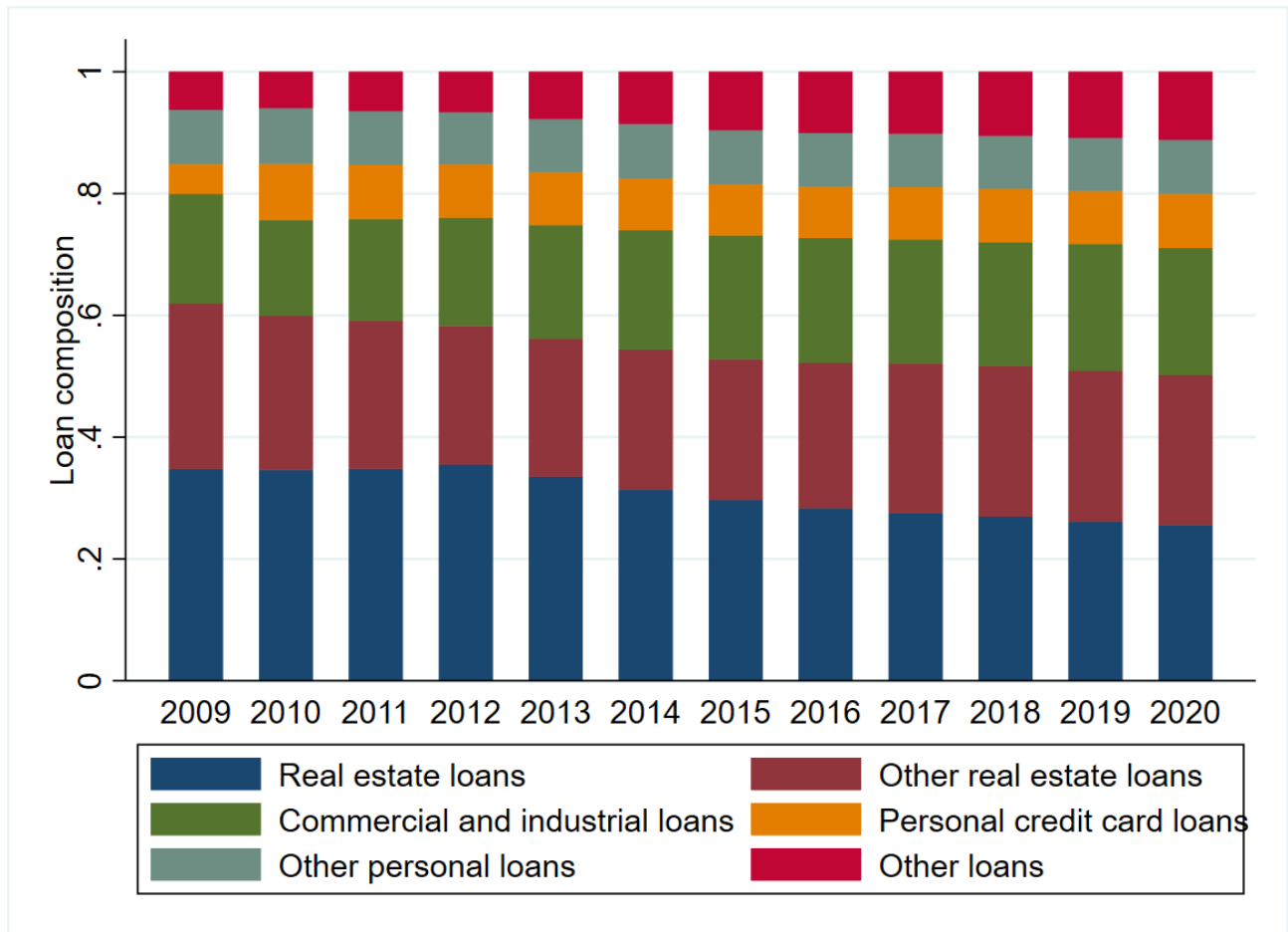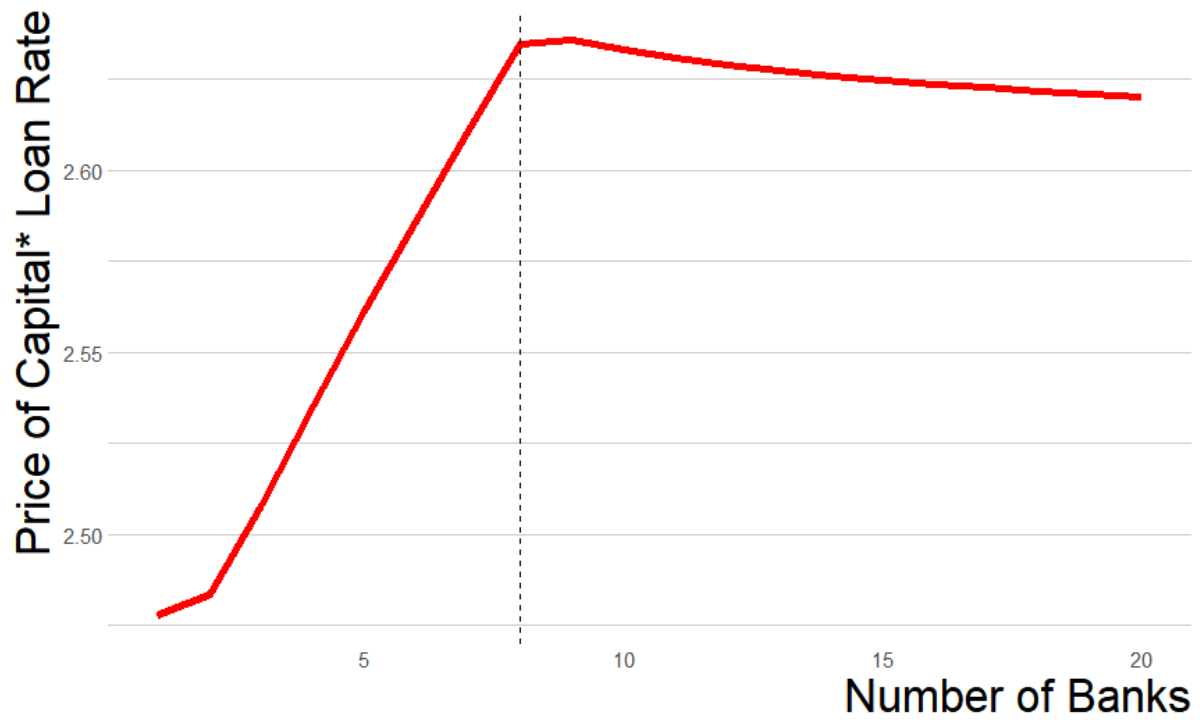
Figure C.2: Loan Composition in the U.S.

Figure D.1: Price of Capital times Loan Rate

Notes: This figure shows the price of capital q times loan rate $1 + r^b$ in the steady state, as a function of M. $q(1 + r^b)$ shapes exactly the same as the fraction of risky loan v_r.
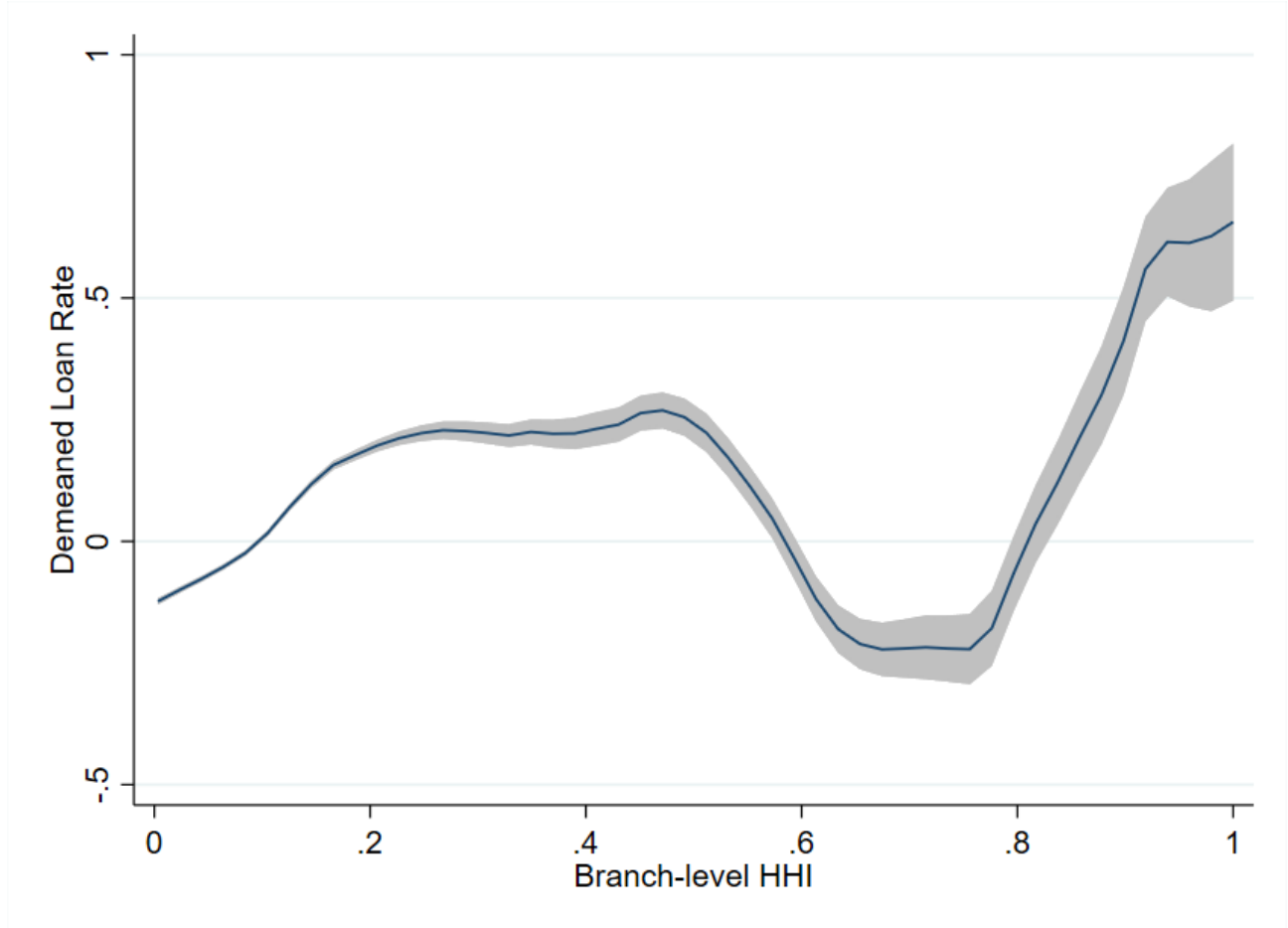
Figure D.2: Local Smoothing between Bank Concentration and Loan Rate (Auto Loan)

Notes: 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.