Bank Concentration, Bank Capital, and Misallocation

Yu Yi*

April 1, 2023

Abstract

U.S. bank concentration, together with the bank capital, have been rising over the last thirty years. Based on the stylized facts, I developed a tractable dynamic model with heterogeneous financially constrained entrepreneurs and an imperfectly competitive banking sector. The model demonstrates that an increase in bank concentration leads to an increase in bank capital and potentially a non-binding capital constraint. I use the model to understand how bank concentration affects misallocation through the interaction between bank concentration and bank capital when the financial market is imperfect, which I refer to as the *"bank capital channel"*. This channel suggests that banks over-accumulate equity capital in terms of allocative efficiency, based on which I discuss implications on regulations.

Keywords: Bank Concentration, Bank Capital, Heterogeneous Agent, Misallocation

---

*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, Ben Moll, Rachel Ngai, Ricardo Reis, Tiancheng Sun, Pengfei Wang, Zhiwei Xu, seminar participants at LSE, and conference participants at AsianFA (2022), AFR (2022), AMES Tokyo (2022), ESAM (2022), MFS (2022), UZH Macro PhD Workshop (2022) for their advice.

1 Introduction

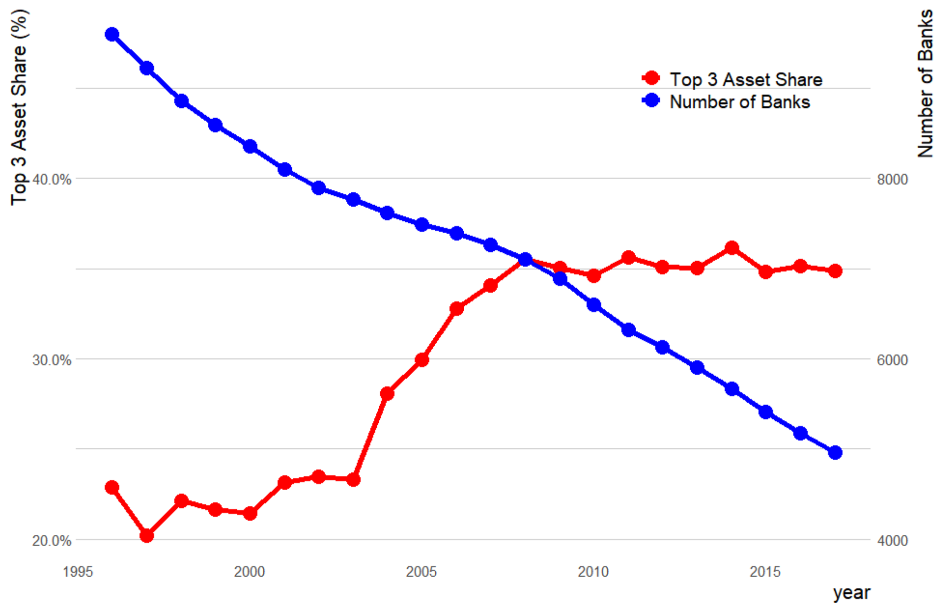
Bank concentration and bank capital are two key concepts in the banking literature, while little work has been done to illuminate their relationship. In the United States, both bank concentration and the regulatory bank capital ratio have been increasing simultaneously, as observed in Figure 1. Specifically,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

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

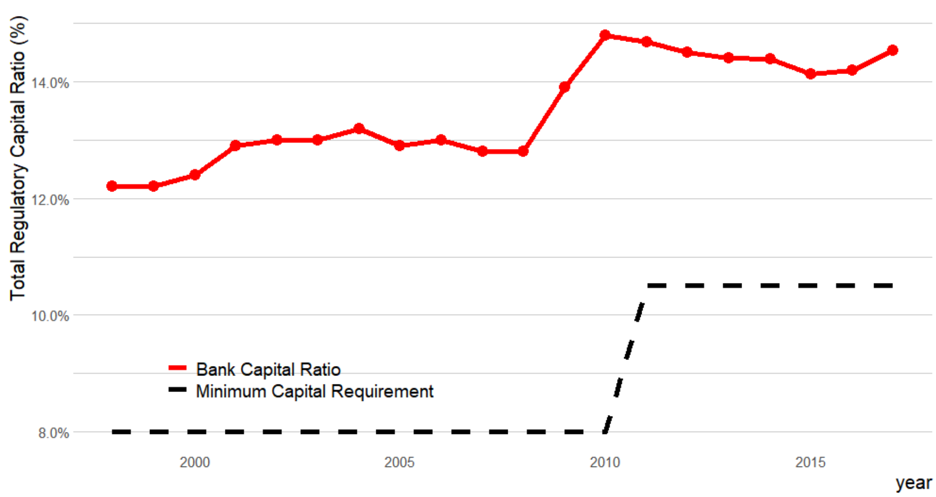
The paper presents two primary findings. Firstly, a higher level of bank concentration leads to a potentially non-binding capital requirement and a higher actual bank capital ratio. This is mainly a result of deposit market concentration. This prediction is supported by micro-level data from the US,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

There are three types of entrepreneurs in the equilibrium, who are classified as borrowing entrepreneurs, lending entrepreneurs, and autarky entrepreneurs, depending on their productivities. Entrepreneurs with the highest productivity level produce and borrow up to their limits, while those at the bottom prefer to hold all their resources in banks. Imperfect competition in the banking sector results in a positive net margin between the loan and

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



Panel A: Number of banks and top 3 asset shares



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

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

Figure 1: In panel (a), the blue line shows the number of banks in the U.S. over years, while the red the line shows the assets of three largest commercial banks as a share of total commercial banking assets over years; in panel (b), the red line illustrates the total regulatory capital ratio in the United States and the black dashed line is the minimum capital requirement.(Source: FRED)

deposit rates, which prompts some entrepreneurs (autarky entrepreneurs) to neither borrow nor lend. Instead, they use their initial holdings to engage in production activities.

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

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

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

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

Related Literature

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

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

This paper adds to the literature on the effects of micro distortions on macroeconomic outcomes (Hsieh and Klenow (2009), Bartelsman et al. (2013)). Particularly, Hsieh and

Klenow (2009) reveals significant discrepancies in the productivity of labor and capital among various agents in China and India. This capital and labor misallocation leads to a reduction in the manufacturing Total Factor Productivity (TFP). In this paper, I identify two major factors contributing to capital misallocation: financial frictions and bank concentration.

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

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

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

2 More Stylized Facts

2.1 Data Description

In this paper, I employ a combination of three distinct data sources to perform our analysis. Firstly, I utilize the Summary of Deposits data from the Federal Deposit Insurance Corporation (FDIC). Secondly, I draw upon bank balance sheet data from U.S. Call Reports, which is made available by the Federal Reserve Bank of Chicago. Finally, I extract additional bank-specific characteristics from the Research Information System (RIS) Database, also provided by the FDIC. In this section, I outline the salient features of each of these datasets.

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

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

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

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

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

2.2 Bank Capital and Bank Concentration

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

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

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

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

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

Capital to Risk Weighted Asset Ratio and Bank Concentration

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

Table 1: This table presents an estimation of the relationship between bank concentration and bank capital, using data at the bank-quarter level covering the period from 1994 to 2020. Specifically, columns (1) and (2) report results using the total capital to risk weighted asset ratio as the dependent variable, while columns (3) and (4) use the Tier 1 capital to risk weighted asset ratio. I additionally control for the return of assets in columns (2) and (4). The Standard errors are clustered at bank level. *** indicates significance at the 0.01 level.

banking sector.

3 Model Environment

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

3.1 Entrepreneurs

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

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

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

$$y_t = z_t k_t$$

Capital is assumed to fully depreciate after production.

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

$$c_t + q_t a_{t+1} \leq \Pi_t \equiv \begin{cases} z_t k_t - q_t(r_t^b + 1)(k_t - a_t) & k_t \geq a_t \\ z_t k_t + q_t(r_t^d + 1)(a_t - k_t) & k_t \leq a_t \end{cases} \tag{2}$$

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

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

$$k_t \leq \lambda a_t, \ \lambda \geq 1 \tag{3}$$

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

3.2 Bankers

The banking sector is characterized by assuming the presence of imperfect competition. Specifically, the economy is assumed to have a total of $M \geq 1$ bankers[2], each competing for the quantity of loans Q_{it}^L and deposits Q_{it}^D à la Cournot. The case where $M = 1$ represents a monopoly bank, whereas in the limit as M approaches infinity, the banking sector is perfectly competitive. At beginning of each period, each banker i is endowed with some equity capital N_{it}. Bankers are risk neutral and have preferences over dividend payouts

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

Banker serves as a financial intermediary and facilitates borrowing and lending between entrepreneurs. The loans are the sole asset on bankers' balance sheet and are financed by equity capital and deposits. Bank equity capital is accumulated through retained earnings. Table 2 summarizes the balance sheet items at the start of each period t.

Assets	Liabilities
Loans (Q_{it}^L)	Deposits (Q_{it}^D)
	Equity capital (N_{it})

Table 2: Bankers' Balance Sheet

Banker i's balance sheet identity can be expressed as

$$Q_{it}^L = Q_{it}^D + N_{it} \tag{4}$$

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

$$c_i^b + q_t N_{it+1} \leq (1 + r_t^b) q_t Q_{it}^L - (1 + r_t^d) q_t Q_{it}^D \tag{5}$$

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

[2] M is an integer.

Bankers also face a minimum capital requirement

$$N_{it} \geq \kappa Q_{it}^{L}, \tag{6}$$

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

3.3 Capital Supplier

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

4 Equilibrium Characterization

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

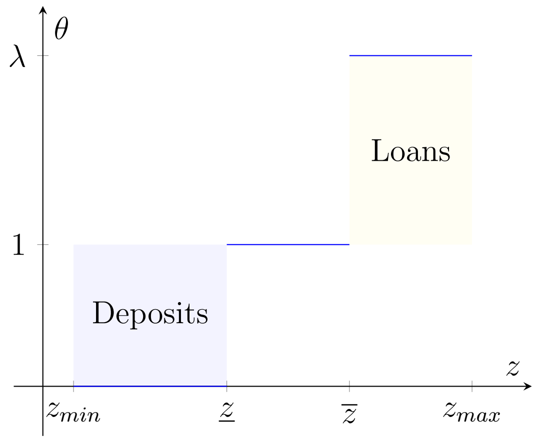
4.1 Entrepreneurs' Side

Owing to the existence of imperfect competition in the banking sector, a net margin $r_t^b - r_t^d$ is levied on the transactions conducted by the bankers. Given the deposit and loan rate, entrepreneurs' financial decisions regarding borrowing or lending are characterized in Lemma 1.

Lemma 1 *There are two productivity cutoffs $\underline{z_t}$ and $\overline{z_t}$ and the capital demand for individual entrepreneur is:*

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

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



Leverage Ratio for Different Entrepreneurs

Figure 2: The blue area is the deposit size and the yellow area is the loan size.

The cutoff property relies heavily on the constant return to scale of the production function. According to Lemma 1, the optimal capital demand decision is at corners: it is zero for entrepreneurs with low enough productivity, maximum amount allowed by the borrowing constraint for those with high enough productivity and initial wealth for those with intermediate level productivity. There are two types of marginal entrepreneurs. For the entrepreneurs with productivity $\underline{z_t}$, the return of each additional unit of capital investment $\frac{\underline{z_t}}{q_t}$ equals the opportunity cost of not depositing that in the bank $r_t^d + 1$; while for those with productivity $\overline{z_t}$, the return of each additional unit of capital investment $\frac{\overline{z_t}}{q_t}$ equals the cost of acquiring that unit $r_t^b + 1$. This heterogeneity in productivity among entrepreneurs generates an endogenous loan demand and deposit supply in the economy, as illustrated in Figure 2.

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

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

$$K_t = \int a_t dH_t(a_t, z_t) \tag{7}$$

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

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

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

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

Lemma 2 *Denote Q_t^L and Q_t^D as the loan size and deposit size respectively. Aggregate quantities $\{Q_t^L, Q_t^D, K_{t+1}\}$ satisfy:*

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

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

$$\begin{aligned} q_t K_{t+1} = \beta\{ \int_{z_{min}}^{\underline{z}} q_t(1 + r_t^d) dG(z_t) + \int_{\overline{z_t}}^{z_{max}} \lambda[z_t - q_t(1 + r_t^b)] \\ + q_t(r_t^b + 1) dG(z_t) + \int_{\underline{z_t}}^{\overline{z_t}} z_t dG(z_t)\} K_t \end{aligned} \tag{10}$$

Equation (8) reveals that the aggregate loan demand is determined by three key factors: the fraction of borrowing entrepreneurs, the borrowing limit and entrepreneurs' initial capital holding. Similarly, the deposit supply is contingent upon the initial capital of lending entrepreneurs, as described by equation (9). Meanwhile, the law of motion for aggregate capital is encapsulated by equation (10). The future wealth of entrepreneurs, $q_t K_{t+1}$, depends on the saving rate β and the net return of entrepreneurs. Specifically, the three terms contained within the brackets on the right-hand side of equation (10) represent the return rates of depositors, borrowers, and autarky entrepreneurs, respectively. Notably, the constant saving rate across all entrepreneurs stems from the log utility functional form and the constant return to scale production function.

4.2 Bankers' Side

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

$$q_t(1 + r_t^d) = \frac{\epsilon_t^d}{\epsilon_t^d + s_{it}^d} q_t \mu_{it} \tag{11}$$

$$q_t(1 + r_t^b) = \frac{\epsilon_t^b}{\epsilon_t^b - s_{it}^b} (\mu_{it} + \kappa q_t \chi_{it}) \tag{12}$$

where $q_t \mu_{it}$ is the multiplier on the balance sheet identity and $q_t \chi_{it}$ is the multiplier on the bank capital constraint. Equation (11) specifies that the deposit rate is determined solely by the marginal benefit of issuing deposits, which is the multiplier on the balance sheet identity. Meanwhile, the marginal cost of issuing loans is reflected in equation (12), as it tightens both the balance sheet identity and the capital constraint by κ. Moreover, the loan and deposit rates are influenced by mark-up and mark-down, which are functions of loan demand elasticity ϵ_t^b, deposit supply elasticity ϵ_t^d, and market shares of loans s_{it}^b and deposits s_{it}^d held by each banker. The Euler equation (13) is derived from the optimal condition for bank capital:

$$q_t = \beta q_{t+1}(\mu_{it+1} + \chi_{it+1}) \tag{13}$$

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

4.3 Steady State Equilibrium

In this section, I will define the symmetric equilibrium and subsequently focus on the steady state. To derive an analytical solution, I impose $\kappa = 0$, and assume a uniform distribution of productivity $U[z_{min}, z_{max}]$. These modeling assumptions do not alter the primary findings of this paper.

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

- *(a) Each entrepreneur maximizes life-time utility given loan rate, deposit rate and the price of capital;*
- *(b) Bankers maximize their life-time utility given* (4), (5), (6) *by competing for loans and deposits;*
- *(c) Bankers choose the same quantities for all assets and liabilities;*

- *(d) Market clearing condition for*
  - *loan market:* $\sum_{i=1}^{M} Q_{it}^{L} = Q_t^L$*;*
  - *deposit market:* $\sum_{i=1}^{M} Q_{it}^{D} = Q_t^D$*;*
  - *capital market:* $\sum_{i=1}^{M} N_{it} + K_t = \overline{K}$*.*

Lemma 3 *Proportion of the autarky entrepreneurs is* $\frac{1}{M+1}$

The implication of Lemma 3 is that the presence of autarky entrepreneurs is contingent upon the level of competition in the banking sector. This is intuitively plausible since banks tend to charge a higher net margin in the presence of high bank concentration, thereby increasing the proportion of autarky entrepreneurs. This straightforward outcome enables me to concentrate on the conduct of borrowing entrepreneurs and lending entrepreneurs in the equilibrium.

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

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

Proposition 1 suggests that the presence of financial constraints and imperfect competition in the banking sector affect agents' incentives to accumulate capital. To comprehend the mechanics behind Proposition 1, it is crucial to examine the primary sources of friction in the model, namely, the imperfect financial market and imperfect competition in the banking sector.

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

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

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

Referring back to Proposition 1, it follows that the capital constraint is not binding if the borrowing limit exceeds $\overline{\lambda}(M)$. Moreover, the monotonicity of the cutoff $\overline{\lambda}(M)$ in M indicates that the capital constraint is not binding when the banking sector is highly concentrated. When the capital constraint does not bind, higher bank concentration and borrowing limits result in a higher bank capital to asset ratio. To understand the positive relationship between financial market perfection and bank capital, one should consider the proportion of borrowing entrepreneurs. As the borrowing limit increases, borrowing entrepreneurs can obtain more loans, reducing both the proportion of borrowers and the heterogeneity of borrowing entrepreneurs. Consequently, bankers can extract a higher return from the borrowing entrepreneurs, which encourages them to accumulate more capital. The primary mechanism driving the positive correlation between bank concentration and bank capital is that in a more concentrated banking sector, the deposit rate decreases, which reduces deposit supply. Banks can raise funds for investment through deposits or capital. The substitution effect between the two liabilities increases bank capital. In the next section, I will provide a quantitative explanation for why the concentration in the deposit market dominates even when there is also loan market concentration.

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

$$Y = Z\overline{K} = (uE[z|\underline{z} \leq z \leq \overline{z}] + \lambda v E[z|z \geq \overline{z}])(\overline{K} - N), \tag{14}$$

where Y represents aggregate output, and Z denotes the average productivity of the economy. The first equality follows directly from the linear production function. In Equation (14), the proportion of autarky entrepreneurs is denoted by u and the proportion of borrowing entrepreneurs is denoted by v. The equation's second equality indicates that only entrepreneurs are capable of producing. The productivity of entrepreneurs is determined by five factors: the weighted average productivity of autarky entrepreneurs with a productivity level between $\underline{z}$ and $\overline{z}$, represented as $E[z|\underline{z} \leq z \leq \overline{z}]$, multiplied by their proportion u, plus the weighted average productivity of borrowing entrepreneurs with a productivity level greater than or equal to $\overline{z}$, represented as $E[z|z \geq \overline{z}]$, multiplied by their leverage ratio λ and their proportion v. At the beginning of each period, the family of entrepreneurs possesses $\overline{K} - N$ units of capital.

Proposition 2 *Suppose λ is finite. As the bank concentration $\frac{1}{M}$ rises, output falls.*

As in previous discussions, the implications of Proposition 2 are discussed with respect to the two primary market frictions. In a frictionless market, output should be $z_{max}\overline{K}$. However, in an incomplete financial market with a perfectly competitive banking sector, the decentralized equilibrium becomes inefficient, as not all capital is allocated to the most productive entrepreneurs. The introduction of imperfect competition in the banking sector further distorts efficiency in terms of output. Bankers with larger market power would charge a wider net margin. In Lemma 3, it is demonstrated that a larger net margin leads to a higher proportion of autarky entrepreneurs, who are characterized as the most inefficient producers, resulting in decreased output. This mechanism is referred to as the "*net margin channel*". Empirical research conducted by Joaquim et al. (2019) has examined this channel, indicating that a rise in bank competition and a reduction of spread in Brazil to global levels could yield an output increase of approximately 5%.

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

The question arises as to whether the "*net margin channel*" is the sole transmission mechanism through which bank concentration affects the efficient allocation. Proposition 1 illustrates a positive correlation between bank capital and bank concentration, which raises the possibility that this relationship may also have a bearing on aggregate output. In order to shed light on this issue, the central planner's problem will be analyzed in the subsequent section. This analysis will offer a deeper understanding of the interplay between bank capital, bank concentration, and the broader macroeconomic performance.

4.4 Optimal Capital Allocation in Production

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

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

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

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

where the first equality is derived from the definition stipulated in Equation (14), while the second equality is given by Lemma 1 and social planner's choice. This equation illustrates that the average productivity of the economy can be expressed as the weighted average productivity of the autarky entrepreneurs and the borrowing entrepreneurs, with a weight of $\frac{1}{M+1}\frac{1}{\kappa_0+1}$ that depends on the bank concentration and social planner's choice.

Proposition 3 (Optimal Capital Allocation) *Denote κ_0^* as the optimal ratio of $\frac{N}{K}$. There exists an optimal capital allocation that satisfies:*

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

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

Proposition 3 proffers valuable insight into the manner in which bank concentration and borrowing limits interact with optimal capital allocation. The optimal capital allocation is positively correlated with bank concentration, as suggested by Proposition 3. This is attributable to the advantage of higher bank capital, which reduces the amount of capital held by autarky entrepreneurs. An increase in bank concentration results in a higher proportion of autarky entrepreneurs, leading to a greater benefit when bank capital increases. Moreover, Proposition 3 implies that a higher borrowing limit results in a higher optimal bank capital. An increase in the borrowing limit has the effect of mitigating the distortion caused by higher bank capital on the average productivity of both autarky entrepreneurs and borrowers. Consequently, the optimal position for bank capital is increased.

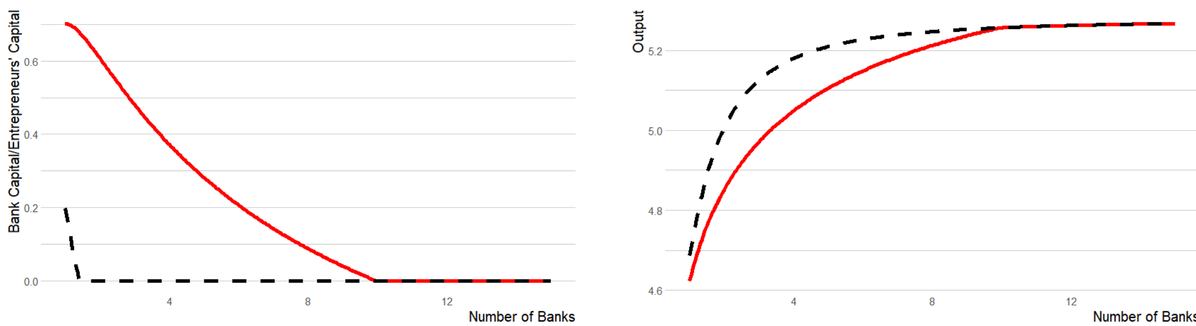
4.5 Bank Capital Channel

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

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

Proposition 4 *Bankers are over-accumulating capital:* $\frac{N^*}{K^*} > \kappa_0^*$

Figure 3 portrays the implications of Proposition 4. The red line signifies the capital ratio and output in the decentralized equilibrium on the left panel, while the right panel depicts the same variable in the centralized equilibrium using a black dashed line. The decentralized equilibrium exhibits a consistently higher level of capital ratio output than the centralized equilibrium, with the intersection of the two lines on the left panel occurring when the natural capital constraint becomes binding. Therefore, a positive wedge arises between optimal output and output in the decentralized equilibrium when the capital ratios in the two scenarios differ.



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

Centralized Equilibrium v.s. Decentralized Equilibrium

Figure 3: The red line is decentralized capital ratio (output) between bankers and entrepreneurs, while the dashed black line is optimal one.

According to Proposition 4, excessive levels of bank capital can lead to inefficiencies in allocation. To elucidate this point, it is beneficial to examine the differences between the objectives of bankers and social planner and identify the pecuniary externalities. Central planner maximizes the output, which is expressed as::

$$Y = uKE[z|\underline{z} \leq z \leq \overline{z}] + vK\lambda E[z|z \geq \overline{z}] \tag{15}$$

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

$$\begin{aligned} c_b &= qr^b(\lambda - 1)vK - qr^d(1 - v - u)K \\ &= vK(\lambda - 1)\frac{r^b}{1 + r^b}\overline{z} - (1 - v - u)K\frac{r^d}{1 + r^d}\underline{z} \end{aligned} \tag{16}$$

The initial row in equation (16) reveals that consumption is subject to the net return of loans and the costs associated with deposits, while the second row is obtained through the replacement of the loan and deposit rates with the productivity of marginal entrepreneurs.

Given the substantial disparity in the objectives of the central planner and the bankers, determining the cause of bank capital overaccumulation may not be a straightforward process. Therefore, it would be useful to compare the factors present in both equations and assessing how differences in each component contribute to distinct motives.

Both bankers and the social planner reap benefits from lending activities. Bankers earn $vK(\lambda-1)\frac{r^b}{1+r^b}\overline{z}$ on loan lending, while the social planner values loans as a means of providing resources to more productive entrepreneurs, reflected in the second element in Equation (15). These two elements differ in three ways. Firstly, bankers place value on profits solely based on loan size, while the social planner values returns from both loans and the self-investment of borrowers, denoted by $\lambda - 1$ and λ, respectively. Secondly, the social planner is not subject to a capital cost when issuing loans, denoted by $\frac{r^b}{1+r^b}$, while the cost for bankers is 1. Lastly, the return on lending loans for bankers is based on the productivity of marginal entrepreneurs, who are indifferent between borrowing and staying autarky. In contrast, the social planner bases their returns on the average productivity of borrowers, represented by $\overline{z}$ and $E[z|z \geq \overline{z}]$, respectively. Accumulated bank capital leads to an increase in lending activities in both centralized and decentralized equilibria, but the social planner derives a higher return and incurs lower costs from lending activities relative to bankers. This inherent conflict between output and profit prompts the social planner to accumulate more bank capital than bankers, thus generate opposite implication to Proposition 4.

Bankers are not concerned with the behavior of the autarky entrepreneurs, which appears in the central planner's problem. (the first element in Equation (15)). Social planner recognizes that accumulating more bank capital entails allocating fewer resources to autarky entrepreneurs. This process follows the same mechanism as Proposition 3, where a higher bank capital corresponds to lower initial entrepreneurial capital and lower initial capital holdings for the autarky entrepreneurs. Hence, the social planner take this into account and accumulate more capital than bankers do.

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

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

5 Quantitative Analysis

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

5.1 Calibration

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

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

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

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

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

$$HHI = \sum_{i=1}^{M} {s_i^d}^2 = \sum_{i=1}^{M} (\frac{1}{M})^2 = \frac{1}{M} \tag{17}$$

where the first equality follows the definition of of HHI and the second equality follows that in the steady state of the symmetric equilibrium, all the bankers constitute $1/M$ market share in both deposit market and loan market. The average HHI in the US from 1994 to 2020 is calculated as the weighted average of branch-level HHI, using branch deposits for the weights, and amounts to 0.1342318. Using Equation (17), I obtain $M = 7.45$. By matching the model's implied bank capital to asset ratio with that of the US in years between 2001 and 2017, I choose λ. A higher value of λ indicates a more efficient financial market in the economy. The model's bank capital to asset ratio is $1 - \frac{\sqrt{1+\lambda \frac{M^2}{2M+1}}}{\lambda - 1}$, while the average bank regulatory capital to risk-weighted assets for the US in years between 2001 and 2017, according to FRED, is 13.71%. Given M, we obtain an implied value of $\lambda = 6.74$.

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

One period in my model corresponds to one year. Following Gali (2005)Gali and Monacelli (2005) and Christiano et al. (2005)Christiano et al. (2005), the discount factor β is calibrated at 0.96, which implies a riskless annual rate of about 4% in the steady state. Additionally, a depreciation rate of $\delta = 0.1$ is adopted to more realistically account for the capital's wear and tear, resulting in an annual depreciation rate of 10%. The baseline model requires a modification, whereby capital suppliers do not provide an exogenous amount of capital each period, but rather the aggregate capital remains constant at $\overline{K}$, normalized to 1. The calibration of all parameters is summarized in Table 3.

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

Table 3: Calibrated Parameter Values

5.2 Policy Implications

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

5.2.1 Deposit Rate Floor

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

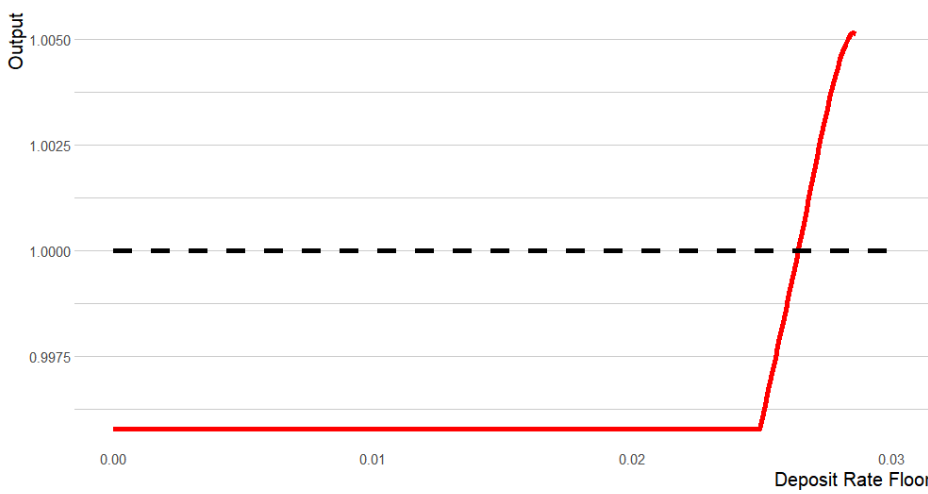
$$\tilde{r}_t^d = Max\{\overline{r}, r_t^d\} \tag{18}$$

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

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

Using the calibrated parameters, Figure 4 shows the impact of the deposit rate floor on bank capital ratio (panel A) and output (panel B). The results reveal that when the deposit rate floor is low, the decentralized equilibrium is attained, and the capital ratio remains significantly higher than the minimum capital requirement, whereas the output is

Panel A: Capital ratio under different deposit rate floors



Panel B: Output (TFP) under different deposit rate floors
Effects of the Deposit Rate Floor

Figure 4: Capital ratio and output under different deposit rate floor is depicted with red lines. The black dashed line illustrates the output level in the central planner's problem, which is normalized to 1.

low. As the policy becomes more restrictive, deposit rate floors begin to take effect, causing a substantial decline in the capital ratio and an increase in output. The responses of output and capital ratio to the deposit rate floor is such that an increase in the deposit rate floor from 2.5% to 2.87% raises output by 1 percent and brings the capital ratio to the minimum requirement. The underlying intuition is straightforward: the deposit rate floor restricts the benefits of holding capital, leading to a decline in bankers' capital and an increase in their output.

In panel (b) of Figure 4, the optimal output level (black dashed line) is normalized to 1. This level is achieved when the social planner allocates a portion of capital to bankers to ensure that the bank capital to asset ratio meets the minimum capital requirement. It is worth noting that the optimal level of banking capital should ideally be zero. Nonetheless, in this instance, an effort has been made to make the centralized and decentralized equilibria comparable. Notably, it is observed that the deposit rate floor can result in an even higher output level than in the social planner's problem. This is due to the deposit rate floor's dual impact of forcing the capital ratio to an efficient level and simultaneously reducing the spread between the deposit rate and loan rate. The resulting decrease in the net margin implies a lower proportion of autarky entrepreneurs and, ultimately, a higher output level.

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

5.2.2 Transaction Cost of Bank Capital

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

$$c_{it}^b + q_t N_{it+1} + C(N_{it+1}) \leq (1 + r_t^b) q_t Q_{it}^L - (1 + r_t^d) q_t Q_{it}^D \tag{19}$$

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

where $C(N_{it+1})$ is the cost must be incurred in the process of accumulating N_{it+1} amount of capital. A linear form $C(N) = cN$, where c is an exogenous constant, is assumed for the transaction cost. As a result, Equation (13) can be rewritten as follows:

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

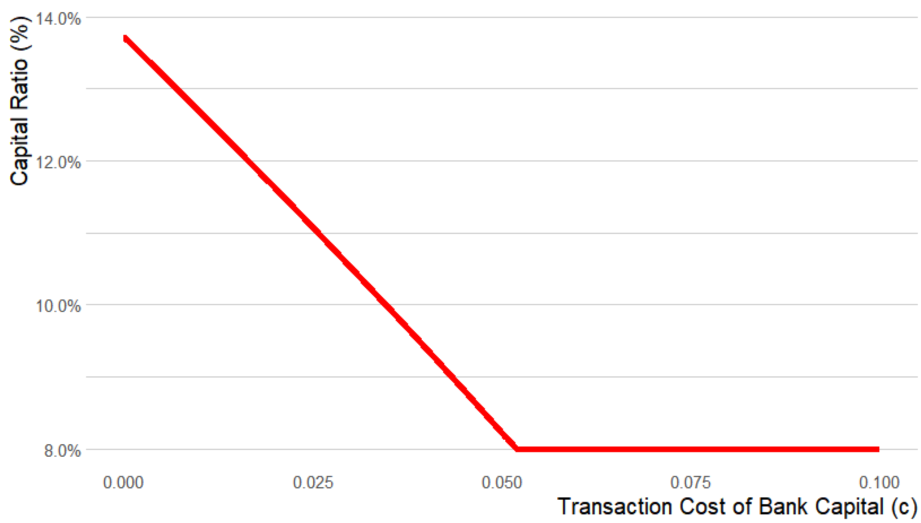
The introduction of transaction costs associated with bank capital may result in reduced motivation of capital accumulation for bankers. Figure 5 depicts the impact of transaction costs on the capital ratio and aggregate output (consumption) under various scenarios. As indicated in panel A of Figure 5, when the cost of holding capital for bankers escalates, the bank capital ratio declines. As the value of c increases from 0 to 0.053, the capital requirement becomes binding.

Panel B of Figure 5 illustrates the correlation between transaction costs and aggregate output. The panel reveals that aggregate output experiences a rise of 0.04% upon the capital requirement becoming binding. However, as the transaction cost increases further, output declines. This can be attributed to the fact that an increase in transaction cost leads to a higher loan rate, which causes a greater proportion of autarky entrepreneurs and decreases output. Additionally, this mechanism also explains why the introduction of transaction costs related to bank capital fails to attain the output level in the centralized equilibrium. Specifically, as depicted in panel B of Figure 5, the red line representing output consistently falls below the black dashed line representing optimal output.

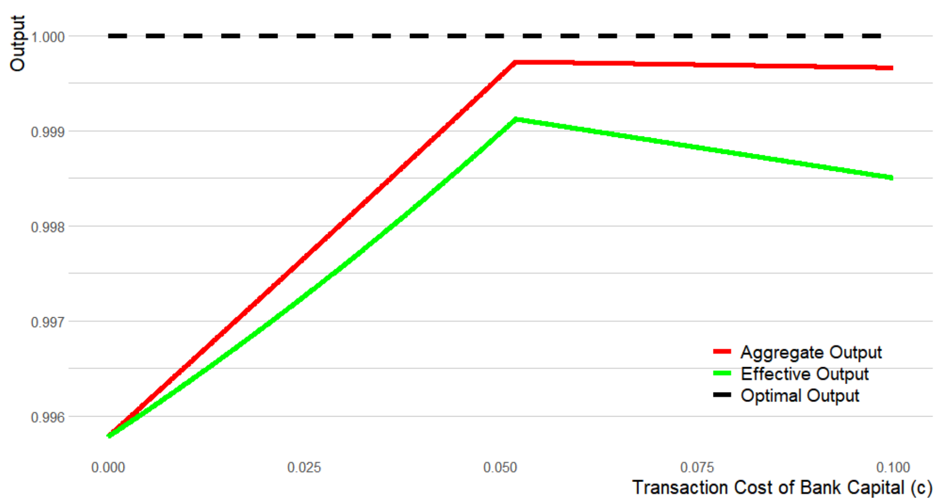
The effective output is also depicted in panel B of Figure 5, which is defined as aggregate output minus the resources that cannot be consumed or saved. Previously, aggregate output and effective output were indistinguishable in the absence of transaction costs on bank capital, as shown in panel B of Figure 5 where the red and green lines coincide for $c = 0$. However, as the transaction cost of bank capital increases, more resources are allocated to its accumulation, and the difference between aggregate and effective output expands. Although effective output follows a similar pattern as aggregate output, it might provide a more precise indication of welfare in this context. Both aggregate and effective output attain their maximum level when the bank capital ratio satisfies the capital constraint.

5.2.3 Capital Requirement Ceiling

Following the financial crisis in 2008, policymakers implemented a minimum bank capital requirement to address the issue of risk exposure. High leverage ratios are known to incentivize banks to take risks, and the imposition of a bank capital requirement serves to mitigate these incentives by putting the bank's equity capital at risk. This paper analyzes



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



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

Figure 5: Capital ratio and output under different deposit rate floor is depicted with red lines. The black dashed line illustrates the output level in the central planner's problem, which is normalized to 1. Green line illustrates the effective output (output minus transaction cost)

safe investments with different returns and, in the absence of risk considerations, examines the impact of changes in bank capital levels on the allocation of resources across different projects, providing insights into the role of bank capital in promoting allocative efficiency.

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

Proposition 5 *When there is no minimum capital requirement and a zero capital requirement ceiling, the capital allocation in the decentralized equilibrium is efficient.*

The proof directly follows by the argument above. Both the capital ratio and output in the decentralized equilibrium is the same as that in the social planners' problem.

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

5.3 Disentangling Bank Deposit and Loan Market Concentration

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

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

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

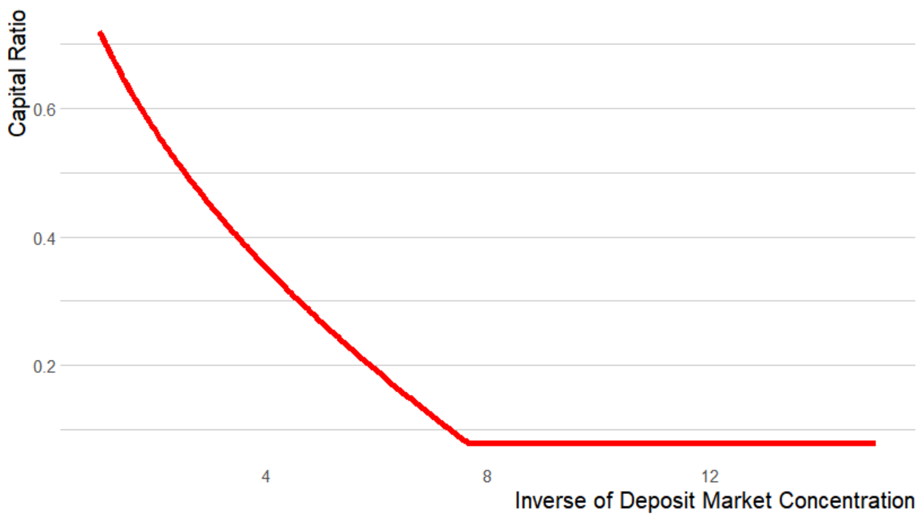
$$\delta + r_t^d = \frac{\epsilon_t^d}{\epsilon_t^d + 1/M_d}\mu_{it} \tag{20}$$

$$\delta + r_t^b = \frac{\epsilon_t^b}{\epsilon_t^b - 1/M_l}(\mu_{it} + \kappa\chi_{it}) \tag{21}$$

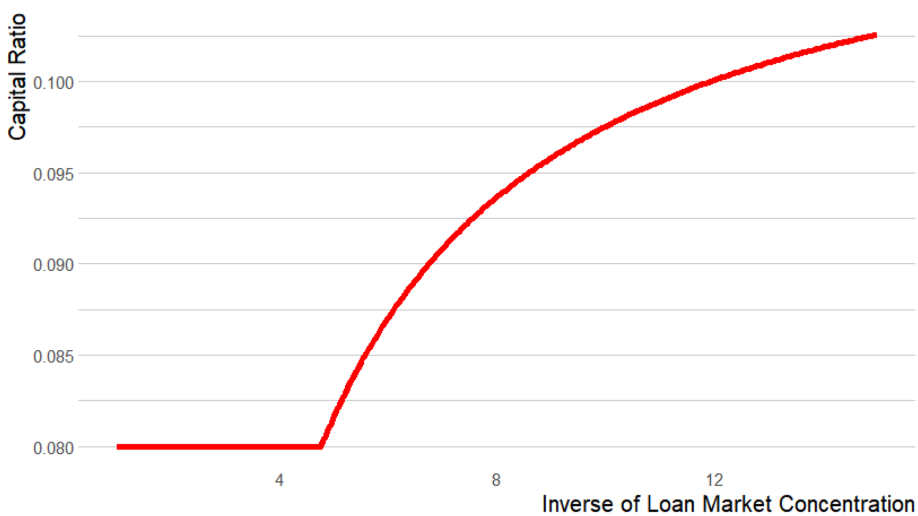
where on the left hand side of (20) and (21) shows the effective deposit rate and loan rate. The market share held by individual bankers in the deposit and loan markets are determined by M_d and M_l.

Case 1 To investigate the impact of deposit market concentration on bank capital, I consider a scenario where loan market concentration remains fixed. Specifically, I set M_l to its calibrated parameter value of 7.45, and analyze the effects of varying M_d on bank capital. As illustrated in panel A of Figure 6, the results demonstrate that bank capital ratio increases as deposit market concentration rises. The observed relationship between bank capital and deposit market concentration in this scenario is precisely the same as that observed in the baseline model. Intuitively, the results suggest that when the deposit market is highly concentrated, bankers are inclined to set lower deposit rates. As deposit returns decline, fewer entrepreneurs are willing to accept deposit contracts, prompting bankers to accumulate more capital. Notably, when deposit market concentration (as measured by HHI) exceeds 0.25, the bank capital ratio may exceed 30%.

Case 2 To investigate the impact of loan market concentration on bank capital, I consider a scenario where deposit market concentration remains fixed. Specifically, I set M_d to a fixed value of 7.45 and examine how changes in M_l affect bank capital. As illustrated in panel B of Figure 6, the red line indicates a negative relationship between loan market concentration



Panel A: Effect on Deposit Market Concentration on Bank Capital



Panel B: Effect on Loan Market Concentration on Bank Capital

Figure 6: Effects of Deposit (Loan) Market Concentration on Bank Capital

and bank capital ratio. This observed correlation is entirely opposite to that observed in Case 1. The results suggest that an increase in loan market concentration leads to a higher loan rate, resulting in smaller loan sizes for entrepreneurs. Due to the scarcity of investment opportunities, bankers accumulate less capital.

The two cases highlight the dominance of deposit market concentration as a driver of increases in bank capital. This conclusion is further supported by a comparison of the magnitudes of the capital ratio changes observed in the two cases, as depicted in the two panels of Figure 6. Specifically, the increase in deposit market concentration leads to a much larger rise in bank capital than the decrease in bank capital resulting from loan market concentration. Notably, in the baseline model, where $M = M_d = M_l$, we observe a positive correlation between bank concentration and bank capital ratio when the bank capital constraint is nonbinding.

6 Conclusion

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendices

A Proofs

Proof for Lemma 1

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

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

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

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

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

Proof for Lemma 2

Proof. Equations (8) and (9) are directly obtained from Lemma 1, given that borrowing entrepreneurs borrow up to the borrowing limits and lending entrepreneurs deposit all their capital in the bank.

For lending entrepreneurs, the budget constraint becomes:

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

For borrowing entrepreneurs, the budget constraint becomes:

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

For autarky entrepreneurs, the budget constraint becomes:

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

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

Thus I obtain

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

which is an equivalent formula of Equation (10). ■

Proof for Lemma 3

Proof. The bellman equation for the banker i is:

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

subject to the balance sheet identity (4), the budget constraint (5), and the minimum capital requirement (6).

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

$$r_t^b + 1 = \frac{(\mu_{it} + \kappa\chi_{it})M\overline{z_t}}{(M+1)\overline{z_t} - z_{max}} \tag{22}$$

$$r_t^d + 1 = \frac{\mu_{it}M\underline{z_t}}{(M+1)\underline{z_t} - z_{min}} \tag{23}$$

$$q_t = \beta q_{t+1}(\mu_{it+1} + \chi_{it+1}) \tag{24}$$

$$\chi_{it+1}(N_{it+1} - \kappa Q_{it+1}^L) = 0 \tag{25}$$

where $q_t\mu_{it}$ is the multiplier of the bank's balance sheet identity. $q_t\chi_{it+1}$ is the multiplier of the capital constraint. Equation (25) is the complementary and slackness condition for the minimum capital requirement.

With the assumption $\kappa = 0$, the combination of Equations (22) and (23) with Lemma 1 leads to the conclusion that $\frac{\overline{z_t} - \underline{z_t}}{z_{max} - z_{min}} = \frac{1}{M+1}$, which implies that the fraction of autarky entrepreneurs is $\frac{1}{M+1}$. ■

Proof for Proposition 1

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

Case 1. Let me first consider the scenario in which the bank capital constraint is binding. In this case, bankers fund loans solely with deposits. Denote the proportion of autarky entrepreneurs and borrowing entrepreneurs as u and v, respectively[6]. From Lemma

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

3, we have $u = \frac{1}{M+1}$, while Equation (4) implies that $v = \frac{M}{\lambda(M+1)}$.

Substituting the formula of u and v into the law of motion for aggregate capital (Equation (10)), I obtain

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

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

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

Therefore, I obtain

$$r^d + 1 = \frac{2\frac{1}{\beta}\lambda(M+1)^2\underline{z}}{[(2M^2+2M+1)\lambda - M^2]z_{max} + [(2M+1)\lambda + M^2]z_{min}} \tag{26}$$

$$r^b + \delta = (r^d+\delta)\frac{\overline{z}}{\underline{z}} \tag{27}$$

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

Case 2. Now suppose that the bank capital constraint is non-binding. By Equation (25), $\chi = 0$. Plugging this into Equation (24), I obtain $\mu = \frac{1}{\beta}$. Then the deposit rate and loan rate become:

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

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

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

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

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

$$v = \frac{-(2M+1) + \sqrt{4M^2+4M+1+(2M^3+M^2)\lambda}}{(M+1)M\lambda} \tag{28}$$

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

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

Equation (22) and (26), the following expression is obtained:

$$\lambda > \frac{M^2 + 4M + 2}{2M + 1} \equiv \lambda(M) \tag{29}$$

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

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

$$\begin{aligned}\frac{N}{N+D} = \frac{\lambda v - \frac{M}{M+1}}{(\lambda - 1)v} &= \frac{\lambda}{\lambda - 1}(1 - \frac{M^2}{-(2M+1) + \sqrt{4M^2 + 4M + 1 + (2M^3 + M^2)\lambda}}) \\ &= \frac{\lambda}{\lambda - 1}(1 - \frac{\sqrt{1 + \lambda\frac{M^2}{2M+1}} + 1}{\lambda})\end{aligned} \tag{30}$$

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

Proof for Proposition 2

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

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

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

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

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

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

The third term in the bracket becomes:

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

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

Output takes the form

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

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

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

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

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

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

Proof for Proposition 3

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

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

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

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

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

Proof for Proposition 4

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

$$
\begin{aligned}
\frac{N^*}{K^*} &= \lambda v - \frac{M}{M+1} \\
&= \frac{-(2M+1) + \sqrt{(2M+1)^2 + (2M^3 + M^2)\lambda}}{M(M+1)} - \frac{M}{M+1} \\
&= \frac{\sqrt{(2M+1)^2 + (2M^3+M^2)\lambda}}{M(M+1)} - \frac{M+1}{M}
\end{aligned} \tag{31}
$$

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

$$
\begin{aligned}
&\frac{\sqrt{(2M+1)^2 + (2M^3+M^2)\lambda}}{M(M+1)} - \frac{M+1}{M} > \frac{\sqrt{\lambda-1}}{M+1} - 1 \\
\Rightarrow &\sqrt{(2M+1)^2 + (2M^3+M^2)\lambda} > M\sqrt{\lambda - 1} + (M+1)
\end{aligned} \tag{32}
$$

Equation (32) becomes $2M + 1 + \lambda M^2 > (M+1)\sqrt{\lambda - 1}$, which always holds because $2M^2(2M+1) > (M+1)^2$. ∎