Bank Capital, Bank Concentration, and Risk Taking

Yu Yi*

November 29, 2022

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 that there is 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 the 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 the loan rate is positive 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

*Department of Economics,London School of Economics (E-mail: Y.YI3@lse.ac.uk). I am extremely grateful for Shengxing Zhang for his continued guidance and support. I would like to thank Maarten De Ridder, Wouter Den Haan, Matthias Doepke, Jonathon Hazell, Kai Li, Ben Moll, John Moore, Martin Oehmke, Ricardo Reis, Mark Schankerman, Pengfei Wang, Zhiwei Xu, Hongda Zhang, seminar participants at LSE, and conference participants at MMF (2022), Xiamen University PhD Macro Workshop for their advice.

1 Introduction

The impact of bank concentration[1] on risk taking has been extensively studied in the literature, but no consensus has been reached, neither theoretically nor empirically, among policymakers or academics. Many argue that reducing bank concentration encourages risk taking by squeezing bank profits and lowering franchise values (Corbae and Levine, 2018)[15]. Alternatively, some researchers contend that a more concentrated banking sector carries more risks (Carlson and Mitchener, 2009[12]). In this regard, bank concentration entails a higher loan rate, which, in turn, results in firms taking on additional risks, thereby increasing instability (Boyd and De Nicolò, 2005[8]).

The question whether bank concentration and risk taking are positively or negatively correlated remains pertinent as markedly different policies are implied from different perspectives. An observed positive correlation between concentration and risk urges policymakers to reduce impediments to competition so as to enhance both efficiency and stability. Those who support a negative correlation between bank concentration and risk emphasize the trade-off between competition and stability, which raises concerns regarding a highly concentrated banking sector and how to improve efficiency without putting the economy at risk using other policy instruments.

This paper emphasizes the role of bank capital in shaping the relationship between concentration and risk-taking. Bank capital in the United States has been shown to be significantly correlated with bank concentration (Yi, 2022[42]). Furthermore, bank capital is essential for economy-wide risks. Based on that, Basel III imposed higher capital requirements after the financial crisis in 2008.

In this paper, I build a tractable dynamic model that sheds light on the impact of bank concentration on risks and allocative efficiency by proposing the *net margin mechanism* and *risk shifting mechanism*. I find the answer depends on whether banks accumulate excess capital above the required minimum: when the bank capital constraint is binding, bank concentration ambiguously affect risk taking, while leads to allocative inefficiency; when the bank capital ratio exceeds the minimum capital requirement, agents are motivated to take risks in a competitive banking environment, and the relationship between bank concentration and output is non-monotonic, with the two mechanisms moving in opposite directions in terms of efficiency.

The model includes heterogeneous entrepreneurs and bankers as key agents. Entrepreneurs

[1] Many countries have seen a rise in bank concentration. For instance, the number of banks in the United States has declined from 10000 in 1997 to 5000 in 2017. In contrast, the top 3 asset share, defined as assets of the 3 largest commercial banks as a share of total commercial banking assets, has grown from 20% in 1997 to 35% in 2017.

are short lived and protected by limited liability. In each entrepreneur's sphere of activity, there are two types of projects - a prudent project and a gambling project, out of which one yields a certain return, and the other yields an excess return only if the project is successful. With limited enforcement and commitment, bankers facilitate the flow of credit amongst different entrepreneurs and compete in both loan and deposit markets à la Cournot.

Depending on their productivities, there are four types of entrepreneurs in the equilibrium: the borrowing entrepreneurs who gamble, the borrowing entrepreneurs who stay prudent, the lending entrepreneurs, and the autarky entrepreneurs. Entrepreneurs at the top of the productivity scale will borrow and produce at their maximum capacity, while making the optimal choice between the two projects. Those at the bottom prefer to deposit all their endowment in banks. 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 produce.

I find that borrowing entrepreneurs with lower productivity are more likely to gamble. This is a result of asymmetric information between bankers and entrepreneurs. Particularly, bankers do not observe entrepreneurs' productivity or project choices, but rather apply the same repayment rate to all the borrowers. With this loan rate, gambling entrepreneurs are protected by limited liability while receiving a lower income. The intuition then follows that, on the one hand, highly productive borrowers receive a large share of the loan return and therefore prefer the prudent project with a higher profit margin; on the other hand, those with low productivity invest in the gambling project and benefit from limited liability when the project fails. This trade-off offers a micro foundation for the *risk shifting mechanism*——a higher loan rate means a higher funding cost, as well as a larger fraction of gambling projects and risky loans.

Bankers internalize entrepreneurs' best responses when they choose the optimal loan rate. Increasing bank concentration could lead to higher interest rates due to the decline in the elasticity of loan demand. Due to the *risk shifting mechanism*, the motivation of bankers with market power to raise loan rates is mitigated. Instead, the accumulation of bank capital could not only reduce the moral hazard problem by lowering the loan rate, but also maximize profit in a highly concentrated banking sector. Bankers hold excess capital above the minimum requirement when bank concentration is large enough, resulting in a non-monotonic relationship between bank concentration and loan rate. A further empirical test of the model prediction reveals a negative correlation between bank concentration and loan rate when Herfindahl-Hirschman Index (HHI) is approximately 0.7.

As the loan rate rises, the *risk shifting mechanism* leads to increased stability and lower efficiency. The interaction between the *risk shifting mechanism* and bank capital, however, induces the relationship between bank concentration and stability to be dependent on the bank capital constraint. As long as the bank capital constraint is binding, there will be lower efficiency and higher risk taking in a more concentrated banking sector. Increasing bank concentration enhances both efficiency and stability as bankers accumulate excess capital above the required minimum.

Apart from the *risk shifting mechanism*, the relationship between bank concentration and risk taking is also affected by the *net margin mechanism*. As the banking sector becomes more concentrated, the wedge between the loan rate and deposit rate widens, resulting in a rising proportion of autarky entrepreneurs. Despite their inefficiency, these entrepreneurs always invest in prudent projects since they only use internal financing. Consequently, as bank concentration increases, the *net margin mechanism* leads to stability and inefficiency.

With both *risk shifting mechanism* and *net margin mechanism* are taken into account, a more competitive banking sector implies higher efficiency and ambiguous risk taking when the bank capital constraint is binding, with the two mechanisms having opposite effects on risk taking. Under the When the bank capital constraint is non-binding, however, a higher bank concentration leads to higher risk taking and a hump-shaped output with the two mechanisms having opposite effects on efficiency.

It is safe to remove impediments to competition when the bank capital constraint is binding. Reducing bank concentration will not have much impact on stability if the bank capital ratio is around the minimum requirement. If the capital ratio is above the minimum capital requirement, reducing bank concentration and raising the minimum bank capital requirement simultaneously can improve efficiency and stability.

1.1 Related Literature

The paper pertains to the literature discussing the impact of bank concentration on risk taking, which has not yet reached a consensus. Bank competition is not the same as bank concentration, whereas bank concentration is suggestive to bank competition. According to one perspective, there is a positive correlation between bank concentration and stability (Hellmann et al., 2000[26]; Beck et al., 2003[5]; Agoraki et al., 2011[1]; Tabak et al., 2015[40]; Jiang et al., 2017[29]; Carlson et al., 2022[11]; Beck et al., 2013[4]), where they find related empirical evidence by providing indirect measures of bank competition and stability. Concentration-fragility view, however, argues that intensifying bank competition stabilizes the economy (De Nicolo et al., 2004[17]; Beck et al., 2006[6]; Carlson and Mitchener, 2009[12]; Craig and Dinger, 2013[16]). My paper addresses this question theoretically and shows that the relationship between these two objects greatly depends on the bank capital constraint: the correlation is insignificant when the bank capital constraint is binding; bank concentration enhances stability when bank capital exceeds the requirement. The model implications are partially in line with the competition-fragility view.

The paper is most related to theories developed by Boyd and De Nicolò (2005)[8], Corbae and Levine (2018)[15], and Martinez-Miera and Repullo (2010)[35]. Boyd and De Nicolò (2005)[8] supports the competition-stability view by arguing that lower lending rates will reduce entrepreneurs' borrowing costs and motivate them to take less risk. In my 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)[15], 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. A U-shaped relationship between bank concentration and stability is produced by Martinez-Miera and Repullo (2010)[35] when their model allows for imperfect correlations among loan defaults. Unlike previous studies, I consider the excess accumulation of bank capital in my paper. Bank concentration will have a different impact on stability depending 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)[28], Black and Strahan (2002)[7], Diez et al. (2018)[18], and Joaquim at al. (2019)[30]. My paper contributes to the 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 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)[3], Kiyotaki and Moore (2019)[31] and Moll (2014)[37]. In particular, Moll (2014)[37] eases the *i.i.d.* assumption of productivity and shows how the persistence of idiosyncratic productivity shock affects misallocation. My research extends their models by including the bankers' problem in this setting and illustrates how bank concentration affects efficiency and stability when the financial market is incomplete.

There are various papers that discuss imperfect competition in the banking sector: Drechsler et al. (2017)[20], Lagos and Zhang (2022)[32], Van Hoose (2010)[41], Corbae and D'Erasmo (2021)[14], and Head et al. (2022)[25]. Drechsler et al. (2017)[20] follows Dixit and Stiglitz (1977)[19] by assuming the representative household substitutes deposits across banks imperfectly. Lagos and Zhang (2022)[32] incorporate bargaining power into the model to account for imperfect bank competition. Based on Burdett and Judd (1983)[10], Head et al. (2022)[25] study how bank concentration impacts monetary policy transmission. Corbae and D'Erasmo (2021)[14] develop a market structure where big banks interact with small fringe banks. My paper is closely related with Van Hoose (2010)[41], in which we both assume that banks compete à la Cournot, but differs at least in two ways: on the one hand, there is imperfect competition in both the deposit and loan market; on the other hand, the elasticities of loan demand and deposit supply are endogenously determined by entrepreneurs' decisions.

This is not the first theory that delves into bank capital: some papers focus on static models, in which bank capital is not a choice but a parameter (Brunnermeier and Koby 2018[9]), some papers impose an exogenous law of motion of bank capital (Li 2019[33]), and others assume that bank capital constraints are always binding (Repullo 2002[38]). My model endogenously determines bank capital by optimizing dividend payouts and equity issuance. This setup allows me to keep track of the relationship between bank concentration and bank capital, as well as a potential non-binding capital constraint.

An extensive body of literature has been developed that pertains to non-binding capital constraints. According to empirical evidence, banks are willing to hold more capital than required and adjust their capital ratio independently of capital regulations. For example, Alfon, Argimon and Bascuñana-Ambrós (2004)[2] show that banks in the U.K. increased their capital ratios in the last decade despite the fact that there was a reduction in the minimum capital requirement. Flannery and Rangan (2008)[22] find that the U.S. banking sector has undergone a dramatic capital buildup and half of large bank holding companies have more than doubled their equity ratios over the last decade. In my paper, the non-binding capital constraint plays an influential role in explaining the impact of bank concentration on stability, as well as the non-monotonic relationship between bank concentration and loan rate. From a theoretical perspective, the mechanism of excess bank capital accumulation is similar to Yi (2022)[42], where he emphasizes the substitution effect between bank capital and deposits. However, due to the *risk shifting mechanism* described in my paper, banks are even more motivated to hold 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 stability. 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. Further, I give policy implications. Section 6 concludes the paper.

Appendix provides the proofs.

2 Environment

Consider a model economy with discrete time and infinite horizon, where time is indexed by $t = 0, 1, 2, \cdots$. The model describes the credit structure in an economy with three types of agents that I call 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. Productivity z is assumed to be identically and independently distributed (*i.i.d.*) and follows an exogenous distribution of $G(z)$. Entrepreneurs are risk neutral and maximize the expected consumption

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

At period t, entrepreneurs of this generation are endowed with two production technologies, a prudent project and a gambling project. The prudent project gives a return of z with each unit of input of capital, while the gambling project yields αz with the probability of p and nothing otherwise. Whether the gambling project succeeds or fails depends on the realization of an idiosyncratic shock. Following the literature, I impose:

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

The above assumption implies that the gambling project will yield a higher return when it succeeds, but a lower expected return than the prudent project. The inefficiency of gambling projects can be explained by assigning a private benefit to borrowing entrepreneurs. Limited liability protects entrepreneurs who invest in gambling projects so that they will die with nothing if the project fails.

At the middle of each period, some entrepreneurs prefer to borrow, while others may lend. I assume that borrowers are unable to commit, and lenders are unable to enforce their promises. To make banks function as financial intermediaries, I assume that bankers have the ability to enforce and commit. The entrepreneurs may obtain external financing from the bankers, and repay their debt at r_t^b once the project has been successfully completed. On the other hand, entrepreneurs deposit their resources in banks and receive a return of r_t^d.

r_t^b and r_t^d stands for loan rate and deposit rate respectively. Entrepreneurs will give birth to offspring after producing and trading in the loan and deposit markets. Entrepreneurs of this generation will consume a certain percentage (s) of their net returns and invest the remainder in capital. The capital is then transferred to the next generation of entrepreneurs and distributed equally among them. Different from Moll (2014)[37], there is no heterogeneity of wealth among the entrepreneurs of the same generation. This homogeneity of wealth is not necessary for me, but it is one way to ensure a non-zero endowment for every entrepreneur, even if their parents leave nothing for them.

Additionally, entrepreneurs face a borrowing constraint

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

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

2.2 Bankers

The key assumption in the banking sector is imperfect competition. To characterize this, I assume there are $1, 2, \cdots, M$ long-lived bankers in the economy, 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 consists of a monopoly bank and when M converges to infinity, the banking sector is perfectly competitive. At the beginning of each period, each banker i is endowed with some equity capital N_{it}. Bankers are risk neutral and derive utility from dividend payouts

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

Bankers act as financial intermediaries and facilitate lending and borrowing between lending entrepreneurs and borrowing entrepreneurs. Using equity capital and deposits, the

[2]Following Van Hoose (2010)[41], I model imperfect bank competition using Cournot competition. It is a simple approach to examine the banking sector between the extremes of perfectly competitive banking ($M = \infty$) and monopoly banking ($M = 1$). The extreme cases under Cournot competition are equivalent to those when applying Bertrand competition (monopoly banking with $M = 1$ and perfectly competitive banking with $M > 1$). However, to generate an intermediate market structure with price competition, additional frictions may be required.

banker issues a loan contract, which could be safe or risky. The fraction of risky loans is denoted by v_{rt}. Bank equity capital is accumulated only through retained earnings.[3] Balance sheet identity of banker i then follows

$$Q^L_{it} = Q^D_{it} + N_{it} \tag{2}$$

The balance sheet items at the beginning of the period t are summarized in Table 1. I assume that each banker can fully diversify the idiosyncratic risks and analyze the equilibrium in regions in which no bankers default on deposits. At the end of period t, bankers' dividend payouts and retained earnings are funded by the return of their operations in the loan and deposit market. The intratemporal decision is simplified to a standard consumption and saving problem, where banker i faces a budget constraint

$$c^b_i + q_t N_{it+1} \leq (1 + r^b_t) q_t (1 - v_{rt}) Q^L_{it} + p(1 + r^b_t) q_t v_{rt} Q^L_{it} - (1 + r^d_t) q_t Q^D_{it} \tag{3}$$

RHS terms represent banker i obtains income—returns from issuing safe and risky loan contract, minus the repayment back to the depositors, which is used for financing the LHS variables—consumption of dividends and accumulation of banker's equity capital. The price of capital is q_t. To simplify equation (3), I define a new variable

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

the intuition of which is the expected probability of loan repayment. By construction, p^e_t represents a weighted average of repayment probabilities between safe loans and risky loans. Equation (3) then becomes

$$c^b_i + q_t N_{it+1} \leq q_t \{(1 + r^b_t) p^e_t Q^L_{it} - (1 + r^d_t) Q^D_{it}\}. \tag{3'}$$

Assets	Liabilities
Safe loans ($(1 - v_{rt}) Q^L_{it}$)	Deposits (Q^D_{it})
Risky loans ($v_{rt} Q^L_{it}$)	Equity capital (N_{it})

Table 1: Balance Sheet.

There is asymmetric information between entrepreneurs and bankers, who are uninformed of the types of entrepreneurs, including their productivity and project choices. As a result

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

of this, bankers charge a loan rate that applies to the entire population of entrepreneurs. However, the amount of loans they can issue is limited by the minimum capital requirement

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

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

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: a *"net margin mechanism"* and a *"risk shifting mechanism"*.

3.1 Entrepreneurs' Side

I will first derive the conditions under which gambling exists in equilibrium. Entrepreneurs, whose objectives are to maximize their expected consumption, choose to borrow when gambling, so as to benefit from limited liability. Otherwise, they prefer to invest in the prudent project that will provide them with a higher expected return. Because of the linearity of the production function, borrowing entrepreneurs who gamble always borrow up the borrowing limits. The incentive compatible condition then follows that borrowing entrepreneurs who gamble should obtain a higher expected return than if they self-finance and produce [4]

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

which yields a lower bound on productivity (z_2)

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

Further, borrowing and gambling should dominate borrowing and staying prudent

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

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

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

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

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

According to Assumption 2, entrepreneurs with productivity between z_2 and z_3 choose to gamble. There are three parameters involved in Assumption 2, namely α, p and λ. The inequality is more likely to hold when α is larger, λ is large or p is larger. Intuitively, the marginal benefit of gambling is higher when excess return α or success probability p is higher. When the asset pledgeability λ is higher, heterogeneity among borrowing entrepreneurs declines. Consequently, bankers are able to extract more profit from borrowers, inducing them to gamble.

Given the deposit and loan rate, entrepreneurs' financial decisions (borrow or lend) are fully characterized in Proposition 1.

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

[4] I neglect the time index here for simplification

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

The productivity cutoffs are defined by $z_1 = q(r^d + 1)$, $\overline{z_3} = Min\{z_3, z_{max}\}$, and z_2 follows Equation (6).

The cutoff property relies heavily on the constant return to scale of the production function. The optimal capital demand decision is at corners according to Proposition 1: it is zero for entrepreneurs with low enough productivity ($z < z_1$), maximum amount allowed by the borrowing constraint for those with high enough productivity ($z > z_2$), and initial wealth for those with intermediate level productivity ($z_1 < z < z_2$). Capital demand distinguishes two types of marginal entrepreneurs. For the entrepreneurs with productivity z_1, the return on each additional unit of capital $\frac{z}{q}$ equals the opportunity cost of not depositing that capital in the bank $r^d + 1$. The entrepreneurs with productivity z_2, however, are indifferent between using external finance to gamble and using their own capital to produce. Assumption 2 indicates that $z_2 < q(1 + r^b)$, therefore, the borrowing entrepreneurs with z_2 will not invest in the prudent project.

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

It is now sensible to refer to the entrepreneurs with productivity below z_1 as the lending entrepreneurs, those with productivity above z_2 as the borrowing entrepreneurs, and those

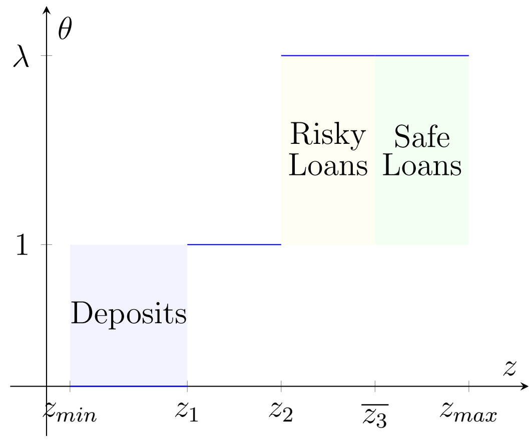


Figure 1: Entrepreneurs' Leverage and Project Choice

with productivity in between as the autarky entrepreneurs. The productivity of lending entrepreneurs is so low that it is not worthwhile for them to produce and instead deposit all their endowment in banks. Borrowing entrepreneurs are willing to borrow up to the asset pledgeability λ with their productivity being high. The introduction of imperfect competition in the banking sector results in the emergence of the third type of entrepreneur. Generally, banks charge a positive net margin between their loan and deposit rates, which induces some entrepreneurs to opt neither to borrow nor to lend. In this regard, entrepreneurs who have a productivity between z_1 and z_2 are referred to as autarky entrepreneurs. Because the autarky entrepreneurs employ 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 project choice of the borrower. There is an extreme case ($\overline{z_3} > z_{max}$) in which all loans are risky. I will discuss how the equilibrium behaves if there are only risky loans, as well as if there are both risky and safe loans. As entrepreneurs of each generation have the same initial wealth, the aggregate entrepreneurial capital demand equals the individual demand.

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

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

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

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

Equation (8) implies that the total loan demand depends on three elements: the proportion of borrowing entrepreneurs, the amount each entrepreneur borrows and entrepreneurs' initial capital holdings. Similarly, the deposit supply is given by lending entrepreneurs' total capital holding as described by equation (9). Equation (10) captures the law of motion for the aggregate entrepreneurial capital demand, where the wealth of entrepreneurs in the next generation $q_t a_{t+1}$ depends on their saving rate s and net return.

3.2 Bankers' Side

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

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

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

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 by κ. The benefits of issuing more loans are influenced by the elasticity of loan demand and the expected probability of loan repayment (second term on the LHS of Equation (12)).

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 $\overline{z_3} = z_{max}$

Based on Proposition 2, in partial equilibrium, when the lending rate is higher, the fraction of risky loans is higher, whereas the expected probability of loan repayment is lower. Following a spike in loan rates, more entrepreneurs are motivated to gamble, resulting in a higher proportion of risky loans. The *risk shifting mechanism* here is very similar to that proposed by Boyd and De Nicolo (2005)[8]. Their risk-incentive mechanism is completely based on the functional assumption of project return, whereas I provide a micro-foundation from the perspective of the entrepreneurs. When $\overline{z_3} = z_{max}$, all loans are risky, $v_r = 1$ and $p_e = p$, and the *risk shifting mechanism* is shut down. As I will demonstrate in the following section, even though *risk shifting mechanism* acts, it is not necessarily the dominant effect in the equilibrium.

Denote the aggregate loan demand elasticity $\epsilon^b = -\frac{\partial log Q^L}{\partial log(1+r^b)}$, the aggregate deposit supply elasticity $\epsilon^d = \frac{\partial log Q^D}{\partial log(1+r^d)}$, and market share of loans and deposits that each banker holds as s_i^b and s_i^d respectively. 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 above two equations indicate that the optimal loan (deposit) rate represents a mark-up (-down) over the marginal cost (benefit) of issuing the loan (deposit). As long as the bankers obtain a greater share of either the loan or deposit market, the markup or markdown will be higher. In a perfectly competitive banking sector where s_i^b and s_i^d converge to zero, there will be no mark-up (-down). The *risk shifting mechanism*, however, generates a new term in Equation (14) that lowers the markup. The reason for this is that bankers know that if they set a loan rate too high, entrepreneurs are more likely to gamble. When the banking sector becomes highly concentrated, bankers' motives to raise loan rates will be mitigated.

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

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

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, where I 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* $\{c^b_{it+1}, N_{it+1}\}_{t=0}^{\infty}$, *a sequence of aggregate quantities* $\{a_{t+1}, Q^D_t, Q^L_t\}_{t=0}^{\infty}$, *a sequence of interest rates* $\{r^b_t, r^d_t\}_{t=0}^{\infty}$, *and a sequence of price* $\{q_t\}_{t=0}^{\infty}$ *such that:*

- *(a) Entrepreneurs maximize expected life-time utility given loan rate, deposit rate and the price of capital;*
- *(b) Bankers maximize their life-time utility given constraints* (2) (3) (5) *by competing for loans and deposits;*
- *(c) Bankers choose the same quantities for all assets and liabilities;*
- *(d) Market clearing condition for*
  - *loan market:* $\sum_{i=1}^{M} Q^L_{it} = Q^L_t$;
  - *deposit market:* $\sum_{i=1}^{M} Q^D_{it} = Q^D_t$;
  - *capital market:* $\sum_{i=1}^{M} N_{it} + a_t = \overline{K}$.

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

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

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

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

When the banking sector is perfectly competitive and there is no risk taking, the model equilibrium is equivalent to Moll (2014)[37] without labor. There is no positive margin and a single cutoff determines who will be creditors and lenders. With a slight deviation ($\alpha p \lessapprox 1$) from the benchmark, bankers will charge a positive wedge between the loan rate and deposit rate, which I interpret as risk premium. Additionally, due to the inefficiency of the gambling project ($\alpha p < 1$), there will be some autarky entrepreneurs. The risk taking motive is therefore undesirable, not only because the gambling project is inefficient, but also because more resources are allocated to inefficient producers.

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

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

where v_a is defined as the fraction of autarky entrepreneurs. The relationship between bank concentration and risky capital depends on two 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 sign of the second element in equation (19) depends on how the proportion of autarky entrepreneurs is affected by bank concentration. If the banking sector is more concentrated, the net margin is wider, and thus there are more autarky entrepreneurs. I refer to this as the *"net margin mechanism"*. The validity of this channel has been established in Yi (2022)[42] by assuming that there is no risk-taking motive and a uniform distribution of productivity. This paper will demonstrate the *net margin mechanism* quantitatively in the following section. Autarky entrepreneurs who use their own money for production always invest in prudent projects. Consequently, the *net margin mechanism* leads to a negative correlation between bank concentration and risk taking.

The first term in equation (19) relates to how bank concentration affects the percentage of risky loans. According to Proposition 2, bank concentration affects the loan rate, which in turn affects the proportion of risky loans through the *risk shifting mechanism*. However, there is still a great deal of uncertainty regarding the impact of bank concentration on loan rates. The third element in equation (19) replies on the sensitivity of bank capital to the number of bankers. In the following section, I will clarify that this term is positive while quantitatively negligible.

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

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

4 Quantitative Analysis

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

4.1 Calibration

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

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

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

where the second equality follows that in the steady state of the symmetric equilibrium, bankers represent a market share of $1/M$ in the deposit market. Following Drechsler el al. (2017)[20], the average deposit market HHI in the U.S. from 1994 to 2020, which amounts to

0.1342318, is calculated as the weighted average of branch-level HHI, using branch deposits for weights. According to Equation (20), M is approximately 7.45. [5]

I have not focused on a particular distribution of productivity in the above sections. Nevertheless, the assumption that firm productivity follows a Pareto distribution has become widely accepted, building on Melitz (2003)[36]. In this paper, I use the bounded Pareto distribution instead of the Pareto distribution to satisfy Assumption 2. The bounded Pareto distribution is characterized by the shape parameter γ, the maximal value z_{max}, and the minimal value z_{min}. z_{min} is normalized to 1. I calibrate z_{max} and γ to match the dispersion of productivity and markups for the US in the sample years. As illustrated in Hsieh and Klenow (2009)[27], the difference between the 75th and 25th percentiles of TFPR is 0.53[6]. The cumulative density distribution function of $log(z)$ is $\frac{z_{min}^{-\gamma}-e^{-\gamma z}}{z_{min}^{-\gamma}-z_{max}^{-\gamma}}$ if the productivity z follows a bounded Pareto distribution[7]. γ is set to be 1.5 to keep the markup around 20% following Liu and Wang (2014)[34].[8]

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

In accordance with Basel III, the parameter κ is used to generate the implied policy requirement. Basel III requires a minimum Total Capital Ratio of 8%. With the addition of the capital conservation buffer, a financial institution is required to hold at least 10.5% percent of risk-weighted assets in capital. As I do not include the risk-based capital constraint in the benchmark model, I simply value κ at 0.08. I will demonstrate in the appendix that the main mechanisms do not change even when risk-based capital constraints are included.

One period in my model corresponds to one year. Following Gali (2005)[23] and Chris-

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

[6] Hsieh and Klenow (2009)[27] distinguishes between TFPQ and TFPR, where the use of the plant-specific deflator yields TFPQ and the use of the industry deflator yields TFPR. Due to the normalization of the price of consumption good, TFPQ and TFQR are equivalent by definition.

[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 $F_X(x) = Pr(X \leq x) = \frac{L^{-\gamma}-x^{-\gamma}}{L^{-\gamma}-H^{-\gamma}}$. Then the cdf. of Y is $F_Y(x) = Pr(Y \leq x) = Pr(log(X) \leq x) = Pr(X \leq e^x) = \frac{L^{-\gamma}-e^{-\gamma x}}{L^{-\gamma}-H^{-\gamma}}$. The probability distribution function is therefore $\frac{\gamma e^{-\gamma x}}{L^{-\gamma}-H^{-\gamma}}$.

[8] Different from Liu and Wang (2014)[34], I introduce an imperfect competition in the banking sector. In response, the markup has been raised such that the shape parameter γ does not have to be as large as in their paper.

tiano et al. (2005)[13], 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[24]). 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: Parameter Values

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

4.2 Equilibrium with only Risky Loans

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

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

The proof directly follows Equation (7), where z_3 converges to infinity when αp is close to 1. In contrast to a prudent project, the gambling project is more costly because the expected return is lower. Under the assumption of $\alpha p \lessapprox 1$, the gap between the expected return of gambling projects and that of prudent projects narrows, causing all borrowing entrepreneurs to gamble.

I set $p = 0.9$ and $\alpha = 1.05$. Figure 2 illustrates the effect of bank concentration on financial stability. As can be seen in Panel b of Figure 2, all loans are risky. Panel (a) of Figure 2 shows the bank capital to asset ratio, which far exceeds the minimum capital requirement when the banking sector is highly concentrated. However, the minimum bank capital requirement is binding when the level of bank competition is high. In accordance with Yi (2022)[42], the empirical evidence indicates that, on the one hand, the bank capital ratio has far surpassed the minimum capital requirements in the U.S., and on the other hand, bank concentration is positively correlated with bank capital. The intuition is straightforward: a more concentrated banking sector will result in a lower deposit rate, as well as a lower deposit supply. On the bank's balance sheet, both deposits and equity capital are liabilities. In a more concentrated banking sector, the bank capital ratio increases due to the substitution effect between the two objects.

Panel (c) of Figure 2 illustrates how intensifying competition among banks leads to an increase in capital allocated to gambling projects. There is a significantly negative relationship between bank concentration and stability regardless of whether the capital constraint is binding. With all loans being risky, the *risk shifting mechanism* is shut down, leaving only the *"net margin mechanism"* to shape the relationship between bank concentration and financial stability. In a more concentrated banking sector, as shown in panels (e) and (f) of Figure 2, bankers charge a higher spread, which in turn results in an increased percentage of autarky entrepreneurs. Those autarky entrepreneurs who use their initial endowments to produce would choose prudent projects. Therefore, bank concentration and stability are positively correlated.

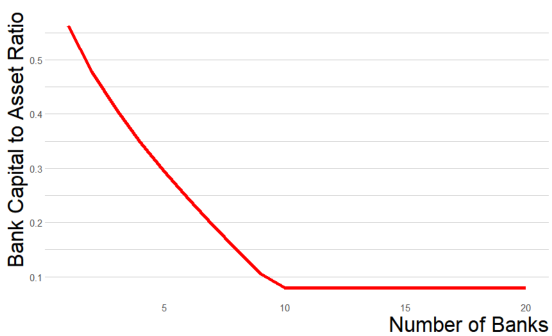
This *"net margin mechanism"* also explains how bank concentration distorts the allocation, as illustrated in panel (d) of Figure 2. As bank concentration climbs up, both the net margin and the fraction of autarky entrepreneurs rise. It is autarky entrepreneurs that are the most inefficient producers, to whom more capital is allocated through an extensive margin when the banking sector is more concentrated.

4.3 Equilibrium with Both Safe and Risky Loans

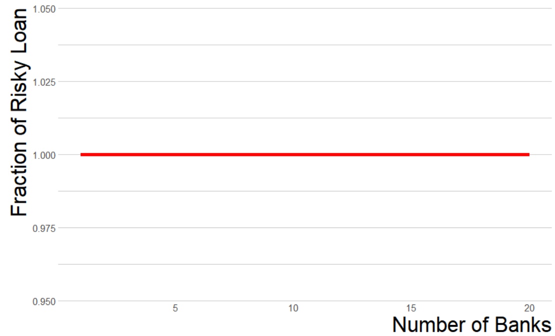
A more general case of the equilibrium is that loans could be either risky or safe. In addition to the the *net margin mechanism*, the *risk shifting mechanism* as described in Proposition 2 will reshape the relationship between bank concentration and financial stability. I set $\alpha = 1.05$ and $p = 0.7$ in this section to enable gambling and prudent projects to coexist in the steady state [9]. Figure 3 presents how bank concentration affects the loan rate.

After considering the *risk shifting mechanism*, it is noteworthy that the correlation between bank concentration and loan rate is non-monotonic. The loan rate could rise as a result of profit maximization when the bank concentration is large. More specifically, as the banking sector becomes more concentrated, the elasticity of loan demand declines. As a consequence of the *risk shifting mechanism*, however, bankers who charge a higher loan rate observe more gambling entrepreneurs and lower expected probability of loan repayment. In response, they will internalize the best response of entrepreneurs and will not set a rate that is too high. As shown in Figure 3, there is a surprisingly negative correlation between bank

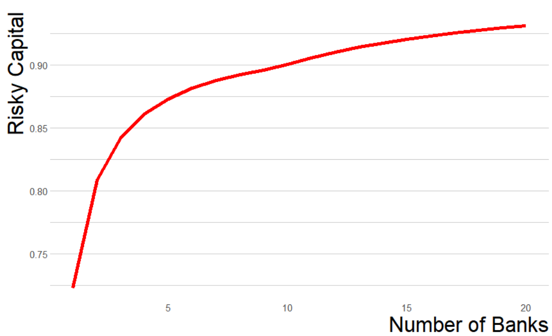
[9] Calibration is performed to match moments from data under this parameter setting.



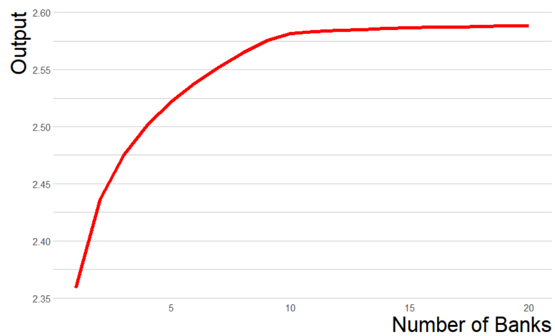
(a) Bank Capital to Asset Ratio



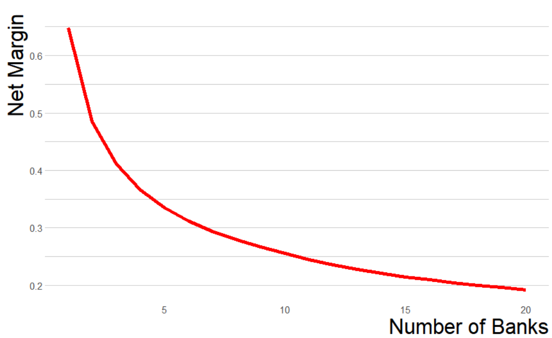
(b) Fraction of Risky Loan



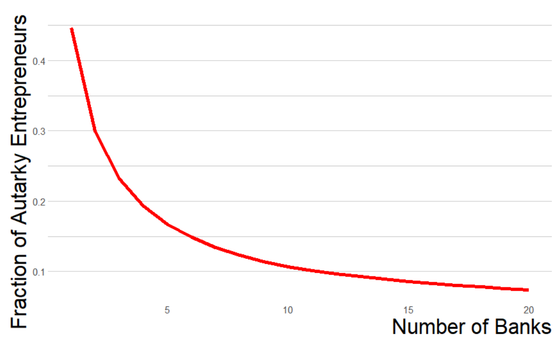
(c) Risky Capital



(d) Output



(e) Risky Capital



(f) Output

Figure 2: This plot presents the relationship between bank concentration (number of bankers) and endogenous variables: bank capital to asset ratio, fraction of risky loan, risky capital, output, net margin, and fraction of autarky entrepreneurs, when all loans are risky. I focus on the comparative statics in the steady state.

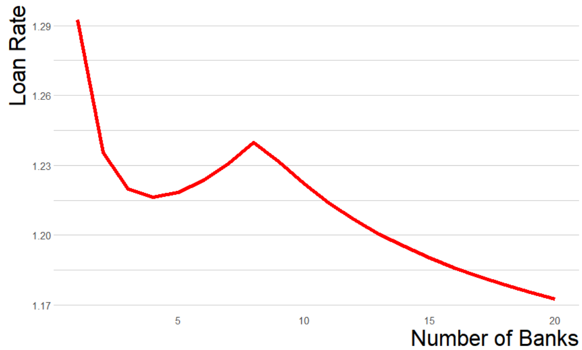


Figure 3: Bank Concentration and Loan Rate

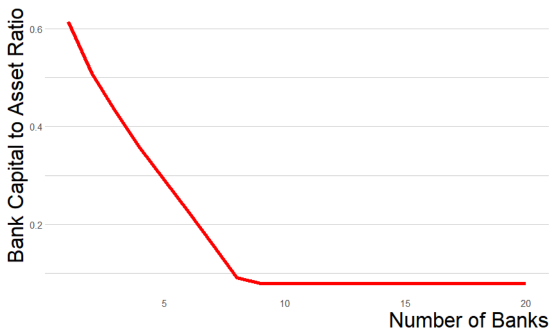
concentration and loan rate when there are approximately 4 to 8 banks in the economy.

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

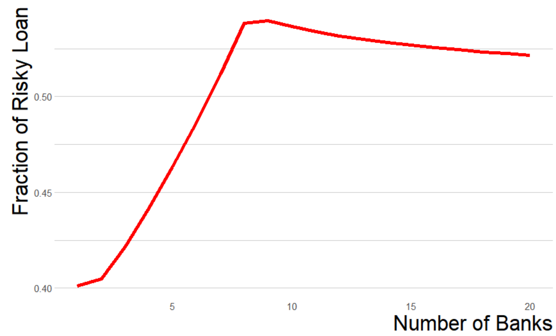
The kink in Figure 3 corresponds to the point at which banks start to accumulate excess capital, as shown in the panel (a) of Figure 4. Bankers with large market power are motivated to hold a capital ratio above the minimum requirement. This motive is further enhanced when considering the *risk shifting mechanism*. Those bankers with more capital may issue more loans, lowering the rate and the fraction of risky loans.

Panel (b) of Figure 4 shows the relationship between bank concentration and the fraction of risky loans. As the banking sector becomes more concentrated, the fraction of risky loans rises when the bank capital constraint is binding, while declines when the bank capital constraint is non-binding. When the bank capital constraint binds, bankers with a high level of market power charge high interest rates, resulting in a higher proportion of risky loans through the *risk shifting mechanism*. When bankers hold excess capital above the required minimum, large bank concentration leads to a lower loan rate and a smaller fraction of risky loans

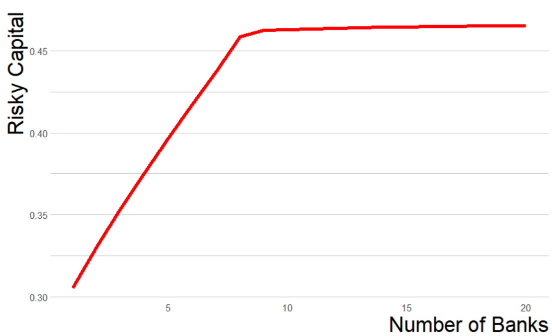
It is relevant to note that the relationship between fraction of risky loan and loan rate in the general equilibrium is non-monotonic. When bank concentration is extremely high, the general equilibrium effect through the price of capital induces the loan rate to rise. However, the return rate of loans in terms of consumption (numeriare) $q(1+r^b)$ is positively correlated with the fraction of risky loans v_r in the steady state. The relationship between $q(1+r^b)$ and M is shown in the appendix. In fact, increasing the price of capital also leads to a higher fraction of risky loans because doing so raises entrepreneurs' external funding cost.



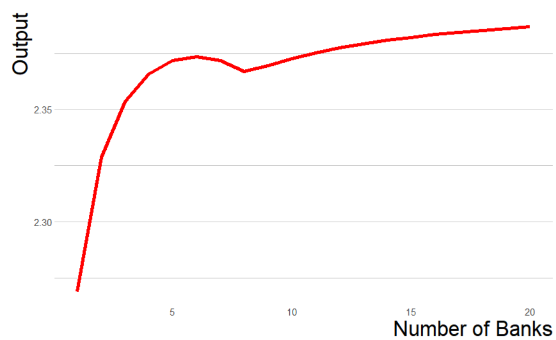
(a) Bank Capital to Asset Ratio



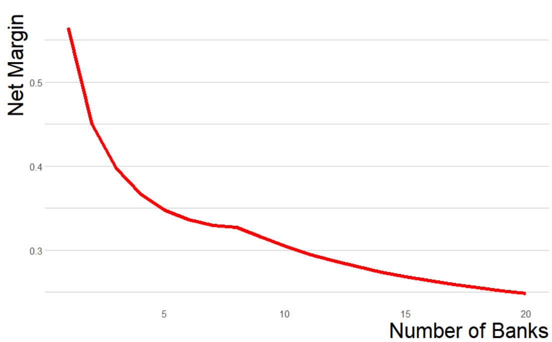
(b) Fraction of Risky Loan



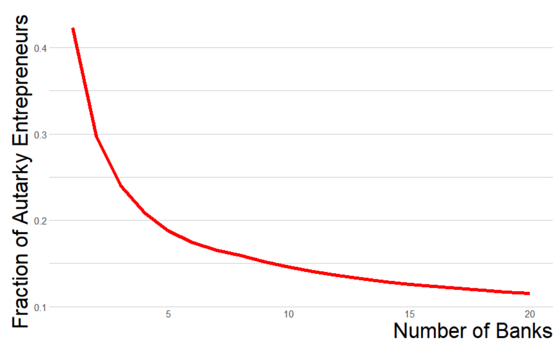
(c) Risky Capital



(d) Output



(e) Risky Capital



(f) Output

Figure 4: This plot presents the relationship between bank concentration (number of bankers) and endogenous variables: bank capital to asset ratio, fraction of risky loan, risky capital, output, net margin, and fraction of autarky entrepreneurs, when loans are either risky or safe. I focus on the comparative statics in the steady state.

4.3.1 Bank Concentration and Risk Taking

The relationship between bank concentration and risks (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, while it is somewhat uncorrelated when the bank capital constraint is binding.

When the bank capital constraint is binding, the loan rate should be higher in a more concentrated banking sector. According to the *risk shifting mechanism*, bank concentration should be positively correlated with risky capital, but this relationship is not evident in panel (c). In this regard, the first element in Equation (19)—the *net margin mechanism*—gives a compensating effect in the opposite direction: as bank concentration increases, the proportion of autarky entrepreneurs will be higher. Autarky entrepreneurs always invest in prudent projects, which results in a decline in risky capital. Based on parameters calibrated to match U.S. moments, the magnitude of the two effects is so similar that there is almost no correlation between bank concentration and risks. As long as the capital ratio exceeds the minimum requirement, however, both mechanisms lead to a negative correlation between bank concentration and risk taking.

4.3.2 Bank Concentration and Output

Panel (d) of Figure 4 illustrates how bank concentration affects output. There is a non-monotonic relationship between the two objects when both the *net margin mechanism* and the *risk shifting mechanism* are considered.

Due to the *"net margin mechanism"*, output should have surged when bank competition is more intense. When bank concentration is high, bankers charge a wider wedge between the loan and deposit rate. In this way, autarky entrepreneurs, who are also the least efficient producers, are allocated with more resources through extensive margin. In fact, Joaquim et al. (2019) demonstrate empirically that if the lending spread falls to the world average, Brazilian output will increase by five percent. Meanwhile, higher bank concentration leads to a higher bank capital ratio in the equilibrium region where bank capital constraint is non-binding. In light of the *risk shifting mechanism*, bankers have even greater incentives to accumulate excess capital ratios, thus reducing the proportion of risky loans. Considering Assumption 1, gambling projects have a lower expected payoff, which mitigates the negative impact of bank concentration on output. The negative relationship between bank concentration and output is even reversed when the number of bankers is approximately 6 to 8. As the number of banks M is calibrated at 7.45 in the above section, this local optimum may be quantitatively significant.

5 Discussions

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

5.1 Supporting Evidence

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

5.1.1 Data Description

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

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

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

RateWatch RateWatch data covers monthly loan rates at the branch level. My sample is from 1994 to 2021. For loan rates, I use one of the most common loans in the sample: auto loans (72 months)[10]. Using this strategy, I am able to eliminate issues associated with

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

observed (and unobserved) heterogeneity among loan products. I will focus on the branches that are actively involved in setting the loan rate.

Following Drechsler et al. (2017)[21], I use HHI to measure bank concentration. I first construct a county-year level HHI, which is measured by the sum of each bank institution's squared deposit market share by county for each year (Equation (20)). To obtain a bank-level HHI, I calculate the weighted average HHI of all the branches under the same bank institution, using branch deposit sizes for weights.

5.1.2 Bank Concentration and Loan Rate Revisited

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

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

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 $\alpha_{s(k)t}$ in the regression is based on Rice and Strahan (2010)[39], where they construct a state-by-state deregulation index.[11] I cluster the standard error at the bank level. The main coefficients of interest in the regression are β_i, where i=1, 2, $\cdots$, 10. The coefficients capture the differential effect of bank concentration on loan rate within deciles of bank concentration. For example, a positive β_1 implies a positive correlation between HHI and loan rate when HHI falls within $(0, 0.1]$. The model predicts that β_i will have both positive and negative values.

I control for bank fixed effect and time fixed effect in Figure 5. As illustrated in Figure 5, the coefficients are positive and significant when HHI lies in $(0, 0.6]$. This implies a positive correlation between bank concentration and loan rate in this region. In contrast, the coefficient becomes negative and significant at 1% level as HHI increases. More specifically, β_7 is -0.27. When HHI increases from 0.6 to 0.7, the loan rate drops by 0.027%. In accordance

[11] Interstate branching was not allowed in the U.S. until the Riegle-Neal Act was enacted in 1994. To mitigate the risks associated with financial institutions, the Dodd-Frank Act entered into force in 2010. However, states are allowed to use the four key provisions contained in IBBEA to restrict or increase the cost of out-of-state entry, based on which Rice and Strahan (2010)[39] construct a bank deregulation index ranging from 0 to 4, with 4 for states with the most strict requirement for entry of out-of-state banks. By adding the state-time fixed effect, I rule out the issues of different deregulation policies across states.

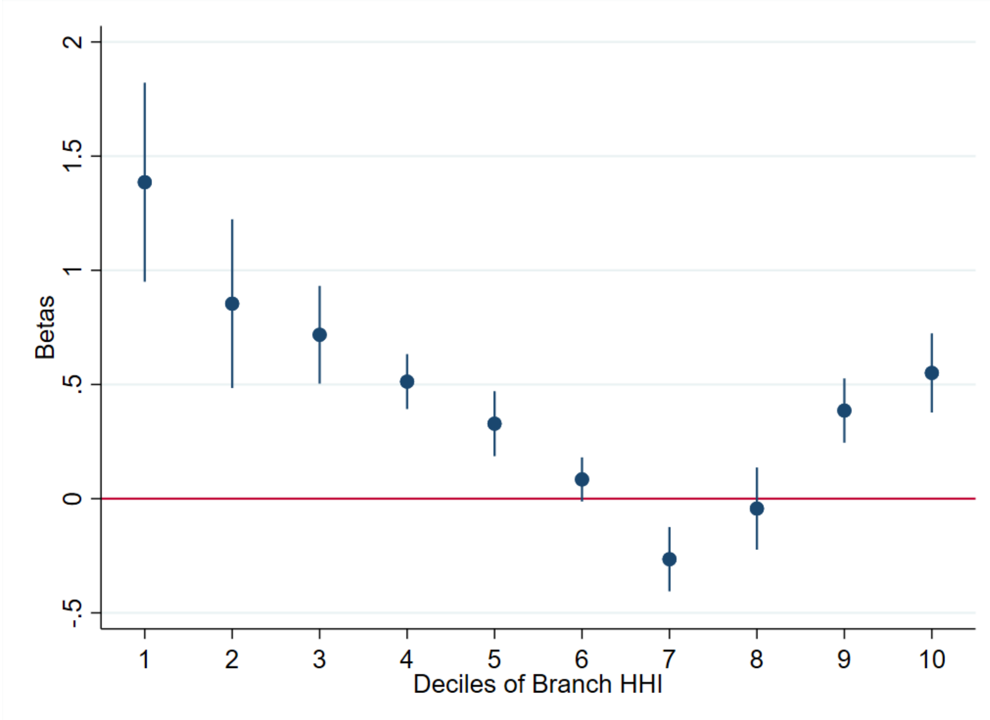


Figure 5: Branch-level HHI and Loan Rate

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.

with the model prediction, this negative correlation can be attributed to *risk shifting mechanism* and non-binding capital constraint. Increasing interest rates increase the likelihood of entrepreneurs gambling. Banks internalize entrepreneurs' decisions and dislike too high a loan rate, even in a highly concentrated banking sector. β_9 and β_{10} are significantly positive. The non-monotonic relationship between HHI and loan rate is consistent with the model equilibrium. In Appendix, I show more specifications in Table 6.

5.1.3 Interaction between Bank Concentration and Bank Capital

In accordance with the model predictions and Table 6, the relationship between bank concentration and loan rate is not monotonic. The non-monotonicity depends on whether the bank capital constraint is binding: when the bank capital constraint is binding, increasing bank concentration always results in higher loan rates, as the elasticity of loan demand falls; when banks accumulate excess capital above the minimum capital requirement, the relationship between bank concentration and loan rate is ambiguous as a result of the *risk shifting mechanism* and the general equilibrium effect. To support the model characterization, I conduct the following regression:

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

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

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

According to table 3, when bank capital ratios are low (high), the effect of bank concentration on loan rate is positive (ambiguous). Observations with high capital ratios are indicative of bankers who accumulate excess capital over the minimum capital requirement in the model. Consequently, $\beta_1 + \beta_2$ being positive and significant is due to a decrease in the elasticity of loan demand when the bank capital constraint is binding. When the bank capital constraint is non-binding, the *risk shifting mechanism* and the general equilibrium effect result in a U-shaped correlation between bank concentration and loan rate, making β_1 insignificant.

5.2 Policy Implications

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

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

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

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.

efficiency simultaneously in this section.

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

Reduced bank concentration contributes to lower efficiency and lower stability when the number of banks is approximately 6 to 8, as illustrated in Figure 4. The short-term effects of intensifying bank competition are therefore not always favorable. There is a local optimum rather than a global optimum when the number of banks reaches approximately 7. Accordingly, policymakers should be confident in reducing the obstacles to bank competition even if they observe a short-term loss of welfare. The argument is relevant because the

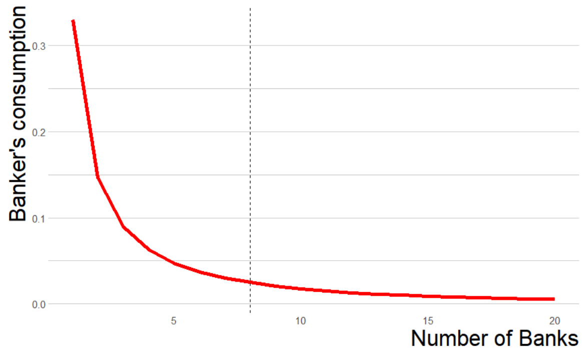


Figure 6: Number of Banker and Bankers' Consumption

calibrated number of banks in the U.S. using HHI is 7.45, which means a small deviation would have a negative but temporary impact.

5.3 Exogenous Variation of Bank Concentration

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

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

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

which equalizes the lifetime utility derived from consumption with the entry cost. It would be useful to examine the relationship between bankers' consumption and the implied number of bankers so that we understand how entry costs affect bank concentration. As illustrated in Figure 6, there is a negative correlation between bankers' consumption and the number of bankers. Consequently, rising entry costs cause few bankers to participate in the banking sector, as a highly concentrated sector assures a higher benefit to entry.

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

6 Conclusion

Throughout this paper, I develop a tractable dynamic model to investigate how bank capital affects the relationship between bank concentration and risk taking. Accumulating excess bank capital when the banking sector is highly concentrated not only enables banks to maximize their profits, but also minimizes the effect of the *risk shifting mechanism*. As a result of the *risk shifting mechanism* together with the *net margin mechanism*, there is a kinked relationship between bank concentration and risk taking, which depends on whether the minimum capital requirement is binding. The model suggests a negative correlation between bank concentration and risk when the capital ratio exceeds the minimum requirement; otherwise, bank concentration has an ambiguous but quantitatively negligible impact on risk. This paper raises concerns about future empirical studies that examine the effects of bank concentration 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. However, for a more concrete analysis, it would be valuable and necessary to explicitly model aggregate risk, in which case I could analyze the financial crisis, financial distress, etc. To keep the model tractable, all terms are real. This model with such rich heterogeneity would be useful for studying monetary policy by introducing price rigidity. I leave all these extensions for future research.

References

[1] Maria-Eleni K Agoraki, Manthos D Delis, and Fotios Pasiouras. Regulations, competition and bank risk-taking in transition countries. *Journal of Financial Stability*, 7(1):38–48, 2011.

[2] Isaac Alfon, Isabel Argimon, and Patricia Bascuñana-Ambrós. *What Determines How Much Capital is Held by UK Banks and Building Societies?* Financial Services Authority London, 2004.

[3] George-Marios Angeletos. Uninsured idiosyncratic investment risk and aggregate saving. *Review of Economic dynamics*, 10(1):1–30, 2007.

[4] Thorsten Beck, Olivier De Jonghe, and Glenn Schepens. Bank competition and stability: Cross-country heterogeneity. *Journal of financial Intermediation*, 22(2):218–244, 2013.

[5] Thorsten Beck, Asli Demirguc-Kunt, and Ross Levine. Bank concentration and crises, 2003.

[6] Thorsten Beck, Asli Demirgüç-Kunt, and Ross Levine. Bank concentration, competition, and crises: First results. *Journal of Banking & Finance*, 30(5):1581–1603, 2006.

[7] Sandra E Black and Philip E Strahan. Entrepreneurship and bank credit availability. *The Journal of Finance*, 57(6):2807–2833, 2002.

[8] John H Boyd and Gianni De Nicolo. The theory of bank risk taking and competition revisited. *The Journal of Finance*, 60(3):1329–1343, 2005.

[9] Markus K Brunnermeier and Yann Koby. The reversal interest rate. Technical report, National Bureau of Economic Research, 2018.

[10] Kenneth Burdett and Kenneth L Judd. Equilibrium price dispersion. *Econometrica*, pages 955–969, 1983.

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

[12] Mark Carlson and Kris James Mitchener. Branch banking as a device for discipline: Competition and bank survivorship during the great depression. *Journal of Political Economy*, 117(2):165–210, 2009.

[13] Lawrence J Christiano, Martin Eichenbaum, and Charles L Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, 113(1):1–45, 2005.

[14] Dean Corbae and Pablo D'Erasmo. Capital buffers in a quantitative model of banking industry dynamics. *Econometrica*, 89(6):2975–3023, 2021.

[15] Dean Corbae and Ross Levine. Competition, stability, and efficiency in financial markets. *Unpublished manuscript*, 2018.

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

[17] Gianni De Nicolò, Philip Bartholomew, Jahanara Zaman, and Mary Zephirin. Bank consolidation, internationalization, and conglomeration: Trends and implications for financial risk. *Internationalization, and Conglomeration: Trends and Implications for Financial Risk*, 2004.

[18] Mr Federico Diez, Mr Daniel Leigh, and Suchanan Tambunlertchai. *Global Market Power and its Macroeconomic Implications.* International Monetary Fund, 2018.

[19] Avinash K Dixit and Joseph E Stiglitz. Monopolistic competition and optimum product diversity. *The American Economic Review*, 67(3):297–308, 1977.

[20] Itamar Drechsler, Alexi Savov, and Philipp Schnabl. The deposits channel of monetary policy. *The Quarterly Journal of Economics*, 132(4):1819–1876, 2017.

[21] Itamar Drechsler, Alexi Savov, and Philipp Schnabl. The Deposits Channel of Monetary Policy. *The Quarterly Journal of Economics*, 132(4):1819–1876, 05 2017.

[22] Mark J Flannery and Kasturi P Rangan. What caused the bank capital build-up of the 1990s? *Review of Finance*, 12(2):391–429, 2008.

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

[24] William M Gentry and R Glenn Hubbard. Entrepreneurship and household saving, 2000.

[25] Allen C Head, Timothy Kam, Sam Ng, and Guangqian Pan. Money, credit and imperfect competition among banks. *Working Paper*, 2022.

[26] Thomas F Hellmann, Kevin C Murdock, and Joseph E Stiglitz. Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? *American Economic Review*, 90(1):147–165, 2000.

[27] Chang-Tai Hsieh and Peter J Klenow. Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, 124(4):1403–1448, 2009.

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

[29] Liangliang Jiang, Ross Levine, and Chen Lin. Does competition affect bank risk? Technical report, National Bureau of Economic Research, 2017.

[30] Gustavo Joaquim, Bernardus Ferdinandus Nazar van Doornik, José Renato Ornelas,

et al. *Bank Competition, Cost of Credit and Economic Activity: Evidence from Brazil.* Banco Central do Brasil, 2019.

[31] Nobuhiro Kiyotaki and John Moore. Liquidity, business cycles, and monetary policy. *Journal of Political Economy*, 127(6):2926–2966, 2019.

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

[33] Jiaqi Li. Imperfect banking competition and financial stability. Technical report, Technical report, Working Paper, 2019.

[34] Zheng Liu and Pengfei Wang. Credit constraints and self-fulfilling business cycles. *American Economic Journal: Macroeconomics*, 6(1):32–69, 2014.

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

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

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

[38] Rafael Repullo. Capital requirements, market power, and risk-taking in banking. *Journal of Financial Intermediation*, 13(2):156–182, 2004.

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

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

[41] David Van Hoose et al. *the Industrial Organization of Banking.* Springer, 2010.

[42] Yu Yi. Bank concentration, bank capital and misallocation. *Working Paper*, 2022.

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

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

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

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:

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

by simply aggregating all the entrepreneurs of different productivities. ■

Proof of Proposition 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:

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

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 rerun regressions that are similar to 21 and 22 with secured business loans in the RateWatch dataset. The total number of observations for secured business loans is 17,282, which is substantially less than the total number of observations for auto loans. Due to the limited data size, I run the following regression

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

where I divide the entire sample into five equal parts and include the interaction terms between HHI and quintile indicators in the regression. Table 4 illustrates that bank concentration has a positive and significant effect on the loan rate when branch-level HHI falls into the second or fifth quintile. Conversely, in other quintiles, there is no significant association between bank concentration and the loan rate.

Based on the model predictions, the effect of bank concentration on the loan rate is more likely to be significantly positive either when bank concentration is low or high. This model explains the positive correlation by considering the channel of the elasticity of loan demand as well as the general equilibrium effect of capital price. Due to the *risk shifting mechanism* in the model, however, the correlation between bank concentration and loan rate should be negative when the bank concentration is in between. The reason for not obtaining

Variables	(1) OLS	(2) OLS	(3) OLS
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0, 0.2]$)	-0.0142	-0.291	-0.256
	(0.649)	(0.698)	(0.669)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.2, 0.4]$)	0.743*	0.657	0.734*
	(0.408)	(0.470)	(0.430)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.4, 0.6]$)	0.216	0.173	0.121
	(0.361)	(0.373)	(0.433)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.6, 0.8]$)	0.250	0.0852	-0.0002
	(0.181)	(0.295)	(0.391)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.8, 1]$)	1.10***	1.12***	0.921**
	(0.363)	(0.399)	(0.424)
Constant	7.10***	7.13***	7.13***
	(0.0560)	(0.0626)	(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

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.

negative estimates in Table 4 might be that the dataset contains too much noise. There is a significant dispersion in the estimate due to the limited number of business loans. The correlation between bank concentration and the loan rate may be significantly negative if the quality of business loans is as good as that of auto loans.

To explain the mechanisms behind the non-monotonicity between bank concentration and loan rate, I run the following regression:

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

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*	0.635*	0.609
	(0.373)	(0.369)	(0.401)
Branch-HHI*Low Capital	0.539**	0.510**	0.455*
	(0.216)	(0.243)	(0.251)
Constant	7.07***	7.07***	7.08***
	(0.0272)	(0.0297)	(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

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 objects 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 D.2, these four loan types accounted for more than 80 percent of the total loan size.

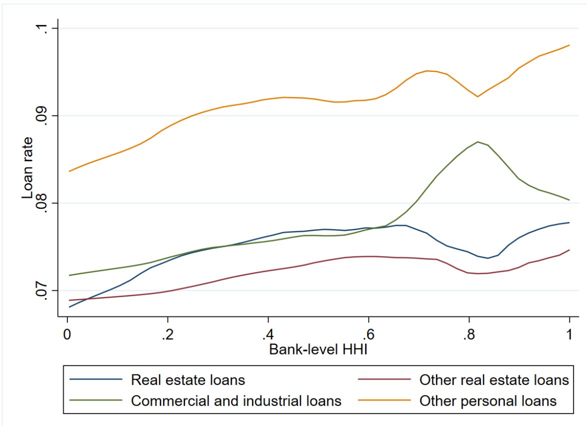
The four sub-figures capture the relationship between bank-level HHI and loan rate in years 2008, 2012, 2016, 2020, where I partially control for the time fixed effect. It is observed from the figure that the loan rate for personal loans is higher than for other loan types. Moreover, the correlation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

D Additional Tables and Figures

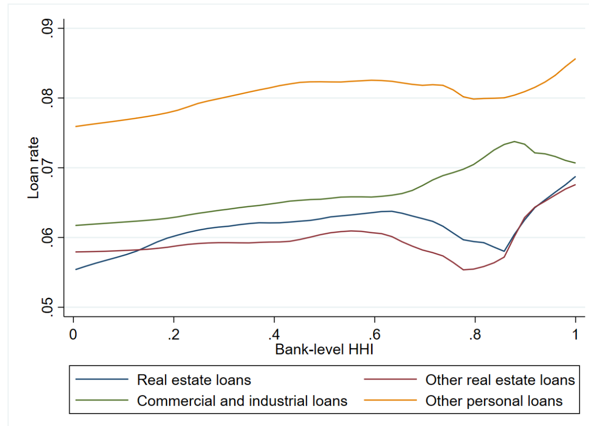
Variables	(1) OLS	(2) OLS	(3) OLS
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0, 0.1]$)	1.386*** (0.222)	0.573*** (0.153)	0.569*** (0.178)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.1, 0.2]$)	0.854*** (0.188)	0.494*** (0.114)	0.491*** (0.129)
Branch-HHI*$\mathbb{1}$(Branch-HHI$\in (0.2, 0.3]$)	0.718*** (0.109)	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

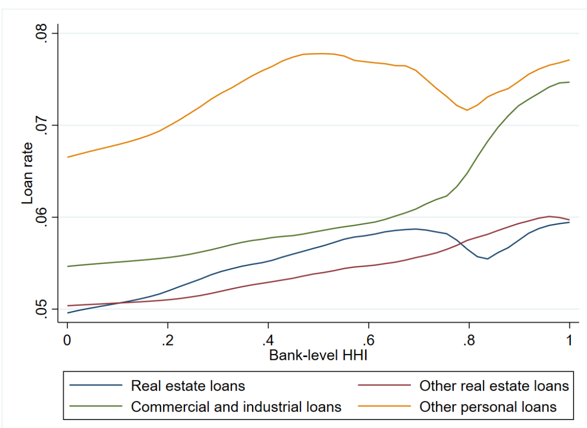
Notes: Table 6 shows the relationship between branch-level HHI and loan rate (Auto 6 years) within different deciles of HHI. The data is at the branch-quarter level and cover from January 1994 to March 2021. Rows 1-10 show the coefficients on the interaction term between HHI and the indicator of HHI 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.



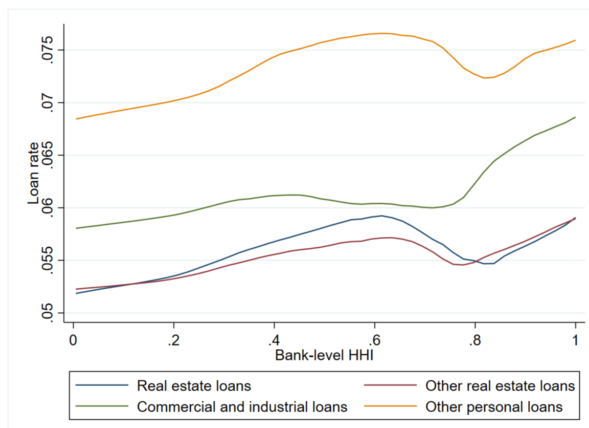
(a) Year 2008



(b) Year 2012



(c) Year 2016



(d) Year 2020

Figure C.1: 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.

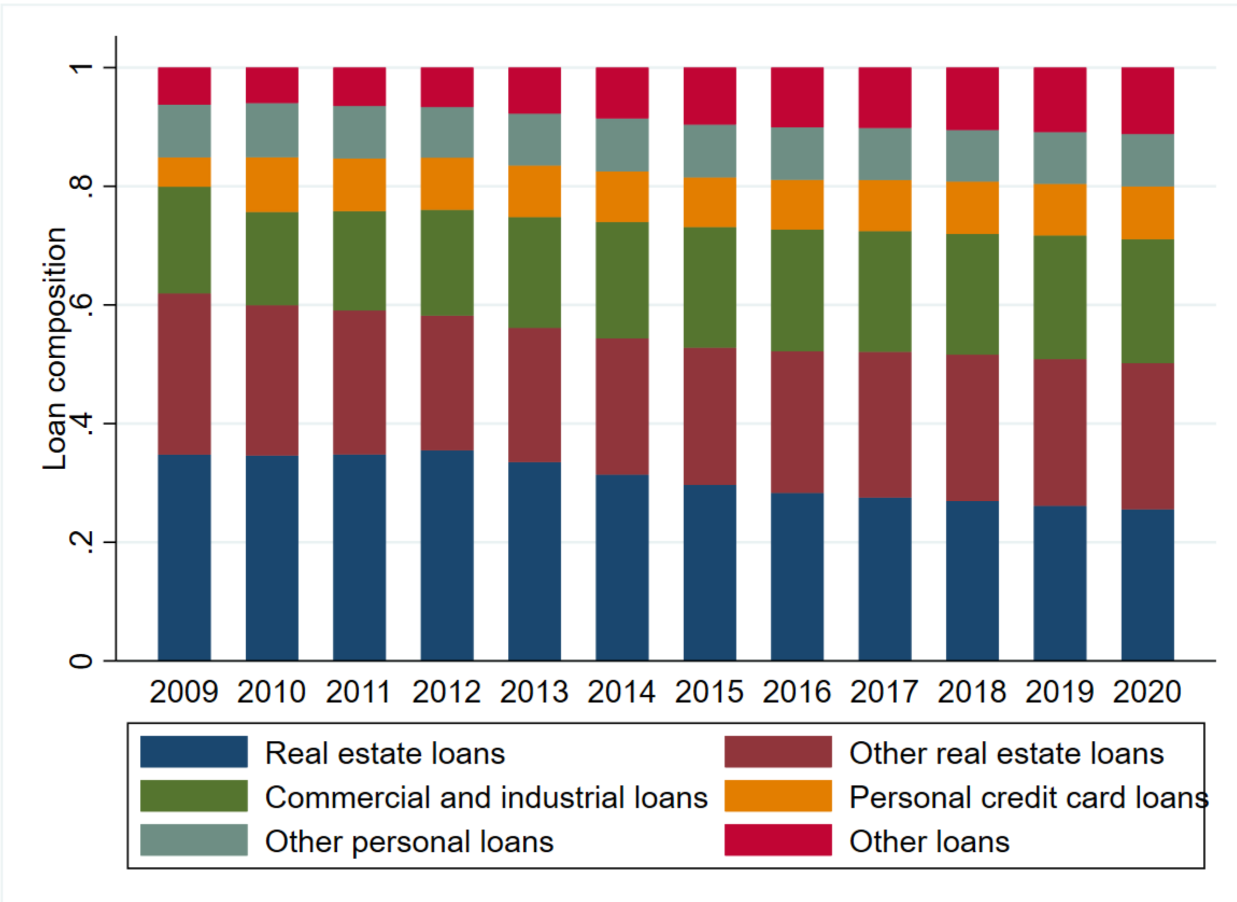


Figure C.2: Loan Composition in the U.S.

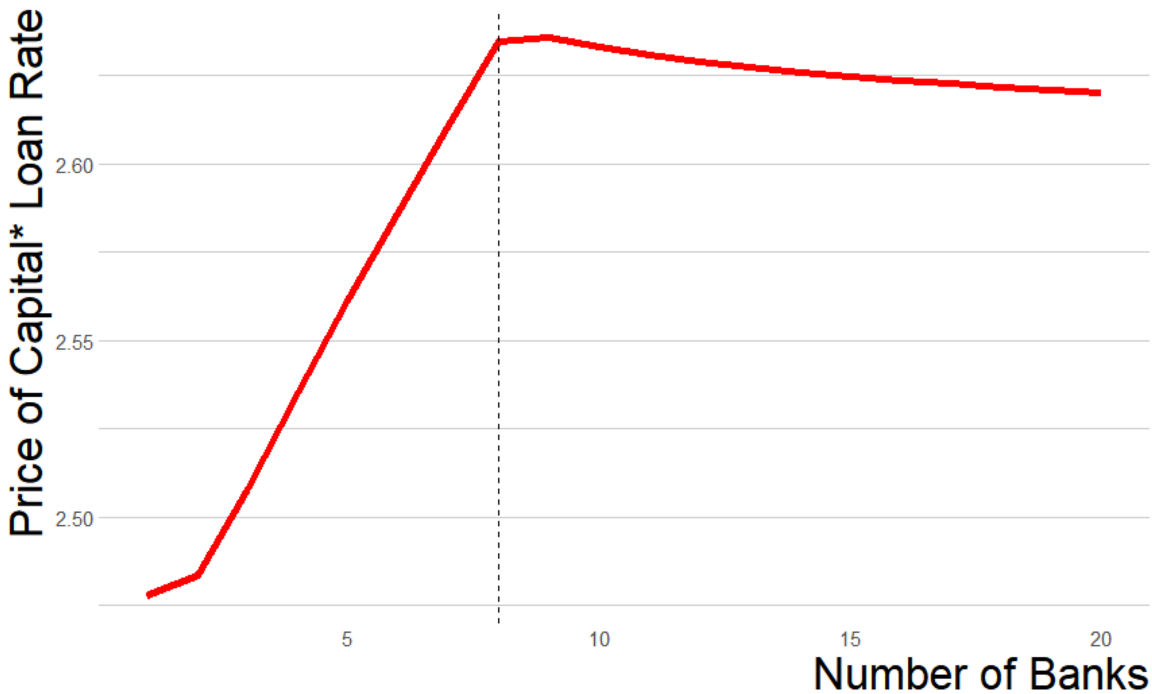


Figure D.1: Price of Capital times Loan Rate

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

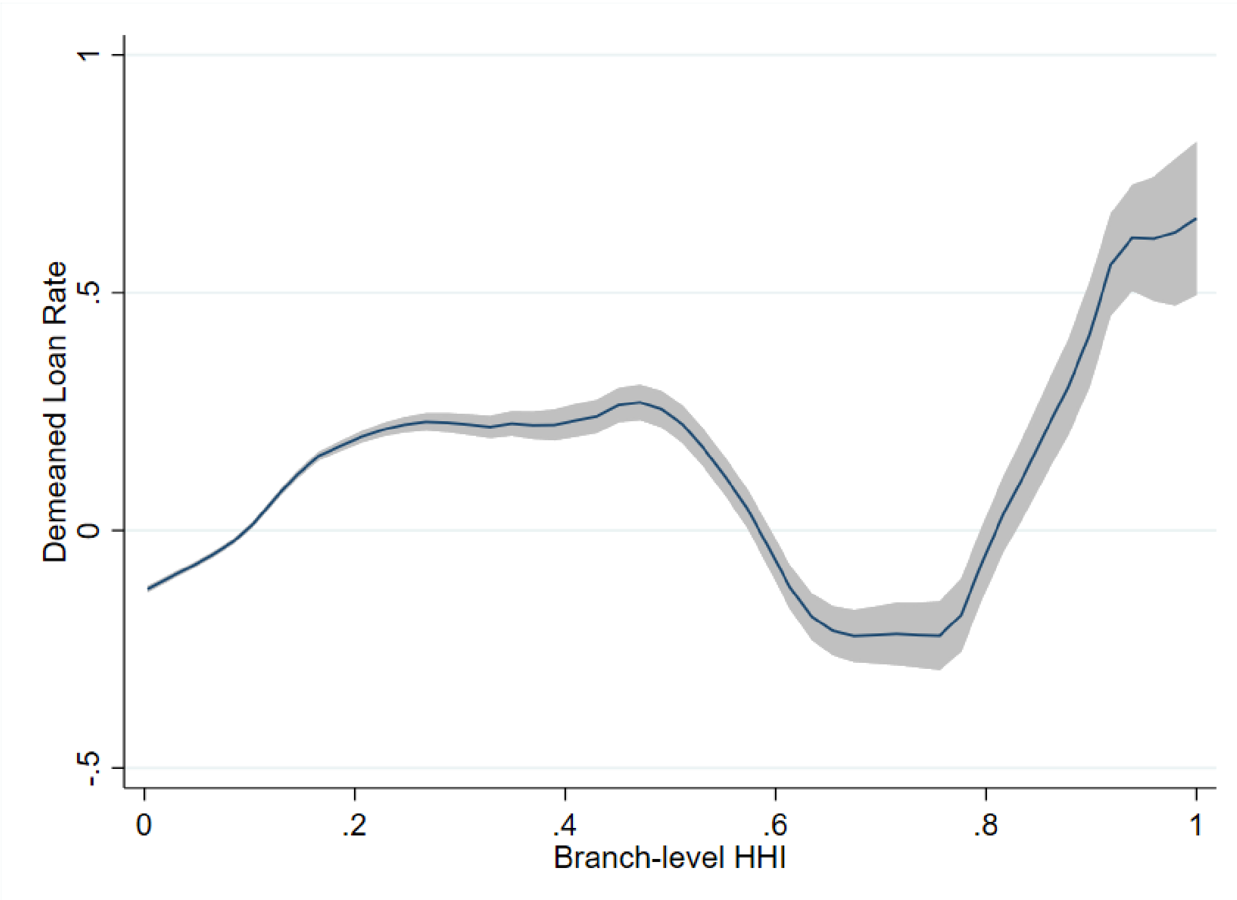


Figure D.2: Local Smoothing between Bank Concentration and Loan Rate

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.