Essays on Firms, Finance, and Macroeconomy

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Dan Su

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Xiaoji Lin, Advisor

June, 2022

Acknowledgements

I thank my advisor Xiaoji Lin and other committee members, Robert Goldstein, Erzo Luttmer, and Andrew Winton, for their continuous help.

Dedication

To my parents and Xiang

Abstract

The primary goal of this dissertation is to understand how the business activities of companies impact the macroeconomy. More specifically, it contains three essays. In the first essay "Rise of Superstar Firms and Fall of the Price Mechanism", I investigate the misallocation implications of corporate internal financing. I introduce product market competition and corporate risk management into a standard continuous-time heterogeneous agent model with incomplete markets. I show that the economy's ability to allocate resources across different agents through the price mechanism is bounded by corporate internal savings as there is no market to equalize the marginal value of internal resources across firms. In other words, corporate cash can help achieve dynamic efficiency across times at the firm level but not static efficiency across individuals at the macro level. More importantly, misallocation – defined as the static resource allocation efficiency across individuals – increases in the new economy where (superstar) firms rely more on internal financing due to the increased earnings risk. Finally, this model can quantitatively match the deteriorating capital allocation efficiency in the U.S. data.

In the second essay "The Rise of (Mega-)Firms with Negative Net Earnings", I document the prevalence of public companies with negative net earnings since the 1970s. The fraction of firms with negative net income has increased sharply from 18% in 1970 to 54% in 2019. Such an increase is mainly driven by the right shifts in the mean, i.e., the increasing popularity of sizable firms that are not profitable. Based on the existing literature on customer capital, I conjecture that the increasing returns-to-scale in the new economy is the main driver behind it. I provide three pieces of supporting evidence. First, earning losses mostly come from the growing customer capital expenses instead of production-related costs, capital investments, or R&D expenditures. Second, cross-sectionally, firms with higher markup tend to have lower net incomes. Third, industries with low marginal production costs, on average, have higher percentages of unprofitable companies.

The last essay "The Macroeconomics of TechFin" is to investigate the business cycle implications of TechFin. Over the past few years, many large technology companies have started lending in the capital markets, i.e., "TechFin". How should we modify

our existing macro-finance theories to accommodate the rise of this new financial intermediary? In this paper, I introduce both a banking sector and a TechFin sector into a continuous-time general equilibrium model with heterogeneous entrepreneurs and incomplete markets. These two financial sectors are identical except for the types of borrowing constraints faced by entrepreneurs. Entrepreneurs borrowing from banks are subject to the standard collateral-based borrowing constraints. In contrast, technology advantages allow the big tech companies to resolve agency costs and perform cash flow-based lending. I use a deep learning neural network approach to obtain global solutions, and the main conclusions are twofold. First, this new TechFin credit system leads to a higher capital allocative efficiency in the steady state. Second, the existence of BigTech lending acts as a propagation mechanism and makes the economy sensitive to the second-moment uncertainty shocks: a small and transitory micro-uncertainty shock can lead to amplified and persistent changes in aggregate outputs. This new financial accelerator mechanism, associated with the new TechFin sector, differs from the classic one (e.g. Bernanke and Gertler, 1989; Kiyotaki and Moore, 1997) in three aspects: micro uncertainty instead of aggregate productivity is the primitive shock; financial friction comes from earnings-based borrowing constraints instead of collateral-based ones; and the feedback loops happen between net worth inequality, instead of net worth level, and asset prices.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Companies nowadays behave very differently from what they did several decades ago. The main topic of my dissertation is to understand why firms are changing their behaviors, as well as the corresponding macroeconomic consequences.

More specifically, this dissertation has three essays. In the first essay "Rise of Superstar Firms and Fall of the Price Mechanism", I investigate the misallocation implications of corporate internal financing. I introduce product market competition and corporate risk management into a standard continuous-time heterogeneous agent model with incomplete markets. I show that the economy's ability to allocate resources across different agents through the price mechanism is bounded by corporate internal savings as there is no market to equalize the marginal value of internal resources across firms. In other words, corporate cash can help achieve dynamic efficiency across times at the firm level but not static efficiency across individuals at the macro level. More importantly, misallocation – defined as the static resource allocation efficiency across individuals – increases in the new economy where (superstar) firms rely more on internal financing due to the increased earnings risk. Finally, this model can quantitatively match the deteriorating capital allocation efficiency in the U.S. data.

In the second essay "The Rise of (Mega-)Firms with Negative Net Earnings", I document the prevalence of public companies with negative net earnings since the 1970s. The fraction of firms with negative net income has increased sharply from 18% in 1970 to 54% in 2019. Such an increase is mainly driven by the right shifts in the mean, i.e., the increasing popularity of sizable firms that are not profitable. Based on the existing literature on customer capital, I conjecture that the increasing returns-to-scale in the new economy is the main driver behind it. I provide three pieces of supporting evidence. First, earning losses mostly come from the growing customer capital expenses instead of production-related costs, capital investments, or R&D expenditures. Second, cross-sectionally, firms with higher markup tend to have lower net incomes. Third, industries with low marginal production costs, on average, have higher percentages of unprofitable companies.

In the third essay "The Macroeconomics of TechFin", I investigate the aggregate implications of BigTech lendings. Over the past few years, many large technology companies have started lending in the capital markets, i.e., "TechFin". How should we modify our existing macro-finance theories to accommodate the rise of this new financial intermediary? In this paper, I introduce both a banking sector and a TechFin sector into a continuous-time general equilibrium model with heterogeneous entrepreneurs and incomplete markets. These two financial sectors are identical except for the types of borrowing constraints faced by entrepreneurs. Entrepreneurs borrowing from banks are subject to the standard collateral-based borrowing constraints. In contrast, technology advantages allow the big tech companies to resolve agency costs and perform cash flow-based lending. I use a deep learning neural network approach to obtain global solutions, and the main conclusions are twofold. First, this new TechFin credit system leads to a higher capital allocative efficiency in the steady state. Second, the existence of BigTech lending acts as a propagation mechanism and makes the economy sensitive to the second-moment uncertainty shocks: a small and transitory micro-uncertainty shock can lead to amplified and persistent changes in aggregate outputs. This new financial accelerator mechanism, associated with the new TechFin sector, differs from the classic one (e.g. Bernanke and Gertler, 1989; Kiyotaki and Moore, 1997) in three aspects: micro uncertainty instead of aggregate productivity is the primitive shock; financial friction comes from earnings-based borrowing constraints instead of collateral-based ones; and the feedback loops happen between net worth inequality, instead of net worth level, and asset prices.

The remainder of the dissertation is organized as follows. Chapter 2 contains the first essay "Rise of Superstar Firms and Fall of the Price Mechanism". Chapter 3 includes the second essay "The Rise of (Mega-)Firms with Negative Net Earnings". Chapter

4 provides the main content for the third essay "The Macroeconomics of TechFin". Proofs, additional figures and tables, and other supplementary materials can be found in the appendix.

Chapter 2

Rise of Superstar Firms and Fall of the Price Mechanism

> "The price mechanism might be superseded if the relationship which replaced it was desired for its own sake."
>
> — Coase (1937), *The Nature of the Firm*

Over the past several decades, we have observed several puzzling macro-finance trends in the data. First, corporate market power, measured as markup, has increased steadily over time (De Loecker, Eeckhout and Unger, 2020). Second, companies rely more and more on internal financing by holding excessive cash on their balance sheets (Bates, Kahle and Stulz, 2009*a*). Third, capital misallocation, measured as the static dispersion of firm-level marginal product of capital, has risen gradually (Hsieh and Klenow, 2018; Bils, Klenow and Ruane, 2021).[1] In this paper, we argue that these three phenomena are deeply connected. Our goal is to provide a theoretical and quantitative framework to explain them jointly.

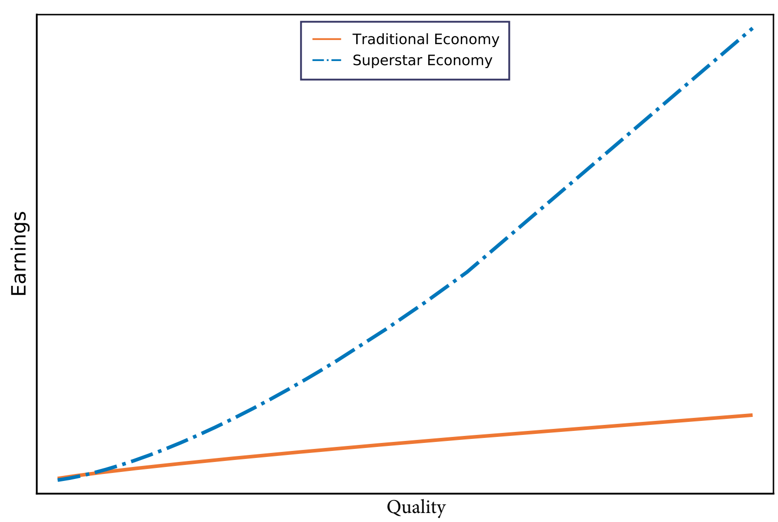
The key model mechanism in this paper can be explained as follows. The primitive drivers behind these trends are some economic fundamental changes from both the demand and supply sides. We will be more specific on the corresponding model setups

[1] Hsieh and Klenow (2018) propose an idea of *reallocation myth*. They find that there has been no improvement in capital allocative efficiency in the U.S. Even after correcting the possible mismeasurement errors, there is a 15% decline in the U.S. allocation efficiency over the past 35 years (Bils, Klenow and Ruane, 2021).

in Section 2.1. The intuition is that nowadays, on the demand side, consumers care more about product quality than quantity. In other words, they are willing to pay more for high-quality products. On the supply side, some new technology (e.g., digitalization) makes firms less costly to serve all the potential customers, which leads to an increased returns-to-scale.

These two primitive shocks can directly impact the level and volatility of corporate earnings. As shown in Figure 2.1, with these two fundamental changes, firms with the best product quality can obtain enormous earnings and dominate the activities in which they engage. This static income *level* redistribution effect has been widely recognized in the existing literature (e.g., Sattinger, 1993; Tervio, 2008; Scheuer and Werning, 2017). However, when firms face uncertainty in their product quality, from a dynamic perspective, superstars are inherently riskier. Compared to the low-concentration traditional economy, in a Superstar Economy, a small variation in product quality can translate into considerable earnings fluctuations, especially on the right-tail side. In other words, these economic fundamental changes are also redistributive in *risk*. As shown in Section 2.1, this feature shows up in the model because corporate earnings become a convex function of the underlying quality in the Superstar Economy. With convexity, both income level and risk are redistributed towards right-tail firms.

Figure 2.1: Risky superstar economy



At the same time, these two economic fundamental changes can also generate both micro- and macro-indirect effects. The micro-indirect impact comes from the fact that the risk-redistribution nature of fundamental changes will substantially and heterogeneously affect corporate *risk* management policy. With external financing costs and precautionary savings incentives, an increased future earnings uncertainty makes firms optimally choose to accumulate more cash and rely more on internal financing. Cross-sectionally, such an incentive is stronger for right-tail firms with more volatile earnings processes. At the same time, this shift in corporate risk management can also generate macro-indirect effects by lowering the aggregate capital allocation efficiency. The intuition is that corporate internal financing behavior is dynamically efficient across times at the individual level but not statically efficient across firms at the macro level. Therefore, misallocation, defined as the static resource allocation efficiency across different individuals, increases when firms rely more on internal financing. In other words, the economy's ability to allocate resources across different firms is being limited by corporate internal financing behaviors. This situation worsens in the Superstar Economy as firms, especially those on the right tail, become more reliant on internal cash holdings due to the increased earnings risk.

To sum up, the punchline in this paper is that misallocation, narrowly defined as the economy's ability to allocate resources across different agents, increases in the new economy with superstar firms. The underlying mechanism comes from the fact that shifts in demand and supply curves can generate direct impacts on the level and volatility of corporate earnings, through which they lead to both micro-indirect impacts on risk management and macro-indirect effects on misallocation. Meanwhile, our paper can be interpreted as one investigating the changes in competition between firms and the market system in the 21st century. Coase (1937) argues that the nature of the firm is a substitution of the market system.[2] Suppose we follow his interpretation and observe the increasing importance of firms in the new economy. In that case, it must be that the relative importance of the market system has been declining. During this process, due to the different natures in resource allocation abilities between firms and market, we should be able to see declining capital allocation efficiency in the data (Hsieh and Klenow, 2018; Bils, Klenow and Ruane, 2021). One crucial difference here is that we focus on the financing side instead of the production side, and we argue that corporate

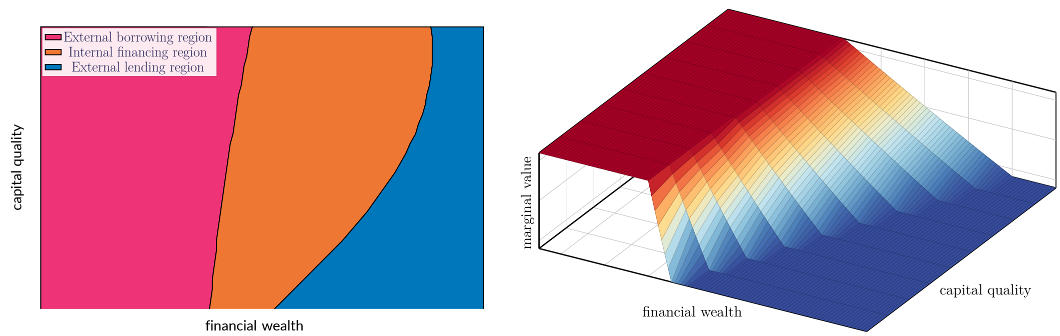
[2]In a free-market economy without firms, the price mechanism alone can work sufficiently well to allocate resources to their highest-valued uses (Smith, 1759, 1776). This is the well-known concept of the invisible hand. However, Coase (1937) and later Williamson (1975) argue that market coordination is unlikely to be the only allocation system within capitalism. The transaction costs of using a market system lead to the emergence of firms as an entirely different sub-economy. Different from the market system, the goal of companies is *not* to eliminate inefficiency in resources allocations.

internal financing is an imperfect substation of external market financing tools.

Roadmap This paper mainly consists of three parts. To begin with, we provide a theoretical model to establish our story formally. Generally speaking, we introduce product market competition and corporate risk management into a standard heterogeneous agent model with incomplete markets. Firms are heterogeneous in capital quality, cash holdings, and external debt positions. Entrepreneurs can improve capital quality through internal investment. However, capital quality must be subject to uninsurable idiosyncratic shocks. For production market competition, on the demand side, we introduce a new parameter named taste for quality, which captures how much people care about quality. Eventually, this parameter determines the prices of goods with different quality. On the supply side, we assume that firms need to pay additional costs when selling products to consumers. The most crucial supply-side parameter is the curvature of the supply curve, which measures how costly it is for firms to expand their operating scale. As for the financing side, we follow the standard risk management framework proposed by Bolton, Chen and Wang (2011) with some slight adjustments. The model has three key implications. First, due to this product market competition, corporate earnings and markup become a function of the underlying capital quality. In addition, the sensitivity of earnings concerning quality depends crucially on these economic fundamental factors. Suppose consumers' taste for quality is strong enough and firms' supply curve curvature becomes flat enough in the new economy. In that case, earnings become a convex function of underlying capital quality. With convexity, both the level and volatility of earnings are substantially higher for firms with better product quality. In other words, shifts in supply and demand curves make both income level and risk redistributed towards right-tail firms.

Second, this risk redistribution makes firms in the new economy optimally choose to hold more cash and rely more on internal financing. In addition, this incentive is more substantial for right-tail superstar firms, as their expected future earnings are riskier. This result shows up in the model because we assume external financing costs, which could be micro-founded as asymmetric information between firms and external investors. With external financing costs, firms find it optimal to target their cash holdings between some upper and lower boundaries. These boundaries are wider for firms with higher future earnings uncertainty. What is interesting here is the aggregate effects: as shown in the left graph in Figure 2.2, these cash control boundaries, arising from individual firms' optimal decisions, segment the whole economy into several different regions and create an endogenous firm-market boundary. Within this boundary, firms rely on internal financing. Outside this boundary, firms use the external financial market for lending and borrowing.

Figure 2.2: Boundary of the invisible hand



Third, an expansion of the internal financing region increases capital misallocation. The underlying mechanism comes from the fact that there is a fundamental difference between internal financing and external financing. As shown in the right graph in Figure 2.2, when firms borrow or lend externally, the uniform price of debt equalizes the marginal cost of capital across firms. However, when firms save internally, the marginal cost of capital becomes unequal because marginal cash value can vary between some upper and lower boundaries. Thus, different from the conventional wisdom that self-financing can undo misallocation (e.g., Midrigan and Xu, 2014; Moll, 2014), in this paper, an expansion of the internal financing region lowers the aggregate capital allocation efficiency. Misallocation increases in the new economy as firms have stronger incentives to save internally.

To follow, we implement several reduced-form empirical investigations on whether some testable predictions derived from our theory actually hold in the data. Our main findings in this part are threefold. To start, superstar firms are indeed riskier as they face more volatile fluctuations in markup. Furthermore, superstar firms on average have the largest degree of capital misallocation. Finally, firms with higher markup are likely to hold more cash on their balance sheets. These three pieces of empirical evidence provide additional support for our previous model mechanism.

Finally, we investigate the model's quantitative implications. We estimate the structural parameters of the model through the simulated method of moments (SMM) approach. In order to evaluate the quantitative performance of the model, we select three facts related to the deteriorating capital allocation efficiency in the U.S. First, the dispersion of the marginal product of capital (MPK) increases sharply among U.S. public firms. Second, the correlation between firm-level total factor productivity (TFP) and external financing dependence has changed from positive to negative, indicating that productive firms become less reliant on external finance. Third, the positive gap between the marginal product of capital and the real interest rate increases over time. These facts are interpreted as signs of increasing market inefficiency as an efficient financial market should lead to zero misallocation, more resources allocated to more productive users, and the marginal return of investment being equal to its marginal cost. When taken to data, our model is able to quantitatively match both the aggregate trends and cross-sectional patterns in the data.

Contributions The contributions of this paper are mainly fourfold. First, we extend academic discussions on superstar firms to the misallocation literature while most existing studies are focused on their impacts on labor share (e.g., Autor et al., 2020) or business dynamism (e.g., De Ridder, 2019). In addition, the existing misallocation focuses more on the markup level channel. In contrast, this paper highlights the relationship among markup risk, internal financing incentives, and capital misallocation. Second, this paper provides a different finance view on the origins of misallocation. We argue that the dispersion of marginal product of capital could come from either borrowing constraints (e.g., Midrigan and Xu, 2014; Buera and Shin, 2013) or an endogenous changes in firms' reliance on the external financial market. Third, we formally establish Coase (1937)'s firm-(financial)market boundary in general equilibrium. More specifically, we introduce corporate risk management framework (e.g., Bolton, Chen and Wang, 2011) into the existing distributional macro framework (e.g., Moll, 2014), and investigate the misallocation implications of corporate internal financing, which has not yet been studied in the existing literature. Fourth, we extend the $R - g$ framework on inequality to allow for the existence of two types of capital returns (two Rs). One is the capital return of entrepreneurs who are still relying on the external financial market to finance their investment, and the other is that of entrepreneurs who are not. The divergence between these two capital returns comes from the risk redistribution nature of economic fundamental changes. More discussions on the existing literature are provided in Section 2.4.

Layouts The rest of this paper is organized as follows. Section 2.1 describes the model setup and provides some preliminary discussions on the underlying mechanism. Section 2.2 presents the reduced-form evidence for the key model predictions. Section 2.3 shows that our model is able to quantitatively match three trends related to the declining capital allocation efficiency in the data. In Section 2.4, we review the existing literature and also discuss the validity of two important assumptions used in this paper. Section 3.4 offers a conclusion.

2.1 Theory

2.1.1 Model Setup

Preference

Consider an infinite-horizon continuous-time economy populated by a unit-mass continuum of entrepreneurs. All entrepreneurs in this economy have the same additive utility function shown as follows:

$$\mathcal{J}_t = \mathbb{E}_t \left[\int_t^\infty e^{-\rho(s-t)} u\left(c_s\right) ds \right], \forall t \geq 0 \tag{2.1}$$

where ρ is the rate of time preference, u is the utility function, and $\mathcal{J}$ denotes the value function. Following Duffie and Epstein (1992) and references thereafter, we introduce the following normalized aggregator f for consumption c and continuation value $\mathcal{J}$ in each period:

$$f\left(c, \mathcal{J}\right) = \frac{\rho}{1-\theta} \frac{c^{1-\theta} - \left[\left(1-\gamma\right)\mathcal{J}\right]^{\frac{1-\theta}{1-\gamma}}}{\left[\left(1-\gamma\right)\mathcal{J}\right]^{\frac{1-\theta}{1-\gamma}-1}} \tag{2.2}$$

where γ determines the coefficient of relative risk aversion and $\frac{1}{\theta}$ measures the elasticity of intertemporal substitution (EIS) for deterministic consumption paths. With this recursive preference, we can separate the effects of risk aversion from EIS.

Entrepreneurs are indexed by their efficient capital ζ, cash holdings ω, and external bond positions b. At each time t, the state of the economy can be summarized as a joint probability density distribution $\Lambda_t(\zeta, \omega, b)$. From now on, we drop the individual and time subscript for simplicity unless otherwise needed.

Product market competition

There is one homogeneous final goods produced by a number of perfectly competitive final-goods producers. Final goods are used as the numeraire here, so its price is normalized to be 1 in each period. These final-goods producers have access to a production technology that transforms bundles of intermediate goods produced by entrepreneurs, i.e., $\{y_i\}_{i\in[0,1]}$ with $y_i \in \mathbb{R}_+$, into the final good $\mathcal{Y}$, with the following production function:

$$\mathcal{Y} = \int_0^1 \mathcal{M}(\zeta_i)^{\phi} y_i(\zeta_i)\, di = \int_0^1 \zeta_i^{\phi} y_i(\zeta_i)\, di \tag{2.3}$$

where $\mathcal{M}$ is a strictly increasing function of capital quality ζ and $\phi \geq 0$ represents the consumer's preference on product quality. In the baseline model, we use the simplest functional form $\mathcal{M}(\zeta) = \zeta$, but our story can be extended to allow for a more general $\mathcal{M}$ function. The key difference between this paper and the existing Dixit-Stiglitz framework is that we assume consumers have a taste for *quality* instead of *variety*. The consumer has a taste for variety if he or she prefers to consume a diversified bundle of goods. In this paper, we do not impose such a condition. As a matter of fact, if $\phi = 0$, then from the consumer's perspective, all the goods with different qualities are perfectly substitutable to each other. In this extreme case, consumers only care about quantities. However, if $\phi \neq 0$, then consumers face a quantity-quality trade-off: there is an imperfect substitution among final goods with different qualities and this degree of substitution is governed by ϕ.[3] With a relatively higher value ϕ, they care more about quality and need to consume a lot more low-quality products in order to get the same utility from having one high-quality product. This quantity-quality trade-off framework is borrowed from Rosen (1981).[4]

[3]For instance, suppose there are two intermediate goods with different qualities of x and y, and $x > y$. Then it is equivalent for the consumer to consume 1 intermediate goods with quality x and $\left(\frac{\mathcal{M}(x)}{\mathcal{M}(y)}\right)^{\phi}$ number of intermediate goods with quality y. In this way, ϕ denotes how much these consumers care about the quality. With Equation (2.3), products with quality-adjustment are still perfect substitutes for final-goods producers.

[4]In the existing literature, this $\mathcal{M}$ function has many different interpretations and applications. For instance, Atkeson and Burstein (2019) interpret this $\mathcal{M}$ as the measure of intermediate goods with frontier technology; Dou, Ji and Wu (2021) view this element as customer base; and Cavenaile and

Another key difference between this paper and the original Dixit-Stiglitz framework is that the price here is a competitive price that clears all the intermediate goods in *different* markets.[5] In other words, prices of intermediate goods are not endogenously choices made by entrepreneurs, but instead prices come from competitive market clearance condition and they reflect customers' taste for products with different qualities: $p(\zeta) = \mathcal{M}(\zeta)^{\phi}$. At the same time, the final-goods producers always make zero profits in equilibrium so that we do not need to consider their consumption decisions. Two conclusions are worth noting here. First, price is increasing in capital quality. Second, changes in customer's taste ϕ can directly affect the (relative) prices of products with different qualities.

Despite that product demand is mostly determined by its capital quality, entrepreneurs still need to decide how many products they want to sell in the market because we assume that serving the customers is costly. In other words, in this model setup, entrepreneurs are doing a Cournot-type competition. More specifically, the entrepreneurs need to pay additional costs of $\Theta(y)$ when they sell y products to the consumers. $\Theta(y)$ here can be interpreted as administrative expenses, sales costs, or production costs that are not directly modelled in this paper. Following the existing literature (e.g., De Ridder, 2019), we assume that $\Theta(y) = f_0 + \xi_0 y^{\frac{1}{\eta}}$, which contains a fixed cost of f_0 and a variable cost component with marginal cost $\xi_0 > 0$ and a scale parameter $0 < \eta \leq 1$. Here $\frac{1}{\eta}$ denotes the curvature of the supply curve, i.e., how costly it is for firms to expand their operating scale. In addition, it could also represent how much the current technology admits the joint consumption or the degree of degradation of services as entrepreneurs increasing their scale. One typical example for changes in $\frac{1}{\eta}$ is digitization: the availability of internet largely reduces the costs for companies to serve a massive number of potential buyers.

In each period, the entrepreneur's optimization problem is to maximize his or her static profits π and it can be summarized as below:

$$\pi \equiv \max_{y} p(\zeta)\, y - \Theta(y) = p(\zeta)\, y - f_0 - \xi_0 y^{\frac{1}{\eta}} \tag{2.4}$$

Roldan-Blanco (2021) use this term as a proxy for the total *perceived* quality of different goods.

[5]It means that for intermediate-goods producers (i.e., entrepreneurs), products with different qualities are entirely *different* goods traded at *different* markets. However, for final-goods producers, after quality adjustments, products with different qualities are perfectly substitutes. We use this setup to generate non-constant markup for firms with different qualities.

Entrepreneurs own and accumulate capital. In addition, they can improve their capital quality through internal investment. Following Brunnermeier and Sannikov (2014) and Di Tella (2017), instead of directly modeling the productivity process, here we assume that capital quality is subject to some random shocks with the following process:

$$d\zeta_t = \left(\bar{\mu} + \iota_t^{\zeta} - \delta\zeta_t\right) dt + \sigma\sqrt{\zeta_t} d\mathcal{Z}_t \tag{2.5}$$

Following the existing literature, capital quality can be interpreted as the efficiency units of capital, and capital quality shocks can be considered as persistent shocks to total factor productivity. Compared to the conventional assumption of productivity shocks, one crucial benefit of using capital quality shock is to save one state variable and thus reduce the computational complexity. Equation (2.5) is a modified Cox-Ingersoll-Ross (CIR) model (Cox, Ingersoll and Ross, 1985), which was used for describing interest rate movements in the asset pricing literature. In this equation, $\bar{\mu}$ is the long-run mean level of entrepreneur's capital quality, ι_t^{ζ} is the entrepreneur's total (dis)investment on capital quality at time t, and δ is the capital quality depreciation rate. $\mathcal{Z}_t$ is the standard exogenous Brownian shock, and it is independent and identically distributed (i.i.d.) across different firms. σ represents the sensitivity of ζ to $\mathcal{Z}_t$. As our purpose is to investigate how shifts in demand and supply curves amplify the impacts of incomplete market frictions on the value of cash, here we assume that entrepreneurs cannot fully hedge their idiosyncratic business risk $\mathcal{Z}_t$.[6]

Following Hayashi (1982) and references thereafter, we assume that when converting final goods into the new capital, entrepreneurs need to pay additional adjustment costs, and they can be specified by a standard quadratic form of $\frac{\kappa_0}{2}\left(\frac{\iota^{\zeta}}{\zeta}\right)^2$, where κ_0 measures the degree of investment inflexibility and $\kappa_0 > 0$.

Corporate risk management

As for the financing side, we assume that entrepreneurs can fund their investment projects with current earnings, internal financing with cash, and external financing

[6] An implicit assumption here is that there is no perfect capital market for entrepreneurs to rent or lease. We could achieve the same result by assuming that entrepreneurs need to pay substantially higher transaction costs when they directly rent or lease their capital, compared to borrowing or lending in the financial market. In this way, entrepreneurs will endogenously choose not to use the capital renting/leasing market.

with instantaneous risk-free bonds.[7] Generally speaking, the model setup here follows the classic cash inventory model in corporate finance literature (e.g., Miller and Orr, 1966; Froot, Scharfstein and Stein, 1993). To introduce the wedge between external and internal financing in a continuous-time setup, as well as investigate the joint behaviors of cash accumulation and investment, we follow the basic setup proposed by Bolton, Chen and Wang (2011) but with some modifications. For instance, transaction costs in Bolton, Chen and Wang (2011) are from external equity financing. Meanwhile, in this paper, we introduce the issuance costs of debt and obtain the simultaneous existence of debt and cash.

External financing More specifically, entrepreneurs have two different ways to save their financial wealth. To begin with, entrepreneurs can lend and borrow in the external financial asset market:

$$db_t = \iota_t^b dt \tag{2.6}$$

b_t denotes the amount of short-term debt issued by each entrepreneur at time t and ι_t^b is the net changes in debt issuance, i.e., $\iota_t^b = b_{t+dt} - b_t$. As mentioned before, for simplicity, we assume that $b_{i,t}$ is issued in short-term and risk-free. In order to make the debt risk-free despite the underlying business risk, we introduce the following two modifications. First, as we can see from the timeline below, we assume that entrepreneurs can observe their next-period capital quality before issuing the short-term debt. This adjustment follows the idea of Kiyotaki (1998) and many works after that. Second, we introduce the following borrowing constraint to make sure that entrepreneurs can only borrow a fraction of their earnings:

$$b_t \leq \beta \frac{\pi_t - \chi_0}{1 + r + \chi_1}, \eta \in [0, 1] \tag{2.7}$$

where π are the entrepreneur's profits as we have defined before. With these two adjustments, all the borrowing and lending in this economy become **credit risk-free**. In some aspects, Equation (2.7) can be interpreted as the earnings-based borrowing constraint (e.g., Lian and Ma, 2021) with external financing costs. $\beta \in [0, 1]$ denotes the tightness of this modified earnings-based borrowing constraint. When $\beta = 1$, we say

[7]Long-term debt financing, long-term employment of managers, and corporate defaults do impact the final results. However, discussions on their quantitative impacts are beyond this paper's scope and hence left for future research.

that the capital market is perfect within its boundary, and entrepreneurs can borrow as much as they are able to pay back next period. When $\beta = 0$, the external financing market is completely shut down, and entrepreneurs can borrow nothing. In this paper, we assume that all the entrepreneurs will face the same β.[8]

Importantly, external financing is costly: every time entrepreneurs borrow or lend externally, they need to pay additional financing costs of $\chi_0 + \chi_1 |b_t|$, where χ_0 and χ_1 are the fixed and variable costs, respectively. These transaction costs could be interpreted in many different ways. First, external financing costs could come from the actual debt issuance costs, such as fees and commissions paid to investment banks, law firms, auditors, and anyone else involved. These costs could be substantial, especially for large syndicated loans. Second, χ_0 and χ_1 could be entrepreneurs' opportunity costs, including the time spent waiting and going through all the external financing process. In some circumstances, these costs could be quite considerable, especially when the economy is in a bad situation. Third, whenever entrepreneurs seek external financing, issues such as asymmetric information and agency problems might arise. In other words, these external financing costs reflect the degree of information asymmetry between companies and outside investors. Finally, those costs should *not* be interpreted as financial frictions. In this paper, by financial frictions, we specifically mean distortions within the financial market system. In contrast, these external financing costs should be broadly considered as the transaction costs of using the market system (Coase, 1937).

Internal financing In addition to the external financial market, entrepreneurs can save in corporate cash to insure themselves. We use $\omega_{i,t}$ to denote entrepreneur i's cash inventory at time t and the cash accumulation evolves as follows:

$$d\omega_t = (\iota_t^\omega - \lambda\omega_t)\, dt, \text{ and } \omega_t \geq 0 \quad \forall t \tag{2.8}$$

where ι_t^ω represents the adjustments on corporate cash balance, thus it can be either positive or negative. As companies cannot have negative cash stock, we impose the non-negativity condition on the level of ω_t, i.e., $\omega_t \geq 0$.

[8] One model extension is to use alternative borrowing constraints, including the natural borrowing constraint and collateral-based borrowing constraint (e.g., Kiyotaki and Moore, 1997). Different types of borrowing constraints will generate quantitatively different results, but the key story, i.e., the shrinking boundary of the external financing market, still holds if we choose a different type of borrowing constraint. Additional quantitative results with different types of borrowing constraints are available upon request.

In Equation (2.8), λ represents the (net) cash carry costs. Although holding cash is beneficial in potentially preventing the firm from current and future underinvestment and external financing costs, corporate cash hoarding is also costly. According to the existing literature, a positive cash carry cost λ could come from many different aspects. First, it could come from the agency costs associated with the free cash problem. For instance, Jensen (1986) argues that managers are more likely to be "empire-building" and willing to take on negative net present value (NPV) investment with free cashflow. Second, λ could come from tax constraints. Cash retention is tax disadvantaged because all the interests earned from cash holdings will be taxed at the corporate tax rate (Graham, 2000). In contrast, debt expenses are tax-deductible. Despite these potential explanations, here we use the reduced-form approach to introduce the cash carry costs, and the carry costs can be summarized as a positive parameter λ.[9] The key mechanism at work is that cash carry costs are linear in its quantity while benefits are increasing (nonlinearly) in future cash flow uncertainty. This trade-off determines the entrepreneur's optimal cash policy and lies at the heart of our story.

Another important feature about Equation (2.8) is this *time-invariant* λ. The implicit assumption here is that the cost of holding cash is immune to changes in its demand. In other words, cash is **completely risk free**. In the model, the reason why cash is different from bond is because there is no such a price for corporate cash so that its demand should be equal to its supply. As a result, the price mechanism will not work when companies hold cash. In reality, corporate cash mostly consists of safe assets. One important feature about safe assets is that their costs and benefits are pre-determined (at least in the eyes of investors) and do not fluctuate according to the changes in the demand. As we will show later, this difference is crucial when we explore the macroeconomic consequences of internal financing. We put more discussions on this assumption in Section 2.4.2.

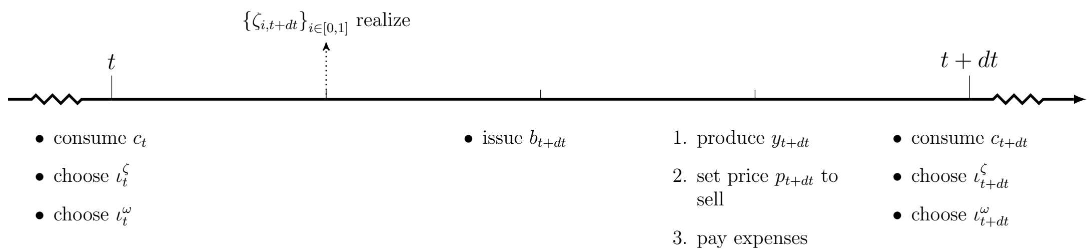
To sum up, the entrepreneur's budget constraint at each period can be stated as in the following equation:

$$c + \iota^k + \iota^\omega - \iota^b = \pi - rb - \mathbf{1}_{b\neq 0}\left(\chi_0 + \chi_1 \left|b\right|\right) - \frac{\kappa_0}{2}\frac{\left(\iota^k\right)^2}{k} \tag{2.9}$$

[9] In the baseline model, we assume that carry corporate cash is costly, i.e., $\lambda > 0$. However, according to Gao, Whited and Zhang (Forthcoming), companies might be able to earn some interests with business sweep programs.

Timeline

The timeline from t to $t + dt$ in this economy can be shown as below:



Right after period t, all entrepreneurs' next-period capital quality $\{\zeta_{i,t+dt}\}_{i\in[0,1]}$ realizes and it becomes a piece of the public information to the whole economy. After that, each entrepreneur needs to determine how much to borrow or lend in the external financial market. Then the firm produces. After production, the entrepreneur needs to pay all the expenses related to debt repayment and issuance costs. Finally, the entrepreneur decides how much to consume, invest in capital quality, and save in corporate cash.

Equilibrium definition

The equilibrium definition can be summarized in Definition 4.1.

Definition 2.1 *A stationary recursive competitive equilibrium consists of prices* $\left\{(p_{i,t})_{i\in[0,1]}, r_t\right\}_{t=0}^{\infty}$ *and allocations* $\left\{\left(y_{i,t}, \iota_{i,t}^{\zeta}, \iota_{i,t}^{\omega}, \iota_{i,t}^{b}, c_{i,t}\right)_{i\in[0,1]}\right\}_{t=0}^{\infty}$ *that satisfy the following conditions:*

1. ***Optimization**: given market prices* $\left\{(p_{i,t})_{i\in[0,1]}, r_t\right\}_{t=0}^{\infty}$*, resource allocations* $\left\{\left(y_{i,t}, \iota_{i,t}^{\zeta}, \iota_{i,t}^{\omega}, \iota_{i,t}^{b}, c_{i,t}\right)_{i}\right.$ *maximize each entrepreneur's life-time utility (4.1) subject to constraints (2.2)-(2.9), and his initial endowment* $(\zeta_{i,0}, \omega_{i,0}, b_{i,0})$.

2. ***Market Clearance**: market prices* $\left\{(p_{i,t})_{i\in[0,1]}, r_t\right\}_{t=0}^{\infty}$ *clear all the markets in this economy*

   - *intermediate goods market:*

$$p(\zeta) = \zeta^{\phi} \tag{2.10}$$

- *bond market:*
$$\int b_{i,t} di = 0 \tag{2.11}$$

- *final goods market:*
$$\mathcal{C}_t + \mathcal{I}_t + \mathcal{G}_t + \mathcal{X}_t = \mathcal{Y}_t \tag{2.12}$$

where $\mathcal{Y}_t$, $\mathcal{C}_t$, $\mathcal{I}_t$, $\mathcal{G}_t$, and $\mathcal{X}_t$ are the aggregated output, aggregated consumption, aggregated "investment", aggregated adjustment costs, and aggregated external financing costs, respectively. Mathematically, they are calculated as follows

$$\begin{aligned}
\mathcal{Y}_t &= \int_0^1 \zeta_{i,t}^{\phi} y_{i,t} di \\
\mathcal{C}_t &= \int_0^1 c_{i,t} di \\
\mathcal{I}_t &= \int_0^1 \iota_{i,t}^{\zeta} di + \int_0^1 \iota_{i,t}^{\omega} di \\
\mathcal{G}_t &= \int_0^1 \frac{\kappa_0}{2} \left(\frac{\iota_{i,t}^{\zeta}}{\zeta_{i,t}} \right)^2 \zeta_{i,t} di \\
\mathcal{X}_t &= \int_0^1 \mathbf{1}_{b_{i,t} \neq 0} \left(\chi_0 + \chi_1 \left| b_{i,t} \right| \right) di
\end{aligned}$$

3. ***Stationary distribution**: the state of the economy $\Lambda_t(\zeta, \omega, b)$ is stationary. At each time t, the transition of $\Lambda_t(\zeta, \omega, b)$ should be consistent with each entrepreneur's optimal policy function $\left\{p^*, l^*, \iota^{k,*}, \iota^{\omega,*}, \iota^{b,*}, c^*\right\}$ and satisfies the following Kolmogorov forward equation[10]:*

$$\frac{\partial \Lambda_t(\zeta,\omega,b)}{\partial t} = -\frac{\partial}{\partial \zeta}\left[\mu^{\zeta,*}(\zeta)\Lambda_t(\zeta,\omega,b)\right] - \frac{\partial}{\partial \omega}\left[\mu^{\omega,*}(\omega)\Lambda_t(\zeta,\omega,b)\right] - \frac{\partial}{\partial b}\left[\mu^{b,*}(b)\Lambda_t(\zeta,\omega,b)\right] + \frac{\partial^2}{\partial \zeta^2}\left[\sigma^{\zeta,*}(\zeta)^2\Lambda_t(\zeta,\omega,b)\right] \tag{2.13}$$

[10] The derivation of Kolmogorov forward equation follows Stokey (2008).

where

$$
\begin{aligned}
\iiint \Lambda_t(\zeta,\omega,b)\, d\zeta d\omega db &= 1,\ \textit{where}\ \Lambda_t(\zeta,\omega,b) \geq 0, \forall t \\
d\zeta = \mu^{\zeta,*}(\zeta)\, dt + \sigma^{\zeta,*}(\zeta)\, dZ &= \left(\bar{\mu} + \iota^{\zeta,*} - \delta\zeta\right) dt + \sigma\sqrt{\zeta} dZ \\
d\omega = \mu^{\omega,*}(\omega)\, dt &= \left(\iota^{\omega,*} - \lambda\omega\right) dt \\
db = \mu^{b,*}(b)\, dt &= \iota^{b,*} dt
\end{aligned}
$$

By definition, stationary equilibrium means that $\frac{\partial \Lambda_t(\zeta,\omega,b)}{\partial t} = 0$*, i.e., Equation (2.13) becomes*

$$
0 = -\frac{\partial}{\partial \zeta}\left[\mu^{\zeta,*}(\zeta)\Lambda_t(\zeta,\omega,b)\right] - \frac{\partial}{\partial \omega}\left[\mu^{\omega,*}(\omega)\Lambda_t(\zeta,\omega,b)\right] - \frac{\partial}{\partial b}\left[\mu^{b,*}(b)\Lambda_t(\zeta,\omega,b)\right] + \frac{\partial^2}{\partial \zeta^2}\left[\sigma^{\zeta,*}(\zeta)^2\Lambda_t(\zeta,\omega,b)\right] \tag{2.14}
$$

2.1.2 Risky Superstar Economy

After describing the basic setup, now we turn to discuss two key model implications. The first key implication is that the endogenous monopolistic competition, combined with the *exogenous* and *homogeneous* stochastic capital quality shocks, generates an *endogenous* and *quality-based non-homogeneous* earnings process for all the entrepreneurs in this economy. More importantly, the difference between these two processes are governed by demand and supply factors. In this way, shifts in the demand and supply curves are redistributive in not only income and wealth but also risks.

Origins of markup We begin with discussing how a firm's markup μ and earnings π are linked to its underlying capital quality ζ. In this paper, the definition of a firm's markup μ is the ratio of profits to total costs, i.e., $\mu = \frac{py}{f_0+\xi_0 y^{\frac{1}{\eta}}}$. Meanwhile, the definition of a firm's earnings π is simply the difference between profits and total costs, i.e., $\pi = py - f_0 - \xi_0 y^{\frac{1}{\eta}}$. We summarize the key results related to the origins of markup in Lemma 2.1. All the proofs are provided in Appendix A.1.

Lemma 2.1 *With the optimization problem shown in Equation (2.4), the markup and earnings for the firm with capital quality* ζ *are shown as follows*

$$\mu = \frac{py}{f_0 + \xi_0 y^{\frac{1}{\eta}}} = \frac{1}{\eta + f_0 \left[\zeta^{\frac{\phi}{1-\eta}} \left(\frac{\eta}{\xi_0}\right)^{\frac{\eta}{1-\eta}}\right]^{-1}} \tag{2.15}$$

$$\pi = (1-\eta)\left(\frac{\eta}{\xi_0}\right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi}{1-\eta}} - f_0 \tag{2.16}$$

Three important conclusions can be drawn from Lemma 2.1. First, both markup and earnings are increasing in capital quality ζ. This outcome is quite intuitive given the model setup used in this paper. To begin with, we assume that consumers in this economy have a taste for quality. As a result, entrepreneurs with better product quality face a higher demand, which makes them able to charge higher prices and sell more products. At the same time, due to the existence of a fixed cost f_0, the total cost per goods is lower if the number of products sold is higher, even if entrepreneurs charge the same price. Hence, demand and supply sides both contribute to markup being increasing in capital quality. Since markup is positively related to earnings, we can obtain the same conclusion for the earnings-quality relationship.

Second, in this paper, the origins of superstar firms come from shifts in demand and supply curves. As Lemma 2.1 shows, the earnings-markup relationship is eventually pinned down by the preference parameter ϕ and some production technology parameters such as η. Changes in supply and demand curves lead to substantial impacts on the distribution of markup. For example, if we live in an economy with low preference over product quality and very costly supply, i.e., $\phi + \eta \leq 1$, then we should expect the markup and earnings to be a concave or linear function of capital quality ζ. In this economy, although markup and earnings are still increasing in capital quality, no single entrepreneur strictly dominates the rest in the market. However, if there are some permanent shifts in demand and supply that makes $\phi + \eta > 1$, then the markup and earnings become a convex function of capital quality ζ. In this new economy, consumers have very strong preference over these high-quality products, and at the same time, it does not cost much for these entrepreneurs to serve all the potential customers. In this situation, superstars can make the best use of their market power and become dominate in the whole industry.

Third, our story here relies crucially on the existence of fixed costs f_0, but we argue that this assumption is quite realistic, especially for the new intangible economy. For example, Hsieh and Rossi-Hansberg (2019) argue that only after paying a fixed cost such as R&D, intangibles like software can be deployed across different markets. In addition, De Ridder (2019) directly model intangibles as inputs that cause a shift from marginal to fixed costs. He finds that this way of interpreting the intangible economy can help us jointly explain the slowdown of productivity growth, the decline in business dynamism, and the rise of market power. More importantly, our model setup is consistent with his empirical finding that there is a significant increase in fixed costs for the U.S. firms. More specifically, among the U.S. public firms, the share of fixed costs in total costs has increased from 13.9% in 1980 to 24.5% in 2015.[11]

Quality-based non-homogeneous earnings process With the previous result, we can easily show that with the endogenous monopolistic competition, the actual earnings process are completely different from the underlying capital quality process. We summarize the key results in the following lemma.

Lemma 2.2 *Given the homogeneous underlying process shown as in Equation (2.5), the actual earnings process is a quality-based non-homogeneous one:*

$$d\pi_t = \underbrace{\left[\pi'(\zeta_t)\left(\bar{\mu} + \iota_t^{\zeta} - \delta\zeta_t\right) + \frac{\sigma^2\zeta_t}{2}\pi''(\zeta_t)\right]}_{drift} dt + \underbrace{\pi'(\zeta_t)\,\sigma\sqrt{\zeta_t}}_{volatility} d\mathcal{Z}_t \tag{2.17}$$

where

$$\pi'(\zeta) = \phi\left(\frac{\eta}{\xi_0}\right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi}{1-\eta}-1} \tag{2.18}$$

$$\pi''(\zeta) = \phi\left(\frac{\phi}{1-\eta} - 1\right)\left(\frac{\eta}{\xi_0}\right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi}{1-\eta}-2} \tag{2.19}$$

As we can see from the lemma above, as long as $\phi + \eta \neq 1$, the earnings dynamics $d\pi$ are different from the underlying capital quality dynamics $d\zeta$: $d\pi$ depends on the current level of capital quality as entrepreneurs with different capital qualities face different π'. Therefore, the expected future earnings and expected volatility of future earnings will be various across firms with different capital qualities. In this way, the curvature of

[11] However, we do not limit our story to intangible capital only because this pattern shows up for manufacturing firms as well. For example, the recent robotic automation creates efficiencies to the operation scale but requires manufacturing firms to pay more on fixed production costs.

earnings-quality relationship $\frac{\phi}{1-\eta}$ not only determines the income distribution but also the risk distribution.[12]

Impacts of shifts in demand and supply curves Now we turn to investigate how shifts in demand and supply can generate a risky Superstar Economy, where both income and risk are redistributed towards entrepreneurs with better capital quality. The key conclusions here can be summarized in the following proposition.

Proposition 2.1 *Given that there is a fundamental change that increases both the taste for quality ϕ and the operation scale η such that $\phi + \eta > 1$, then the economy transits from a traditional economy into a Superstar Economy. This fundamental change brings the following two consequences:*

1. *earnings $\pi(\zeta)$ are convex in capital quality ζ; entrepreneur's output $y(\zeta)$ is also a convex function of capital quality ζ if $\phi > \frac{1-\eta}{\eta}$*

2. *in steady state, drift and volatility components of the entrepreneur's earnings process $d\pi$ are both increasing in capital quality ζ*

The formal proof is provided in Appendix A.1 and here we focus on the main intuition. As can be seen from Proposition 2.1, increases in consumers' taste for quality and companies' scale of operation will move the whole economy towards a winner-take-most extreme. In this new economy, people have stronger preference over the best-quality products. At the same time, it is not costly for these firms with the best products to serve all the buyers. Therefore, with these changes in demand and supply, output and earnings are more distributed towards the right-tail firms and they will dominate in the economy.

[12] Our conclusion here does not depend on the specific choice of the underlying capital quality process. The generalization can be best illustrated with Ito's lemma. Assume that each entrepreneur's ability ζ follows any stochastic differential equation $d\zeta = f(\zeta)\,dt + g(\zeta)\,d\mathcal{W}$, and earnings π are a function of ζ, i.e., $\pi = \pi(\zeta)$. By Ito's Lemma, we can derive the entrepreneur's earning as the following stochastic process:

$$d\pi = \pi'(\zeta)\,d\zeta + \frac{1}{2}\pi''(\zeta)\,g(\zeta)^2\,dt = \left[\pi'(\zeta)\,f(\zeta) + \frac{1}{2}\pi''(\zeta)\,g(\zeta)^2\right]dt + \pi'(\zeta)\,g(\zeta)\,d\mathcal{W} \tag{2.20}$$

Based on the equation above, as long as π is not a linear function of ζ, i.e., π' is not independent of ζ, regardless of the underlying capital quality process, the actual earnings process that eventually affect entrepreneur's investment and cash holding decisions will be a quality-based and non-homogeneous one.

Mathematically, the rise of this risky superstars comes from the fact that shifts in demand and supply curves make earnings become a convex function of capital quality. More specifically, due to the convexity of π with respect to ζ, two implications arise directly from the earnings process (2.17). On the one hand, $\pi'(\zeta)$ on the drift coefficient shows that star firms can become superstars, which is also the common focus of superstar effects in the existing literature. On the other hand, $\pi'(\zeta)$ on the diffusion coefficient means that superstar firms are inherently riskier.[13] Therefore, changes in economic fundamentals affect both the drift and diffusion parts simultaneously, which means that superstar firms can take most but at the cost of bearing more earnings uncertainty in the future. In this way, these fundamental changes are redistributive in both income and risks. In other words, superstar firms are not merely large versions of small firms as they face the highest expected future earnings but also the most volatile income fluctuations. This unique characteristic of a Superstar Economy will substantially change the entrepreneur's cash holdings and investment decisions.

2.1.3 Endogenous Firm-Market Boundary

The second key model implication is that entrepreneur's optimal choice between internal and external financing leads to an endogenous firm-market boundary. The main results related to entrepreneur's optimal risk management policy can be summarized in Proposition 2.2.

Proposition 2.2 *Given the model setup, the optimal policy of an entrepreneur with capital quality ζ involves a quality-based downward control boundary $\overline{\Omega}^{\zeta}$ and upward control boundary $\underline{\Omega}^{\zeta}$. The existence of these control boundaries splits the whole economy into three different regions:*

1. ***External lending region**: within this region, the entrepreneur has accumulated enough cash and decides to lend in the external financial market, i.e., $\omega = \overline{\Omega}^{\zeta}$ and $b < 0$. The Hamilton-Jacobi-Bellman (HJB) equation for this external lending*

[13] The main focus is to argue that superstar firms today are riskier than their counterparts in the 1980s. Whether superstar firms are riskier than small firms today depends on the functional form of the underlying capital quality process.

region is

$$0=\max_{\iota^{\zeta},c}\left\{\begin{array}{l} f\left(c,\mathcal{J}\right)+\left(\bar{\mu}+\iota^{\zeta}-\delta\zeta\right)\mathcal{J}_{\zeta}+\frac{\zeta\sigma^{2}}{2}\mathcal{J}_{\zeta\zeta}-\zeta\sigma^{2}\pi'\left(\zeta\right)\mathcal{J}_{\zeta b}+\frac{\zeta\sigma^{2}}{2}\left(\pi'\left(\zeta\right)\right)^{2}\mathcal{J}_{bb}\\ -\left[\pi'\left(\zeta\right)\left(\bar{\mu}+\iota^{\zeta}-\delta\zeta\right)+\frac{1}{2}\pi''\left(\zeta\right)\zeta\sigma^{2}-c-\iota^{\zeta}-rb-\chi_{0}+\chi_{1}b-\lambda\overline{\Omega}^{\zeta}-\frac{\kappa_{0}\left(\iota^{\zeta}\right)^{2}}{2\zeta}\right]\mathcal{J}_{b}\end{array}\right\} \tag{2.21}$$

where c and ι^{ζ} satisfy the following first-order conditions

$$f_{c}\left(c,\mathcal{J}\right) = -\mathcal{J}_{b} \tag{2.22}$$

$$-\frac{\mathcal{J}_{\zeta}}{\mathcal{J}_{b}}+\pi'\left(\zeta\right) = 1+\kappa_{0}\frac{\iota^{\zeta}}{\zeta} \tag{2.23}$$

2. ***External borrowing region**: in this region, entrepreneur holds a constant amount of cash and chooses to borrow externally, i.e., $\omega=\underline{\Omega}^{\zeta}$ and $b>0$. The corresponding HJB equation for this external borrowing region is*

$$0=\max_{\iota^{\zeta},c}\left\{\begin{array}{l} f\left(c,\mathcal{J}\right)+\left(\bar{\mu}+\iota^{\zeta}-\delta\zeta\right)\mathcal{J}_{\zeta}+\frac{\zeta\sigma^{2}}{2}\mathcal{J}_{\zeta\zeta}-\zeta\sigma^{2}\pi'\left(\zeta\right)\mathcal{J}_{\zeta b}+\frac{\zeta\sigma^{2}}{2}\left(\pi'\left(\zeta\right)\right)^{2}\mathcal{J}_{bb}\\ -\left[\pi'\left(\zeta\right)\left(\bar{\mu}+\iota^{\zeta}-\delta\zeta\right)+\frac{1}{2}\pi''\left(\zeta\right)\zeta\sigma^{2}-c-\iota^{\zeta}-rb-\chi_{0}-\chi_{1}b-\lambda\underline{\Omega}^{\zeta}-\frac{\kappa_{0}\left(\iota^{\zeta}\right)^{2}}{2\zeta}\right]\mathcal{J}_{b}\end{array}\right\} \tag{2.24}$$

where c and ι^{ζ} satisfies the following first-order conditions:

$$f_{c}\left(c,\mathcal{J}\right) = max\left\{-\mathcal{J}_{b},-\mathcal{J}_{b}|_{b=\beta\frac{\pi-\chi_{0}}{1+r+\chi_{1}}}\right\} \tag{2.25}$$

$$1+\kappa_{0}\frac{\iota^{\zeta}}{\zeta} = \frac{\mathcal{J}_{\zeta}}{max\left\{-\mathcal{J}_{b},-\mathcal{J}_{b}|_{b=\beta\frac{\pi-\chi_{0}}{1+r+\chi_{1}}}\right\}}+\pi'\left(\zeta\right) \tag{2.26}$$

3. ***Internal financing region**: within this region, firms finances their investment using internal funds only, i.e., $b=0$ and $\underline{\Omega}^{\zeta}<\omega<\overline{\Omega}^{\zeta}$. The HJB equation for this internal financing region is*

$$0=\max_{\iota^{\zeta},c}\left\{\begin{array}{l} f\left(c,\mathcal{J}\right)+\left(\bar{\mu}+\iota^{\zeta}-\delta\zeta\right)\mathcal{J}_{\zeta}+\frac{\zeta\sigma^{2}}{2}\mathcal{J}_{\zeta\zeta}+\zeta\sigma^{2}\pi'\left(\zeta\right)\mathcal{J}_{\zeta\omega}+\frac{\zeta\sigma^{2}}{2}\left(\pi'\left(\zeta\right)\right)^{2}\mathcal{J}_{\omega\omega}\\ +\left[\pi'\left(\zeta\right)\left(\bar{\mu}+\iota^{\zeta}-\delta\zeta\right)+\frac{1}{2}\pi''\left(\zeta\right)\zeta\sigma^{2}-c-\iota^{\zeta}-\lambda\omega-\frac{\kappa_{0}\left(\iota^{\zeta}\right)^{2}}{2\zeta}\right]\mathcal{J}_{\omega}\end{array}\right\} \tag{2.27}$$

where c and ι^{ζ} satisfy the following first-order conditions

$$f_c(c,\mathcal{J}) = \mathcal{J}_\omega \tag{2.28}$$

$$\mathcal{J}_\zeta = \left[1+\kappa_0\frac{\iota^\zeta}{\zeta}-\pi'(\zeta)\right]\mathcal{J}_\omega \tag{2.29}$$

In addition, the accompanying boundary conditions consist of the Neumann boundary conditions in the ζ-dimension

$$\mathcal{J}_\zeta(\zeta_{min},\omega,0) = 0,\ \forall\omega \tag{2.30}$$

$$\mathcal{J}_\zeta(\zeta_{max},\omega,0) = 0,\ \forall\omega \tag{2.31}$$

the Neumann boundary conditions in the ω-dimension

$$\mathcal{J}_\omega\left(\zeta,\underline{\Omega}^\zeta,0\right) = 1+r+\chi_1,\ \forall\zeta \tag{2.32}$$

$$\mathcal{J}_\omega\left(\zeta,\overline{\Omega}^\zeta,0\right) = 1+r-\chi_1,\ \forall\zeta \tag{2.33}$$

$$\mathcal{J}_{\omega\omega}\left(\zeta,\overline{\Omega}^\zeta,0\right) = 0,\ \forall\zeta \tag{2.34}$$

and the Neumann boundary conditions in the b-dimension

$$\mathcal{J}_\omega\left(\zeta,\underline{\Omega}^\zeta,0\right) = -\mathcal{J}_b\left(\zeta,\underline{\Omega}^\zeta,0\right),\ \forall\zeta$$

$$\mathcal{J}_\omega\left(\zeta,\overline{\Omega}^\zeta,0\right) = \mathcal{J}_b\left(\zeta,\overline{\Omega}^\zeta,0\right),\ \forall\zeta$$

Boundary of the invisible hand Results in Proposition 2.2 can be understood from the perspective of economic inaction. Generally speaking, whenever behaviors entail a fixed cost, inaction is the norm (Stokey, 2008). In this paper, due to the (fixed) transaction costs of external financing, entrepreneurs find it optimal not to participate in any market-based lending or borrowing activities within some regions. As the entrepreneur's optimization problem is a standard cash inventory model, the optimal policy consists of multiple control barriers (Miller and Orr, 1966; Bolton, Chen and Wang, 2011). The proof for these boundary conditions is shown in the appendix. What is interesting here is the "unintended" consequence: these control boundaries, arising from individual entrepreneur's optimal decision, segment the whole economy into several different regions and create an endogenous firm-market boundary. To formally establish this firm-market boundary from the aggregate perspective, we introduce the following definition on the boundary of the invisible hand.

Definition 2.2 *The* ***boundary of the invisible hand*** *is made of a set of upward and downward control boundaries* $\left\{\overline{\Omega}^i, \underline{\Omega}^i\right\}_{i\in[0,1]}$*. For each entrepreneur* $i \in [0,1]$*,* $\overline{\Omega}^i$ *and* $\underline{\Omega}^i$ *are the solutions to the Neumann boundary conditions (2.32)-(2.34) of a nonlinear partial differential equation (2.27). At time* t*, the area controlled by the price mechanism, i.e., the fraction of entrepreneurs outside the internal financing region, can be calculated as follows:*

$$\Psi_t \equiv \int \left(1 - \mathbb{I}_{\overline{\Omega}^i<\omega^i<\underline{\Omega}^i}\right) di = \iiint \left(1 - \mathbb{I}_{\overline{\Omega}^\zeta<\omega<\underline{\Omega}^\zeta}\right) \Lambda_t \left(\zeta, \omega, b\right) d\zeta d\omega db \qquad (2.35)$$

where $\mathbb{I}$ *is an indicator function.*

The idea behind Definition 2.2 is from Coase (1937)'s work: the transaction cost of using the price mechanism leads to the existence of firms as a different sub-economy. Although Coase (1937)'s idea is quite refreshing, it did not get enough attention until Williamson (1979, 1981) formally establishes a formal theoretical framework with the incomplete contracts approach. In contrast, in this paper, we introduce the transaction cost in a reduced-form but add it to a heterogeneous agent general equilibrium framework so that we can investigate the aggregate implications of firm-market boundary. The other major difference between this paper and the existing literature is that most studies are focused on the firm's production boundary. However, what companies do is not only produce goods, but also perform actively in the financial market. Therefore, in this paper, we formally introduce the boundary between firms and the financial market. We argue that corporate internal financing tool is an imperfect substitution of external financing tools. Although corporate cash can help achieve dynamic efficiency across times at the firm-level, it cannot obtain static efficiency across agents at the macro-level. As a result, we have unequalized marginal cost of capital within this internal financing region, which is directly linked to capital misallocation.

With Definition 2.2, we can use Equation (2.35) to keep track of the area disciplined by the price mechanism within the model. There are indeed two different sets of allocation systems in this economy: firms and the price mechanism. Intuitively, Ψ measures the fraction of firms outside the internal financing region. One of the main interests in this paper is to investigate how these changes in demand and supply curves affect the relative size of Ψ through their impacts on the earnings process. More importantly, we should expect Ψ to shrink in the Superstar Economy. The intuition is the following. The value of internal financing comes from the entrepreneur's ability to access and restructure its financing at a low cost. As a result, cash value increases in future cash flow uncertainty. In a risky Superstar Economy with the income and risk redistribution, financial flexibility becomes more valuable, especially for superstar firms. Therefore, we should expect that these firms will rely more on internal financing, and the boundary of the invisible hand will shrink. As it is challenging to get the closed-form solutions, we perform a quantitative exercise to estimate the magnitudes of these impacts.

HJB equation In this part, we briefly explain the determinants of an entrepreneur's HJB equation. Please refer to the appendix for details on explaining boundary conditions. Although this economy has three assets, i.e., productive capital, corporate cash, and risk-free debt, due to these endogenous control boundaries, firms are performing an optimal portfolio selection with at most two assets within each region. This setup simplifies the model and the numerical solution to a large extend. Here we use results in the internal financing region as an example. Analyses on the other two regions are similar.

Within the internal financing region, entrepreneurs finance their investment out of the cash inventory. The HJB equation is shown in Equation (2.27). In this equation, the $\mathcal{J}_\zeta$ term on the right-hand side is the marginal effect of net capital investment on firm value and the $\mathcal{J}_{\zeta\zeta}$ term represents the effect of investment risk on firm value. Similarly, $\mathcal{J}_\omega$ denotes the cash value, and $\mathcal{J}_{\omega\omega}$ is the effect of the volatility of cash holdings on firm value. At the same time, the cross-partial derivative term $\mathcal{J}_{\zeta\omega}$ represents the effects of additional investment on the business marginal cash value. $\mathcal{J}_{\zeta\omega}$ is expected to be positive as additional investment increases the capital quality and thus the underlying risk faced by the entrepreneur.

Compared to the standard work in this branch of literature (e.g., Bolton, Chen and Wang, 2011; Wang, Wang and Yang, 2012), what is new in Equation (2.27) is threefold. First, there are capital quality shocks in this economy, and this setup adds two additional terms ($\mathcal{J}_{\zeta\zeta}$ and $\mathcal{J}_{\zeta b}$) to the HJB equation as any capital investment also contains additional risk. In contrast, other works in this literature assume the i.i.d productivity shocks within $(t, t + dt)$ period. In this way, there is only one $\mathcal{J}_\zeta$ term in their HJB equations. Second, as pointed out before, the endogenous monopolistic competition generates a quality-based non-homogeneous earnings process for entrepreneurs. Therefore, we can not rescale the state variable and change a PDE into an ODE problem as we lose the homogeneity in this economy. In other words, we can no longer treat superstar firms merely as a large version of small firms. Third, capital investment ι^{ζ} also shows up in the $\mathcal{J}_{\omega}$ term here, which means that the entrepreneur's investment behavior will directly affect the value of cash. The intuition here is that any capital investment behavior is likely to change an entrepreneur's capital quality in the economy, thus affecting both the expected level and volatility of future earnings.

First-order conditions Now we turn to explaining the first-order conditions derived from the entrepreneur's optimal decision. Again, we use the internal financing region as an example. The optimal consumption choice is shown in Equation (2.28), where the marginal utility of consumption should be equal to the marginal value of cash. This outcome is intuitive as entrepreneurs can freely choose to save their earnings as cash for the future or consume at this moment. Therefore, the marginal benefits of these two choices should be equal at the optimum.

At the same time, the first-order condition concerning investment is shown in Equation (2.29). The convexity of the physical adjustment cost implies that the model's investment decision admits an interior solution. On the left-hand side is the marginal benefits of investment, i.e., $\mathcal{J}_{\zeta}$. $\mathcal{J}_{\zeta}$ is also called marginal q, which represents the marginal benefit of adding one unit of capital. On the right-hand side is the marginal cost of investment. Generally speaking, the marginal cost of investment has mainly two components. To begin with, in order to increase the capital quality by one additional unit, the entrepreneurs need to pay $1 + \kappa_0 \frac{\iota^{\zeta}}{\zeta}$ out of their pocket. As in this region, entrepreneurs finance their investment fully out of cash. Therefore, the actual marginal costs are the product of $1 + \kappa_0 \frac{\iota^{\zeta}}{\zeta}$ and marginal cash value $\mathcal{J}_{\omega}$. Meanwhile, different from the existing literature, in Equation (2.29), investment will also increase the capital quality of firms, which directly affects the degree of earnings uncertainty. Therefore, the entrepreneur's investment decision will increase the marginal value of cash by a number of $\pi'(\zeta)$, and the net marginal cost of investment can be calculated as $\left[1 + \kappa_0 \frac{\iota^{\zeta}}{\zeta} - \pi'(\zeta)\right] \mathcal{J}_{\omega}$.

In the complete market benchmark, the optimal marginal product of capital $\mathcal{J}_{\zeta}$ should be equal to the marginal cost of adjusting the capital stock $1 + \kappa_0 \frac{\iota^{\zeta}}{\zeta}$. In contrast, Equation (2.29) shows that two different kinds of distortions prevent these two indicators from being equalized in equilibrium. First, in incomplete markets with external financing transaction costs, the marginal value of cash is no longer equal to one, and this corporate cash value will affect entrepreneurs' optimal investment. Second, due to the monopolistic competition, entrepreneurs face a quality-based non-homogeneous earnings process. In this way, additional investment will also affect the capital quality, which will increase the amount of risk faced by entrepreneurs. Increased riskiness leads to changes in the marginal value of cash. In addition, the positive cross-sectional correlation between the marginal value of capital $\mathcal{J}_\zeta$ and the marginal value of cash $\mathcal{J}_\omega$ is also a new characteristic of the model. With these elements, corporate investment becomes less sensitive to marginal q, especially for these superstar firms.

2.2 Reduced-form Evidence

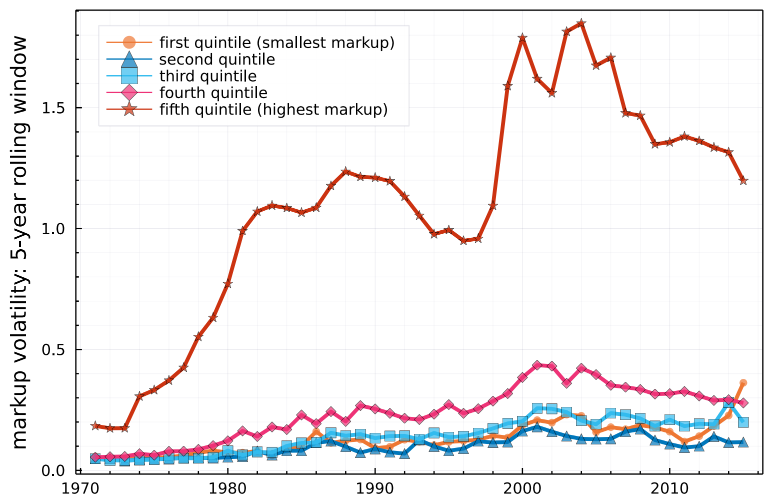
Here we provide some empirical evidence that is consistent with the model implications derived in the previous section. Of course, all these reduced-form empirical results only suggest correlation instead of causality.

The data used in this section and also in the following quantitative analysis section come primarily from the Compustat North America Fundamentals dataset. Explanations on variable constructions in detail are presented in Appendix A.2.

2.2.1 Risky Superstars

The first testable implication from our model is that superstar firms are riskier than the other firms. Here we provide some empirical evidence to support this claim. In Figure 2.3, we plot the time series of markup volatility for five groups of firms with different levels of markup. Consistent with the model, we define superstar firms to be those with the highest markup. We obtain the results in Figure 2.3 using the following steps. First, for each individual company in each period, we compute its markup volatility by using a five-year rolling window. Second, in each year, we classify all the firms with available markup measures into five different quantiles according to their markup level. Finally, we calculate the annual average markup volatility for each quantile of firms. As shown in Figure 2.3, on average, the group of firms with the highest markup level also has the largest degree of markup volatility. Over the sample period, the average markup volatility for the lowest quantile group is 0.129. In contrast, superstar firms have an average markup volatility as high as 1.097, which is a substantially higher number. In addition, there is a clear upward trend until 2004 in the measured degree of riskiness for the superstar firms. It indicates that compared to the rest of the firms, superstars are over time becoming much riskier, which is also consistent with the model implications.[14]

Figure 2.3: Risky superstars



Notes: This figure presents the time-series plots of the markup volatility for five groups of firms with different levels of markups. Main data source for this figure is the Compustat North American Annual data file. Firm-level markup is estimated with Loecker, Eeckhout and Unger (2020)'s production cost function approach, and markup volatility is estimated as a five-year rolling window.

Two caveats are worth noting for interpreting the results here. First, our finding seems to contradict with Herskovic et al. (2016), which found that large firms have small total return or sale growth volatility. This difference comes from the fact that we have different definitions of superstar firms and riskiness. In Herskovic et al. (2016), large firms are defined as firms with more total assets. In contrast, we define superstar firms as those with higher markup. Moreover, Herskovic et al. (2016) measure the riskiness of firms by using the total return volatility or sales growth volatility. In contrast, we adopt markup volatility as the measure for riskiness.

Second, there is a crucial difference between what the model actually implies and what we are able to observe from the data. In the model, we argue that superstar firms

[14]In addition, our empirical finding here on the positive association between market power and cash flows riskiness is consistent with some recent papers such as Dou, Ji and Wu (2021) and Dou, Ji and Wu (Forthcoming).

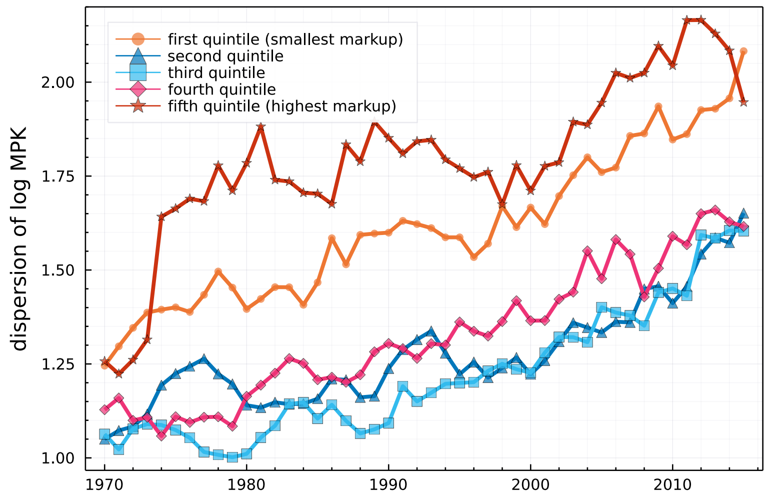
should carry out more risk management because they are inherently riskier *ex ante*. However, the model does not imply that after adopting these risk management policies, superstar firms are still riskier *ex post*. In addition, in practice firms could develop different abilities to hedge their idiosyncratic risks with various financial instruments, which could lead to diametrically opposite conclusions if we focus on different aspects of an equilibrium outcome. For example, the low turnover rate of dominant firms could come from their successful risk management instead of the fact that they are not risky.

2.2.2 Markup and Misallocation

The second testable implication is that dynamics of misallocation are different if we look at firms with different markup levels. In Figure 2.4, we again classify all the firms into five different quintiles as what we did in Figure 2.3. This time, however, we investigate the dispersion of *mpk* within each group of firms. As shown in Figure 2.4, superstar firms not only have on average the largest degree of misallocation, but they have also experienced the rapidest decline in capital allocation efficiency since the 1970s. This result is counter-intuitive if we attempt to understand this phenomenon from the traditional financial friction perspective, which seems to suggest that superstar firms face more financial frictions over time. Still, it is unlikely to happen in reality because we have since the 1980s gone through decades of financial deregulation; however, if we understand this empirical pattern from the previous model, misallocation is increasing for superstar firms because they rely less on external financing. Since the price mechanism will not work for firms within the internal financing region, capital allocation efficiency is declining among these firms.

According to Figure 2.4, another group that also has a relatively higher capital misallocation is the one with smallest markup. One possible explanation for this outcome is exactly the classical financial friction and misallocation story (e.g., Midrigan and Xu, 2014), where finance frictions generate dispersion in the returns to capital because firms cannot borrow as much as they want. The coexistence of these two types of firms indicates that self-financing could have different implications for misallocation. If firms save internally due to financial frictions, then this behavior *reduces* misallocation as it allows firms to invest out of the financial constraints. However, if companies self-finance due to the transaction costs of external financing, then self-financing *increases* misallocation. Therefore, we need to demystify origins of corporate savings when investigating their impacts on misallocation. In addition, it also indicates why it could be misleading if we

Figure 2.4: Markup and misallocation



Notes: This figure presents the time-series plots of misallocation for five groups of firms with different levels of markups. Main data source for this figure is the Compustat North American Annual data file. Firm-level markup is estimated with Loecker, Eeckhout and Unger (2020)'s production cost function approach. Misallocation is defined as the dispersion of firm-level log marginal product of capital mpk and mpk is calculated as the log difference between firm's reported sales (Compustat series $SALE$) and the total net value of property, plant, and equipment (Compustat series $PPENT$).

simply use corporate cash holding as a proxy for measuring the firm-market boundary. Increasing corporate cash-hoarding behaviors could come either from changes in the distortions within the market system, or from movements of the boundary.

One supporting piece of evidence for this claim is that if we investigate the dynamics of misallocation among firms with different cash-to-asset ratios, it turns out that there is not any significant difference among them. As shown in Figure A11 in the appendix, when we classify firms into different groups according to their cash holdings, misallocation patterns are similar among them. This result further confirms the previous hypothesis that the origin of self-financing matters for its impact on misallocation.

2.2.3 Markup and Cash Holding

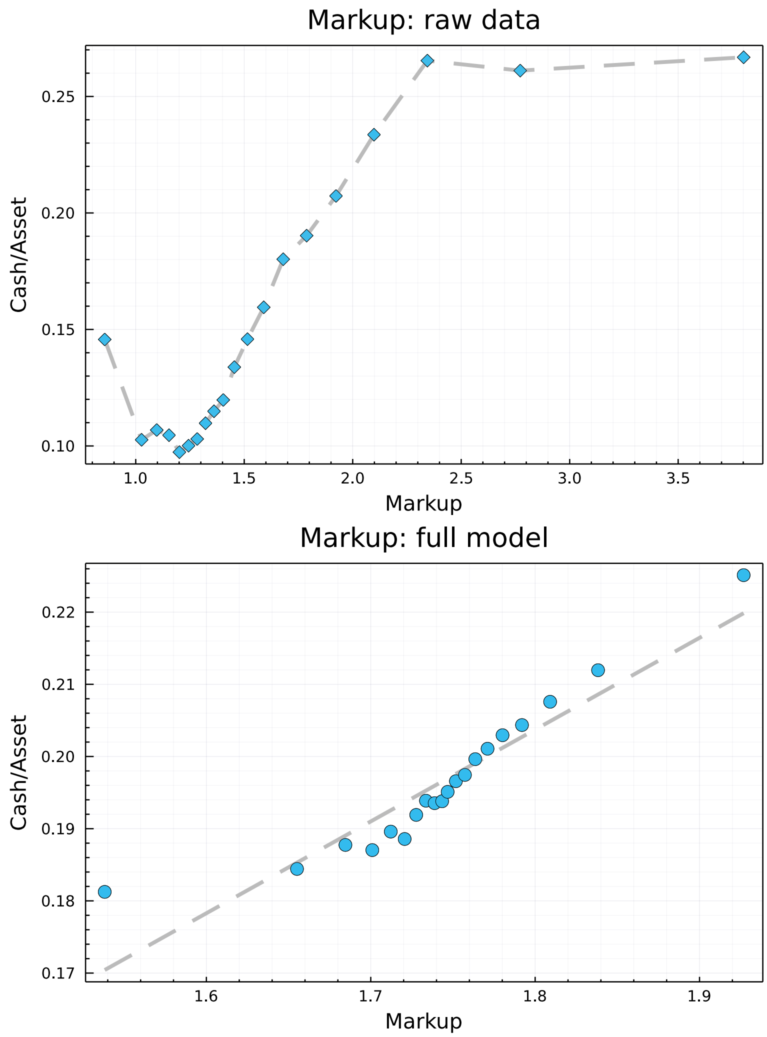
The third testable implication is that cross-sectionally, there should be a positive relationship between firm-level markup and cash holdings. In the previous model, we show that the endogenous quality-based earnings process leads to both the rise of superstar firms and their inherent riskiness, which generates a positive correlation between the firm's markup (or marginal q) and cash value. This cash-markup story distinguishes this paper from all the existing works on explaining why companies hold cash. Therefore, here we provide some related reduced-form evidence for this prediction. Following the existing literature, we measure firm-level cash holdings by using the cash-to-asset ratio and obtain the firm-level markup estimate by using De Loecker, Eeckhout and Unger (2020)'s approach. In addition, we use both the classic Tobin's q and Peters and Taylor (2017)'s Total q measures as proxies for firm-level marginal q. In the main context, we only discuss the empirical results related to markup-cash relationship. q-cash results are provided in the appendix.

Initial evidence with Binscatter plots

Before turning to the formal regression analysis, we can exploit the advantage of Binscatter plots to help clarify the (possibly nonlinear) relationship between cash and markup. Figure 2.5 presents the main result. We provide two graphs: one is plotted with the raw data and the other with a preferred model specification used in the later regression analysis. Our goal with these plots is twofold. First, one can easily eyeball whether there is a positive relationship in the data and, more importantly, whether such a relationship has some degree of nonlinearity. Second, we can use the Binscatter plot to show how fitted the values of a regression equation for better interpretations and evaluations on the model specifications.

As shown in the top panel of Figure 2.5, in the raw data, on average, a firm's cash-to-asset ratio is increasing in its markup, provided that its markup is not too low. Therefore, the data does suggest that on average firms with higher markups hold more liquid cash reserves, which is consistent with the theoretical implication derived previously. At the same time, it shows that this relationship is not always monotone cross-sectionally. Among the firms with the lowest markup, their cash holdings are decreasing in markup. This pattern for the left-tail firms is consistent with the financial frictions story, which argues that firms accumulate cash because they have a higher probability of default or are more likely to face market frictions such as borrowing constraints. As a result, there exists a negative relationship between a firm's cash holding and its financial wealth. Figure 2.5 suggests that this financial frictions story can help make sense of the cash-markup relationship for the left-tail firms, while our story can better explain cash holding behaviors for firms in the right-tail distribution.

Figure 2.5: Tobin's q, total q, markup, and corporate cash holdings: Binscatter plots



Notes: This figure presents the Binscatter plots between corporate cash holdings and markup. Main data source for this figure is the Compustat North American Annual data file. Firm-level cash-to-asset ratio is measured as the ratio of cash and short-term investments (Compustat series *CHE*) to firm's lagged total assets (Compustat series AT). Firm-level markup is estimated with Loecker, Eeckhout and Unger (2020)'s production cost function approach.

The Binscatter plots for cash-Tobin-q and cash-Total-q relationships are provided in Figure A9 and Figure A10, respectively. Based on these two graphs, on average, firms with relatively higher q hold more cash on their balance sheets.

Regression analysis

Table 2.1 shows the regression results for investigating the relationship between markup and cash holdings. We adopt two-way fixed-effects regressions to estimate the impacts of markup on corporate cash holding behaviors. The general model specification in Table 2.1 is shown as follows:

$$\text{cash}_{i,t} = \alpha + \beta \times \mu_{i,t} + (\gamma \times \mu_{i,t}^2) + \Gamma X_{i,t} + \delta_i + \eta_t + \varepsilon_{it}$$

Throughout this section, i and t refer to firm and year, respectively. cash is the firm's cash-to-asset ratio, while μ represents the empirical measure for firm-level markup. We are primarily interested in the sign and statistical significance of the estimated coefficient β. However, as observed in the previous Binscatter plots, there might be some nonlinearity in the relationship between markup and cash holdings. Hence, in some model specifications, we also add the square markup term μ^2 as an additional independent variable. X represents a group of firm-level control variables that could affect corporate cash holdings. Following the standard practice in the existing literature, we include some other firm-level indicators such as return of assets, tangibility, investment, size, profitability, R&D, book leverage, and dividend payout. In addition, we introduce both firm and year-fixed effects to account for the unobserved firm and year characteristics. All standard errors are clustered at the firm level.

Columns (1) - (9) in Panel A of Table 2.1 present the baseline results using fixed-effect regression model, with slight differences in the use of control variables in each column. In the last three columns, we have included all the firm-level controls. The difference between the last three columns comes from the choice of fixed effects. In Column (10), we control for firm and year fixed effects. In Column (11), we include 3-digit NAICS industry and year-fixed effects. In the last column, we introduce industry, year, and industry-year fixed effects. The preferred model specification is the one used in Column (10).

Based on the results shown in Table 2.1, we find that in all model specifications, the estimated coefficients of markup enter with a positive sign at the 1% significance level. This result suggests that firms with higher markup are associated with higher cash-to-asset ratios, which provide additional support for our model. In addition, we can use Binscatter plots to help evaluate how fitted the values of a regression equation are. A good rule of thumb is that if the binned scatter points are close to the regression line,

Table 2.1: Reduced-form evidence: markup and cash holdings

	Cash/Asset											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
markup	0.0092***	0.0080***	0.0269***	0.0086***	0.0054***	0.0080***	0.0113***	0.0080***	0.0083***	0.1177***	0.1783***	0.1763***
	(3.012)	(9.910)	(34.457)	(10.583)	(6.626)	(9.961)	(9.377)	(9.922)	(9.585)	(24.798)	(39.656)	(38.899)
markup square	-0.0002									-0.0160***	-0.0251***	-0.0251***
	(-0.422)									(-18.619)	(-29.197)	(-29.032)
return of assets		-0.0000								0.0002	0.0031***	0.0030***
		(-0.071)								(0.863)	(10.100)	(9.852)
tangibility			-0.3761***							-0.5582***	-0.5386***	-0.5385***
			(-132.893)							(-108.150)	(-118.906)	(-117.821)
investment				-0.0767***						0.0741***	0.2033***	0.2064***
				(-18.388)						(9.165)	(20.450)	(20.854)
size					-0.0106***					-0.0003	-0.0050***	-0.0049***
					(-24.885)					(-0.537)	(-17.237)	(-16.717)
profitability						-0.0003***				-0.0010***	-0.0022***	-0.0022***
						(-3.289)				(-3.683)	(-5.790)	(-5.813)
R&D							-0.0165***			-0.0122***	0.0369***	0.0344***
							(-15.186)			(-8.715)	(21.741)	(20.426)
book leverage								0.0001***		0.0001*	-0.0002***	-0.0002***
								(4.901)		(1.800)	(-3.772)	(-3.376)
payout									0.0020	0.0040	0.0046	0.0047
									(1.180)	(1.623)	(1.394)	(1.406)
	Fixed effects											
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		
Industry (naics3)											Yes	Yes
Industry × Year												Yes
N	198,539	198,537	198,532	195,365	198,539	197,840	101,013	198,539	175,219	90,479	92,091	92,091
Adjusted R^2	0.635	0.635	0.667	0.637	0.636	0.634	0.692	0.635	0.618	0.725	0.336	0.347

the standard error is small, and the estimation is accurate. As is shown in the bottom panel in Figure 2.5, the binned scatter points are largely fitted around the regression line, and they are not very dispersed except for two extreme data points. This graph indicates that the model specifications are reasonably good, and the estimated results are indeed reliable.

In Table A1 in the appendix, we provide the corresponding regression results for the q-cash relationship. Similarly, Table A1 shows that across different model specifications, we can always find a positive and significant association between a firm's cash-to-asset ratio and its q, regardless of the choice on Tobin's q or Total q.

Generally speaking, these three pieces of reduced-form evidence provides additional support for the underlying mechanism in the model section. Now we turn to quantitative analysis.

2.3 Quantitative Analysis

This section outlines how the model is parameterized and then investigates the model's quantitative implications. To start, we discuss the external calibration and structural estimation strategy used in this paper. Next, we show the extent to which our model is able to match both the aggregate trends and cross-sectional patterns in the data. In order to evaluate the quantitative performance of the model, we choose three facts related to the declining capital allocation efficiency in the U.S. It turns out that the model is able to quantitatively match these trends. Finally, we investigate the relative importance of supply factors, demand factors, and financial friction parameters in matching the data.

2.3.1 Parametrization

The model is calibrated at an annual frequency. To reduce the computational burden, we externally calibrate a subset of parameters, then consider the rest estimated within the model. For estimating those structural parameters, we adopt the SMM approach (e.g., McFadden, 1989; Nikolov and Whited, 2014), as there are no closed-form solutions. In addition, following some recent studies, we choose 2000 as the midpoint and split the historical dataset into two different subsamples: the 1980-1999 subsample is interpreted as the traditional economy, and the 2000-2015 subsample is labeled as the superstar economy.

External calibration

Those externally calibrated parameters are shown in the top panel in Table 2.2. The rate of time preference ρ is set to be 0.046. The risk aversion γ is calibrated to 4.0, and the elasticity of intertemporal substitution $\frac{1}{\theta}$ is chosen with a value of 0.5. In addition, the cash carry cost is set to be 1%. All these numbers are standard in the existing literature (e.g., Wang, Wang and Yang, 2012; Bolton, Chen and Wang, 2011).

Some other parameters have their natural data counterparts, so we will use the data to calibrate them. The capital depreciation rate δ can be computed with the fixed asset tables (FAT) obtained from the U.S. Bureau of Economic Analysis (BEA). Consistent with the previous model, we do not limit our analysis here to intangible capital only. Instead, we will calculate the average depreciation rate for all assets. Of course, this paper focuses on firms, so we will only use the capital stock and depreciation data for private sectors. The construction details are as follows. To begin with, by calculating the ratio of depreciation expenses to capital stock, we compute the depreciation rate for intangible and tangible assets, respectively. Then we calculate the annual weighted average depreciation rate by using the relative importance of each asset as the weight. Lastly, we compute the average annual depreciation rate for these two subsamples. The

Table 2.2: Model calibration and estimation

Part A. External Calibration

Parameter	Description	Traditional Economy 1980-1999	Superstar Economy 2000-2015	Source/Reference
ρ	rate of time preference	0.046		
γ	risk aversion	4.0		Wang, Wang and Yang (2012)
θ	EIS reciprocal	2.0		
λ	cash carry cost	0.01		Bolton, Chen and Wang (2011)
δ	capital depreciation rate	0.053	0.056	BEA-FAT
η	operating scale	0.48	0.64	
f_0	fixed production cost	0.11	0.32	Compustat
$\overline{\mu}$	capital quality: long-run mean	1.48		
σ	capital quality: volatility	0.76		

Part B. Internal Estimation

Parameter	Description	Traditional Economy 1980-1999	Superstar Economy 2000-2015	Difference
ϕ	taste for quality	0.43	0.56	+ 0.13
ξ_0	variable production cost	0.94	0.26	-0.68
κ_0	investment adjustment cost	1.20	1.30	+0.10
χ_0	fixed external financing cost	0.37	0.55	+0.18
χ_1	variable external financing cost	0.053	0.088	+0.035
β	tightness of borrowing constraint	0.22	0.29	+0.07

depreciation rate in the traditional economy is 0.053, and the number in the superstar economy is 0.056. Although the depreciation rate is substantially higher for intangible capital and its total stock is increasing over time, there is negligible difference in aggregate depreciation rate between these two subsamples because physical capital still accounts for the majority of total assets.

We use the Compustat dataset to calibrate those parameters on production technology and capital quality. The long-run mean and volatility of the underlying capital quality process are calibrated to match the mean and standard deviation of sales obtained from Compustat. For better interpretation, we rescale the average output to be one here. In order to estimate operating scale η and fixed production cost f_0, we follow De Ridder (2019)'s approach by using firm-level information on markup, revenue, and profit. More specifically, based on the entrepreneur's optimization problem, fixed costs and operating scale can be estimated with the following equations:

$$f_0 = \left(1 - \frac{1}{\mu}\right) py - \pi = \left(1 - \frac{1}{\text{markup}}\right) \text{SALE} - \text{IB} \tag{2.36}$$

$$\eta = \frac{\xi_0 y^{\frac{1}{\eta}}}{py} = \frac{\text{COGS} - f_0}{\text{SALE}} \tag{2.37}$$

Following the standard literature, we measure firm-level revenue by using total sales (Compustat data variable *SALE*). In addition, we obtain the total production cost data by using Compustat variable *COGS*, which contains all the direct costs involved with production. The operating profits π are measured with income before extraordinary items (Compustat data variable *IB*). As the firm-level price information is not available in Compustat, we cannot directly estimate variable production cost ξ_0. Therefore, we will estimate it with the structural approach. The estimated fixed costs (after output rescaling) are 0.11 and 0.32 for the traditional economy and new economy, respectively. Moreover, the estimated operating scale has also increased from 0.48 to 0.64 in the sample period. This upward trend in fixed costs and operating scale is consistent with other related work (e.g., De Ridder, 2019; Hoberg and Phillips, 2021).

One implicit assumption behind our calibration strategy is that the primitive stochastic capital quality process does not change when the society transitions from the traditional to the superstar economy, i.e., $\bar{\mu}$ and σ are the same for these two subsamples. This condition implicitly assumes that firms become superstars not because they become super-productive but because they benefit from the permanent changes in supply and demand curves. This assumption is consistent with some empirical results from the existing literature. For example, Gutiérrez and Philippon (2019) find that superstar firms have not become more productive despite the increasing market valuation. Gabaix and Landier (2008) find that the recent rise in CEO compensation is an efficient equilibrium response to the increase in firms' market value, rather than resulting from increasing agency issues and managerial skills.

Internal estimation: SMM-MCMC approach

The rest of the parameters are jointly calibrated with the SMM approach by targeting some moments in the data. We choose the Markov Chain Monte Carlo (MCMC) method because it can generate faster convergence by bouncing between parameter and state vector draws. However, the cost of using this MCMC algorithm is a more restricted assumption on the distributions of observation errors.

Six parameters are calibrated through this SMM-MCMC approach: quality taste ϕ, marginal production cost ξ_0, investment adjustment cost κ_0, fixed external financing cost χ_0, variable external financing cost χ_1, and tightness of borrowing constraint β. The data moments we choose are mean and dispersion of markup, investment-to-output ratio, and cash-to-output ratio. In addition, we include the relative markup ratio of the 90th percentile to the median value in these two subsamples. It turns out that quality taste parameter ϕ is sensitive to this data moment, so including it can help identify it more accurately.

Table 2.3: Target moments and parameters uncertainty

Part A. Goodness of Fit

Moments	Traditional Economy: 1980-1999		Superstar Economy: 2000-2015	
	Data	Model	Data	Model
Average markup	1.42	1.31	1.54	1.48
Dispersion of markup	0.55	0.58	0.65	0.62
Markup distribution: P90/P50	1.60	1.42	1.70	1.72
Aggregate investment/output ratio	0.083	0.077	0.060	0.064
Dispersion of investment/output ratio	0.21	0.30	0.11	0.10
Aggregate cash/output ratio	0.06	0.08	0.13	0.14
Dispersion of cash/output ratio	0.44	0.57	0.55	0.61

Part B. Parameters Uncertainty

Parameter	Traditional Economy: 1980-1999				Superstar Economy: 2000-2015			
	p10	p90	S.D.	Median	p10	p90	S.D.	Median
ϕ (taste for quality)	0.186	0.672	0.187	0.432	0.346	0.759	0.161	0.556
κ_0 (investment adjustment cost)	0.466	1.433	0.376	0.941	0.183	0.338	0.0597	0.260
ξ_0 (variable production cost)	0.565	1.844	0.492	1.204	0.789	1.829	0.411	1.296
χ_0 (fixed external financing cost)	0.161	0.591	0.167	0.372	0.249	0.833	0.227	0.555
χ_1 (variable external financing cost)	0.0320	0.0744	0.0167	0.0528	0.0559	0.122	0.0258	0.0880
β (tightness of borrowing constraint)	0.149	0.286	0.0535	0.218	0.188	0.390	0.0796	0.288

In the top panel in Table 2.3, we provide the calibration targets and the model response. As shown in this panel, the model fits the data moments reasonably well. The bottom panel in Table 2.2 presents the median values of these structurally estimated parameters. We choose median instead of mean so that the result suffers less from some extreme outcomes. In addition, we can take full advantage of this MCMC approach to build the corresponding confidence intervals and posterior distributions for each of these estimates. In Figure A14 and A15 in the appendix, we plot the posterior distributions, which provide the histograms of the 5000 parameter draws of the estimated model. The corresponding summary statistics for each parameter are provided in the bottom panel in Table 2.3. For instance, in the traditional economy, the median estimated value of

ϕ is 0.432, and its 10-90 percentile range is from 0.186 to 0.672. In contrast, with the subsample dataset on the superstar economy, ϕ is estimated to have a median value of 0.556, and its 10-90 percentile range is from 0.346 to 0.759. Therefore, when taking the previous model to the data, it shows that people indeed have stronger preferences for product quality in today's economy. At the same time, the estimated marginal product cost has declined from 0.941 to 0.260. One possible reason behind is certain scale-related technical changes such as digitization, which allows individual firms to easily serve a large group of buyers with nearly zero marginal cost. The estimated 10-90 percentile range for this marginal cost parameter ξ_0 is from 0.466–1.433 and 0.183–0.338 for two subsamples, respectively.

2.3.2 Cross-Sectional Validation

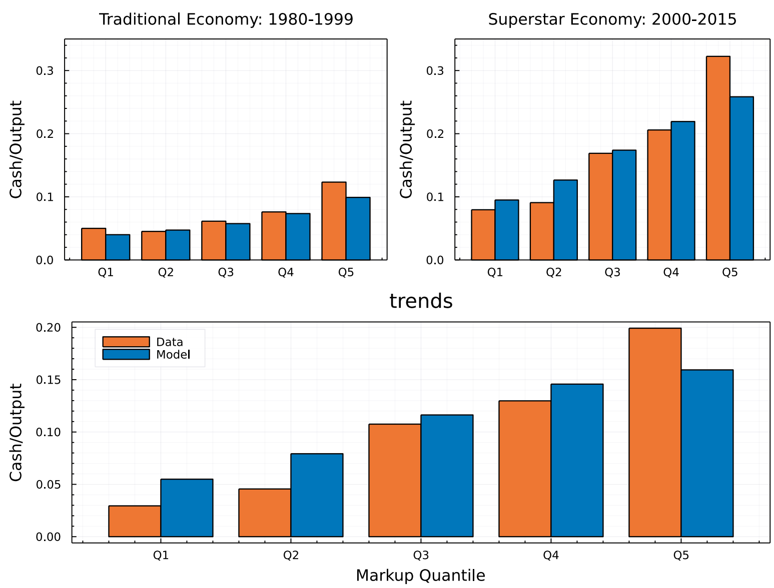
In this section, we provide some cross-sectional evidence to validate the model and the choice of parameters. We focus on the cross-sectional results on cash-to-output ratios and investment-to-output ratios. It is worth noting that when parameterizing the model, we do not directly target these moments. Therefore, the goodness of fit on these moments can be informative for evaluating the model's validity.

We first look at how the model fits the cross-sectional cash-to-output ratios. The results are presented in panel A in Figure 2.6, where we include both the data and model for better comparison. The data part is created as follows. To begin with, for each year between 1980 and 1999, we split all the firms into five different groups according to their markup level. Throughout this section, Q1 represents firms with the smallest markup, and Q5 means the group of firms with the highest values of markup. After that, we compute the firm-level cash-to-output ratios, then take averages for each group in each year. Finally, we compute the subsample average for each group throughout the subsample period. At the same time, the model part is generated by simulating a panel of 5,000 firms. The firms will behave optimally according to the first-order conditions derived in the previous section. Then we classify all the firms into five different groups according to their markup level, and compute the model-implied average cash-to-output ratios for each group. Finally, the bottom picture in Figure 2.6 is obtained by computing the difference in these numbers between the traditional and the new economy.

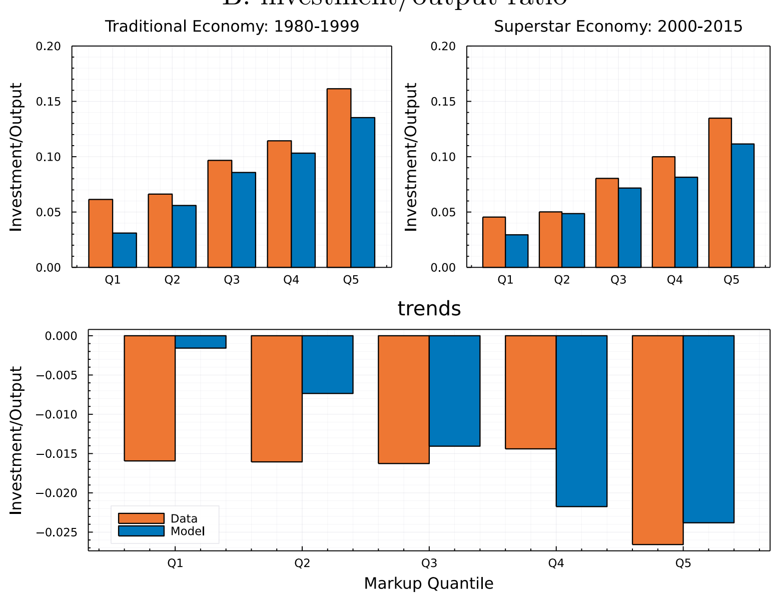
Figure 2.6 shows that our model is able to quantitatively match the cross-sectional patterns of cash-to-output ratios in both subsamples. Our model implies that all the firms in different groups tend to save more in the Superstar Economy, especially for

Figure 2.6: Cross-sectional validation: data v.s. model

A. cash/output ratio



B. investment/output ratio



the firms with the highest markups. This cross-sectional pattern is consistent with what we find in the data. As previously explained, the underlying mechanism for this outcome is that the shifts in demand and supply curves raise both the expected profits and uncertainty in the future; therefore, this risk redistribution channel makes all firms save more internally, especially for the superstar firms. In terms of magnitudes, what is less satisfying here is that the model tends to over-estimate cash-to-output ratio for the firms with the smallest markup, and under-estimate it for the firms with the highest markup. One possible reason is that we assume all firms face the same borrowing constraint. In reality, although, this assumption may not hold. In addition, the reason why the superstar firms in our model do not accumulate more cash is because they have the option of lending. However, in reality, most firms need to overcome considerable information costs for lending in the financial market, which gives them more incentives to hold safe assets instead.

We can also investigate whether the model can quantitatively match the investment-to-output ratio cross-sectionally. The results of comparing the model to the data are presented in panel B of Figure 2.6, which is generated in the same way as we did in panel A. The model can quantitatively fit the cross-sectional patterns of investment-to-output ratios in both subsamples, especially for firms with medium and large markup levels. When transitioning from the traditional economy to the Superstar Economy, the investment-to-output ratio decreases for the firms with highest markups in both the data and the model. In the data, the change in investment-to-output ratio is -0.026, and the corresponding result generated from the model is -0.025. As discussed before in the model section, superstar firms do not have incentives to invest due to their increased risk.

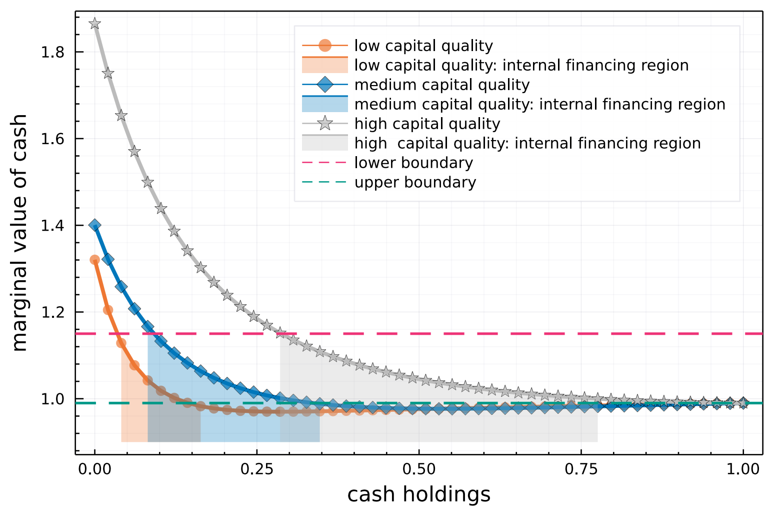
This pattern of the relationship between investment and output is similar to the empirical findings in Kilic, Yang and Zhang (2019), which document a negative cross-sectional correlation between the firm's investment and profitability. However, their explanation is different from this paper: they argue that firms with higher cash flow duration have lower discount rates, leading them to invest more despite having lower current profitability. In addition, our result here also speaks to the secular stagnation literature. Summers (2013) and many other works point out that one puzzling phenomenon in today's economy is that firms have strong incentives to save but no incentives to invest. The income and risk redistribution story in this paper can help understand this puzzle from a new perspective.

2.3.3 Policy Functions

Now we turn to discussing the entrepreneur's optimal investment and cash holding decisions.

Figure 2.7: Policy functions

A. cash holding



B. investment

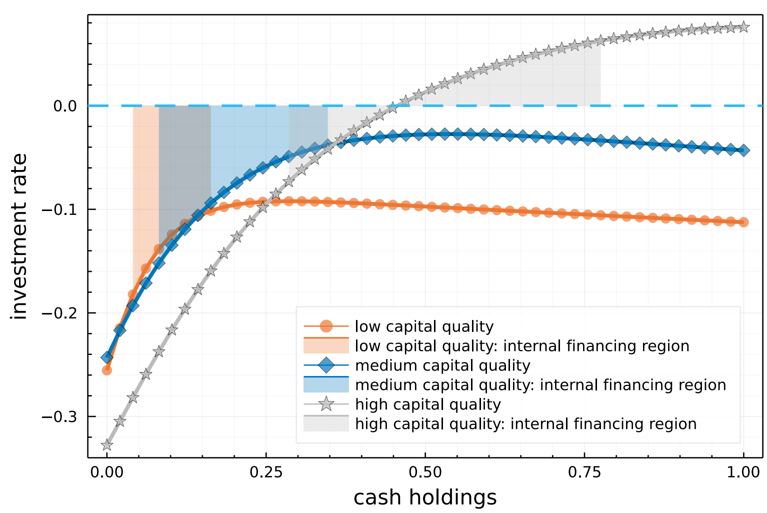


Figure 2.7 presents the numerical solutions with estimated parameters from the Superstar Economy subsample. In panel A, we plot the marginal value of cash with

respect to its level for firms with different capital qualities. Three conclusions are worth noting from this graph. First, the marginal value of cash is always decreasing as cash becomes more abundant, and this conclusion applies regardless of the capital quality level. The underlying reason is that firms with less cash are more likely to borrow from the financial market and hence need to pay more external financing costs. As a result, the marginal value of cash is higher for cash-scarce firms. In addition, this decreasing characteristic is important here as it ensures the concavity of firm value within this internal financing region, which guarantees the uniqueness of optimal policy. Second, although the marginal values of cash on the lower and upper boundaries are the same for entrepreneurs with different capital quality, the corresponding internal financing regions are heterogeneous. The same marginal value on boundaries come from the price mechanism at work. For entrepreneurs on the boundary, they are indifferent between internal and external financing. Given the fact that all the entrepreneurs need to have the same marginal value of cash when they seek external financing, this rule also applies to those on the boundary. Hence, heterogeneous entrepreneurs share the same marginal value of cash on the boundaries. Still, the resulting internal financing regions are different because the relationship between marginal cash value and cash level is different for firms with different capital quality. This result shows up in our model because the level and curvature of this relationship depend on the expected level and uncertainty of future earnings. In a Superstar Economy, the earnings process is quality-based and non-homogeneous. Therefore, firms with different quality levels will have different internal financing regions. Third, the range of internal financing region is increasing in capital quality. As we showed previously, compared to firms with low quality, a Superstar Economy faces more earnings uncertainty and hence places more value on cash. As a consequence, they will have a wider internal financing region. In this way, the dispersion of marginal value of financial wealth will be higher among these superstar firms, which is the origin of misallocation in this paper. This mechanism is quite different from the existing literature: in most studies, in fact, the importance of cash comes from financing frictions as cash keeps the firm away from costly liquidation and external financing. In this paper, by contrast, the superstar firms want to accumulate more cash not because they are more likely to be financially constrained, but because they face a more volatile earnings process in the future.

Panel B of Figure 2.7 plots the optimal investment-to-capital ratios for firms with different levels of capital quality. Two conclusions can be drawn from this graph. First, optimal investment is lower for firms with less cash on hand. Second, the sensitivity of investment to cash level is stronger for firms with relatively higher markup. The first conclusion is quite intuitive because when cash level is sufficiently low, firms need to cut more investment so that they can prevent costly bankruptcy and external financing costs. At the same time, this disinvestment behavior incurs some capital adjustment costs. Therefore, the optimal investment rate is pinned down by the trade-off between cash carry cost and disinvestment cost. The second conclusion is closely related to the higher cash flows uncertainty faced by high-quality entrepreneurs. The incentives of underinvestment are stronger for these firms because they have more desire to hold cash.

2.3.4 Quantitative Performance

Here we inspect the aggregate implications with M.I.T. shocks. By definition, an "M.I.T. shock" is an unexpected shock that hits an economy at its steady state, leading to a transition towards a new one. To better evaluate the quantitative performance, we choose three facts related to the declining capital allocation efficiency in the data. Then we use the general equilibrium model developed in Section 2.1 and the previous parameterization strategy to quantify its underlying mechanism and investigate whether the model is able to jointly explain these aggregate trends.

The three facts documented in this section are: increasing dispersion of firm-level marginal revenue return to capital; negative correlation between firm-level TFP and net finance dependence; and increasing gap between aggregate marginal product of capital and real interest rate. We interpret them as signs of declining capital allocation efficiency, because an efficient financial market should generate zero dispersion of firm-level investment return, more resources allocated towards productive users, and marginal return of additional investment being equal to its marginal cost.

Self-financing and "misallocation"

Fact The first fact is that the measured static dispersion of firm-level marginal product of capital, defined as misallocation, has been increasing steadily since the 1980s. Following David, Schmid and Zeke (2019), we measure it by calculating the standard deviation of log marginal return to capital (mpk), and mpk is computed as the log difference between the firm's revenue and capital stock. For the baseline result, we use the firm's reported sales (Compustat series $SALE$) as the proxy for firm-level revenue and the total net value of property, plant, and equipment (Compustat series *PPENT*) as the proxy for the firm's (physical) capital stock. As shown in Figures A1 and A2 in the appendix, using alternative measures (e.g., including intangible capital) yields similar patterns.[15]

The solid orange line in panel A of Figure 2.8 presents the time-series plot of the baseline measure on misallocation. For this full sample, each year, we include all the firms available in the U.S. Compustat dataset except for those in the financial and utility sectors. As the graph shows, there is a significant increase in the degree of capital misallocation among all the public firms. In the full-sample dataset, the degree of misallocation has increased by 30.7% over the last forty years. This result is surprising as these public firms should have improved access to the financial market over time; therefore, if the measured misallocation is indeed increasing in an advanced economy like the U.S., it is necessary to seek for explanations beyond financial frictions.

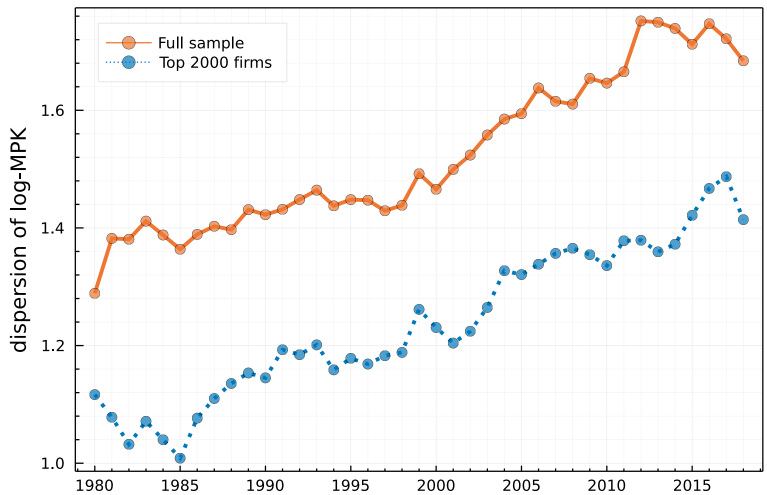
One possible concern for our baseline result is that we measure misallocation with an unbalanced panel. The frequent entry and exit of different firms could contribute to measurement errors. One way to alleviate such concern is to fix the number of firms over the sample period. The blue dotted line is obtained with a similar approach but only for a subset of large firms. For each year, we only include the top 2000 firms according to their market capitalization. Then for this subset of firms, we calculate the degree of capital misallocation using the same approach as before. Again, Figure 2.8 tells us that the U.S. economy has experienced a secular increase of 26.6% in the measured degree of misallocation, even after fixing the number of firms in the sample.

In addition, in panel B of Figure 2.8, we create seven small but balanced panels for each of the seven decades between 1950 and 2020. Specifically, for each ten-year time window, e.g., 1950-1959 or 1960-1969, we create a balanced panel where we only include the firms that are continuously present within this decade. For each year, we again calculate the *mpk* dispersion. As this graph reveals, before the 1980s, the degree of misallocation for a balanced panel of firms declines over time, possibly resulting from an improving financial market or from maturing business operations. For instance, in

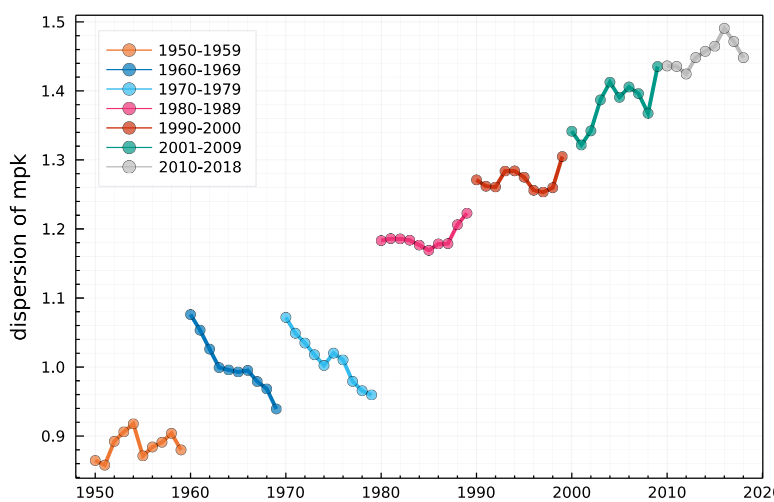
[15] One important caveat is that marginal product is proportional to average product only under the assumption of a Cobb-Douglas production function. Bils, Klenow and Ruane (2021) provide a way to correct mismeasurement by imposing two assumptions. The first assumption is that firm-level markup does not change with productivity shocks. The second assumption is that measurement errors with respect to revenues and inputs are orthogonal to the true marginal product. However, the reason why we do not follow Bils, Klenow and Ruane (2021) to measure misallocation is because the first assumption in their paper does not hold in this paper's story.

Figure 2.8: Increasing "misallocation"

A. unbalanced panels



B. seven small balanced panels



Notes: This figure presents different measures on the time-series of capital misallocation degree among U.S. public firms. Misallocation is defined as the dispersion of firm-level log marginal product of capital *mpk*. The orange solid line in the top graph represents the full-sample result, where we include all the firms available in that year except for those in the financial and utility sectors. The blue dotted line in the same graph shows the measured misallocation among top 2000 firms in that year according to their market capitalization. In the bottom graph, we present measured misallocation degrees for seven small and balanced panels for each of the seven decades between 1950 - 2020. For each balanced panel, we include firms continuously present within this decade. Firm-level *mpk* is measured with David, Schmid and Zeke (2019)'s method using Compustat North American data file. *mpk* is calculated as the log difference between firm's reported sales (Compustat series $SALE$) and the total net value of property, plant, and equipment (Compustat series $PPENT$).

the 1970-1980 balanced subsample, the degree of misallocation in the U.S. economy shows a decline of 10.5%; however, this pattern changes substantially after the 1980s. For the 1980-1989 and 1990-1999 panels, the measured misallocation is relatively stable over time and increases only slightly. In contrast, after 2000, even for a balanced panel of firms, *mpk* dispersion has increased substantially: the misallocation degree has increased by 6.5% within a short ten-year window. In summary, if we interpret the dispersion of firm-level *mpk* as a sign of misallocation, the capital has, since the 1980s, been increasingly misallocated.

Quantitative performance Table 2.4 summarizes the results of comparing these macro-finance trends in the data and those obtained from the model. In the data, misallocation is measured the same way we did for Figure 2.8, which is the dispersion of *mpk*. In the model, misallocation is obtained by calculating the steady-state dispersion of log $\mathcal{J}_{\zeta}$. According to Table 2.4, our model is able to quantitatively fit the data in a reasonably good way. In the data, the degree of misallocation has increased by 0.22, from 1.41 to 1.63. In contrast, the degree of misallocation implied by the model has increased by 0.31, from 1.18 to 1.49. As explained previously, misallocation goes up in the model is due to the expansion of the internal financing region. From a macro perspective, when the internal financing region expands, it means that an individual-entrepreneur-led allocation system is replacing a price-mechanism-governed one, which deteriorates the capital allocation efficiency on the aggregate level.

Table 2.4: Quantitative results

Macro-finance trends	Traditional Economy: 1980-1999		Superstar Economy: 2000-2015		Trends	
	Data	Model	Data	Model	Data	Model
degree of "misallocation"	1.41	1.18	1.63	1.49	+0.22	+0.31
correlation between TFP and net finance	0.036	0.018	-0.128	-0.113	-0.164	-0.131
MPK - r	5.27%	3.93%	10.27%	9.08%	+5.00%	+5.15%
area disciplined by the price mechanism	-	83.40%	-	72.52%	-	-10.88%

The underlying mechanism in this paper still belongs to a finance perspective on misallocation. However, the story here is different from the standard explanations in the existing literature. For instance, Midrigan and Xu (2014) investigate the role of financial frictions on misallocation from both an extensive and intensive margin perspective. In their paper, the reason why finance frictions generate misallocation is twofold. First, it distorts entry and technology adoption decisions. Second, the existence of borrowing constraints may generate differences in the returns to capital across individual producers. In addition, Buera, Kaboski and Shin (2011) show that financial frictions account for a substantial part of the observed cross-country differences in aggregate TFP. Similarly, Gopinath et al. (2017) study the role of the size-dependent borrowing constraints, combined with the decline in the real interest rate, might generate the increasing dispersion of MPK. Generally speaking, these papers are focused on the inefficiency *within* the market system. In contrast, this paper focuses on the role of a moving firm-market boundary on misallocation. Our story here is also different from David, Schmid and Zeke (2019), which attempt to explain misallocation from an asset pricing perspective. David, Schmid and Zeke (2019) argue that as firms differ in their exposures to systematic risks, the dispersion in MPK could come from the heterogeneity in firm-level risk premia. In contrast, this paper attempts to explain the dynamics of misallocation from a corporate finance perspective. Changes in corporate risk management policies could lead to the changing dynamics of misallocation.

Our explanation here is similar to the risk and inaction story in Bloom (2009) and Ackerberg, Caves and Frazer (2015). A standard result from this type of model is the existence of an inaction region. However, Bloom (2009) and Ackerberg, Caves and Frazer (2015) focus on the fixed costs in labor hiring and capital investment. Therefore, in their works, we will see inaction regions in employment and investment while here we have fixed costs in external financing. Hence the inaction in this paper means not using the external financial market. The shifts in supply and demand curves will affect the size of this inaction region through their impacts on the earnings process. As a result, they generate some aggregate impacts on misallocation.

The key implication from our exercise here is that zero misallocation should not be the optimal policy target if we allow for the existence of firm-market boundary. Since Hsieh and Klenow (2009), any dispersion of factor return across firms is usually considered as a barrier to the efficient allocation of resources. Therefore, one should expect zero misallocation if we manage to eliminate all the distortions within the market system. The existence of corporate internal financing, however, creates a firm-market boundary on the financing side so that increasing misallocation defined from Hsieh and Klenow (2009)'s perspective could come from either the increasing inefficiency within the market system or the shrinking boundary of the market system. Unless the government could take all the internal resources from firms, zero misallocation should not be the policy target.

Eclipse of the public markets

Fact Since the insightful observation by Jensen (1989), there has been a growing body of literature discussing whether the role of capital markets has changed over time. The second fact is closely related to a simple question: does the market always allocate capital to the most productive users, and has this role changed over time?

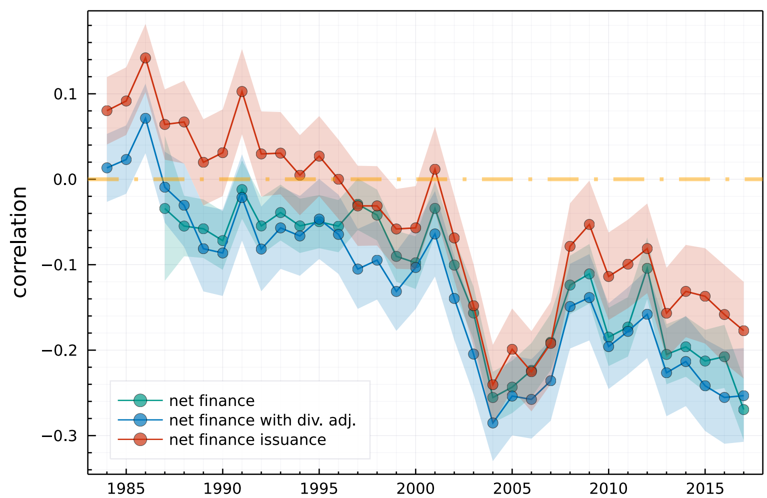
The answer to this question can be found in Figure 2.9, where we calculate the annual cross-sectional correlation between the firm's net finance dependence and its productivity. In the baseline result in panel A, we follow Imrohoroglu and Tuzel (2014) to measure the firm-level total factor productivity (TFP) and Frank and Yang (2019) to obtain three different measures on the firm's net finance dependence. In market economies, a critical role of the financial market is to allocate resources to their most efficient uses. While this correlation should be positive, it does not prove true according to the data. As shown in Figure 2.9, this correlation has changed from positive to negative over the past several decades. For instance, at the beginning of the 1980s, the estimated correlation between firm-level TFP and net finance issuance is 0.1, and it is significant at the 99% confidence level. However, after the late 1990s, such a correlation has become negatively significant for most of the time, which suggests that in today's economy, on average, the total debt and equity do not flow to the most productive firms. As a matter of fact, less productive firms actually rely more on external financing. This conclusion does not depend on how we measure firm-level productivity. In panel B, we estimate the firm-level TFP with alternative approaches, including the Olley-Pakes (Olley and Pakes, 1996), Levinsohn-Petrin (Levinsohn and Petrin, 2003), Wooldridge (Wooldridge, 2009), and Ackerberg-Caves-Frazer (Ackerberg, Caves and Frazer, 2015) methods. This graph shows that using different measures of net finance dependence and firm-level TFP will generate slightly different patterns, but the key message, i.e., productive firms becoming less reliant on external financial markets, still holds in the data.[16]

The pattern documented in Figure 2.9 is consistent with the hypothesis that there seems to be an eclipse of the public markets in the U.S. (Jensen, 1989; Doidge et al., 2018). Our findings here also complement some previous studies in the literature. For instance, using an industry-level U.S. dataset, Lee, Shin and Stulz (2020) find that after

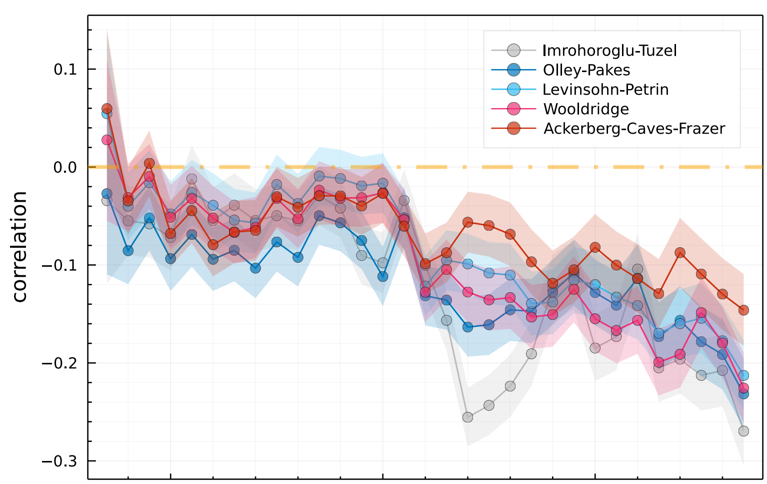
[16] In the baseline analysis, we use Spearman rank correlation, but using Pearson correlation generates similar results. One advantage of using the rank correlation is that it is much less sensitive to potential outliers.

Figure 2.9: Eclipse of the public markets

A. different measures of net finance



B. different measures of firm-level TFP



Notes: This figure presents the cross-sectional correlation between firm-level TFP and firm-level net finance dependence. We use five different approaches to measure firm-level TFP, including Imrohoroglu-Tuzel (Imrohoroglu and Tuzel, 2014), Olley-Pakes (Olley and Pakes, 1996), Levinsohn-Petrin (Levinsohn and Petrin, 2003), Wooldridge (Wooldridge, 2009), and Ackerberg-Caves-Frazer (Ackerberg, Caves and Frazer, 2015) methods. In addition, we follow Frank and Yang (2019) to obtain three different measures on the firm's net finance dependence: total net finance; net finance with dividend adjustments; and net finance issuance. Main data source for this figure is from Compustat North American Annual data file.

2000, industries with low Tobin's q receive more funding from capital markets than those with high Tobin's q. Using a machine learning method for estimating firm-level productivity, Frank and Yang (2019) find that finance typically flows away from high productivity firms as more productive firms tend to lend rather than borrow from capital markets. Doidge, Karolyi and Stulz (2017) find that the number of public firms in the U.S. has fallen substantially in the last several decades, and many firms have started to repurchase their equity. In addition, Bils, Klenow and Ruane (2021) find that there is a modest downward trend in the U.S. allocative efficiency, even after correcting for measurement error.

Quantitative performance The second row in Table 2.4 presents the data-model comparison in the estimated correlation between firm-level external finance dependence and TFP. The correlation in the data is measured the same way we did for Figure 2.9 with the net finance issuance and Imrohoroglu and Tuzel (2014)'s TFP measurement. In the model, such correlation is obtained by calculating the cross-sectional correlation between external finance dependence $\frac{b}{y}$ and firm-level capital quality ζ. According to Table 2.4, our model can quantitatively match the data fairly well. In the data, the estimated correlation has decreased by 0.164, from 0.036 to -0.128. In the model, such correlation has decreased by 0.131, from 0.018 to -0.113.[17] The reason why productive firms rely less on external financing in a Superstar Economy is that these superstar firms need to face more volatile earnings in the future. The increased future cash-flow uncertainty prompts these productive firms to save more internally.

The underlying mechanism here is similar to Jensen (1989)'s hypothesis and more recently Doidge et al. (2018)'s work. Jensen (1989) observes the decline in the number of public firms in the U.S., and he argues that agency problems between shareholders and managers can make public corporations a less inefficient form of organization. Doidge et al. (2018) extend Jensen (1989)'s hypothesis by distinguishing traditional firms with tangible capital and young firms with intangible capital. They argue that intangible capital has more financing issues in the public market as intangible capital has limited collateral value and more information asymmetry issues, so its value depends more on the market environment. Therefore, they argue that the recent decline in the number of listing firms is not a short-term phenomenon. Instead, it indicates that intangible-capital-intensive firms might be better suited for financing through private sources than

[17] The data on net finance issuance in Compustat is not available until 1984. Therefore, the initial year for comparing the model to the data is chosen to be 1984.

through public capital markets. Our story in this paper echoes their hypothesis, and provides some quantitative evidence for their claim.

More importantly, the result here points out that the economic environment in Hsieh and Klenow (2009) relies on one crucial assumption: all firms are exogenously assumed to borrow from the market. Only in this way, a frictionless capital market can improve allocation efficiency by moving resources from firms with low marginal product to firms with high marginal product. The effectivenss of the market system will be weakened if firms endogenously choose internal financing.

Finally, the result here provides a novel perspective to the recent discussion on the disappearing public firms. Although the focus here is on the optimal reliance on external financing instead of firms' decisions on being public or private, the underlying mechanism in this paper could help us understand why there is a significant decline in the number of publicly-listed companies in the U.S.. In the existing literature, possible explanations include regulatory burdens (e.g., Dambra, Field and Gustafson, 2015; Ewens, Xiao and Xu, 2021), intangible capital (e.g., Kahle and Stulz, 2017), private equity (e.g., Ewens and Farre-Mensa, 2020), economies of scale (e.g., Gao, Ritter and Zhu, 2013), and mergers and acquisitions (e.g., Eckbo and Lithell, 2021). In contrast, here we argue that it could come from some economic fundamental changes that permanently change the risk of corporate earnings.

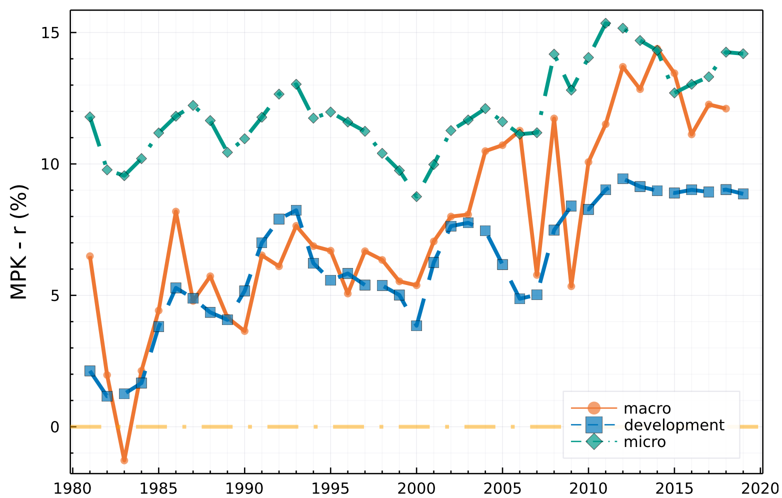
Two rates of return

Fact The third fact is a well-known fact in the macroeconomics literature. Many recent studies (e.g., Farhi and Gourio, 2018) find that there is a puzzling phenomenon about the U.S. economy: aggregate returns to business capital (MPK) have been either stable or growing, but the real interest rate (r) has been declining. As a result, the gap between MPK and r has been stably increasing over time. Figure 2.10 summarizes this well-known fact with three different measures on aggregate MPK. The macro approach measures the aggregate MPK by following Gomme, Ravikumar and Rupert (2011) and using data from the National Income and Product Accounts (NIPA).[18] The development approach adopts Feenstra, Inklaar and Timmer (2015)'s method and the Penn World

[18] Caballero, Farhi and Gourinchas (2017) have adjusted Gomme, Ravikumar and Rupert (2011)'s estimates of the capital stock for intangible intellectual property products (IPP). The main conclusion still holds with this capital stock adjustment.

Table dataset.[19] The micro approach follows Grullon, Larkin and Michaely (2019) and measures the aggregated return of assets with the Compustat dataset. Despite the difference in the absolute magnitudes, a secular upward trend in the difference between MPK and r in the U.S. is obvious, across all of these approaches.

Figure 2.10: MPK minus r



Notes: This figure presents the time-series difference between aggregate marginal product of capital (MPK) and real interest rate in the U.S.. MPK is measured with three different methods. The macro approach measures the aggregate MPK by following Gomme, Ravikumar and Rupert (2011) and using data from the National Income and Product Accounts (NIPA). The development approach is to measure the aggregate MPK with Feenstra, Inklaar and Timmer (2015)'s approach and Penn World Table dataset. The micro approach follows Grullon, Larkin and Michaely (2019) and measures the aggregated return of assets with the Compustat dataset. One-year real interest rate is measured by subtracting inflation expectations from nominal Treasury yields, and inflation expectations are measured as median consumer price inflation expectations from the Philadelphia Fed survey of professional forecasters.

The reason we interpret this gap as a piece of evidence for capital allocation inefficiency is the following. In a standard macroeconomics model, MPK is seen as the return of additional investment, and the real interest rate is interpreted as the cost of additional investment. Any wedges between marginal return and marginal cost should be considered as a sign of market inefficiency because the amount of capital should be

[19]Feenstra, Inklaar and Timmer (2015) (and Penn World Table 9.1) provide a new cross-country comparable measure on the real internal rate of capital return. However, the result does not change if we follow Caselli and Feyrer (2007)'s original approach of measuring aggregate marginal product of capital.

allocated such that the marginal cost equals the marginal benefit. In other words, this increasing wedge suggests an imperfectly competitive financial market.

Of course, the actual cost of investment should also include capital's depreciation rate (δ) and risk premium (RP). Therefore, the total cost of capital in this case should be $r + \delta + \text{RP}$. we plot the time series of this adjusted return-cost gap in Figure A8 in the appendix, and it is clear that the upward trends are still there for all the three different measures. In addition, we can still observe that the gap is positive or changes from negative to positive for two out of these three measures.

Quantitative performance The third row in Table 2.4 presents the estimated trend in the MPK & r gap both in the data and in the model. The data part is obtained the same way as in Figure 2.10. In the model, MPK is obtained by calculating the average value of $\mathcal{J}_\zeta$, while r is simply the equilibrium interest rate. Table 2.4 shows that in the data, the measured MPK & r gap has increased by 5.00%, from 5.27% to 10.27%. Meanwhile, our model suggests that this gap has increased by 5.15%, from 3.93% to 9.08% over the past forty years. The model result is consistent with what can be observed in the data. The underlying mechanism for this upward trend in the model is related to the distinction between internal cash value and external debt value. For firms who finance themselves externally, their marginal product of capital is closely related to the external interest rate. However, for firms financing internally, their marginal product of capital is more connected to the internal cash value rather than to the external interest rate. The expansion of the internal financing region indicates that cash becomes more valuable to firms, which generates the increment in the MPK & r gap. Although there is a growing body of literature on the divergence between return on productive capital and the interest rates, the explanation here is distinct. For instance, Farhi and Gourio (2018) point out that the rising market power, risk premia, and intangible capital are all important for our understanding of some recent macro trends, including this MPK & r gap. Karabarbounis and Neiman (2018) argue that the gap between measured capital income and estimates of the required compensation of capital is most likely explained by measurement error in the cost of capital. Crouzet and Eberly (2020) provide a $Q+$ framework to link together the rents and intangible capital. In this way, three elements can contribute to the difference between average q and marginal q for physical capital: rents to physical capital, the value of installed intangible capital, and the rents to intangible capital. Therefore, Crouzet and Eberly

(2020) argue that increasing rents and intangible capital contribute to the rising gap between MPK and r.

The result in Table 2.4 also speaks to the growing inequality literature, and in particular to Piketty (2013)'s $r-g$ framework. Piketty (2013) argues that the relationship between the return to capital r and the economic growth rate g is particularly important for understanding the economic dynamics. What is new in our paper is that there are actually two types of Rs: one represents the marginal return of entrepreneurs who still rely on external financing, and the other is the marginal return of those who do not. If all the firms are financed externally, we should expect these two Rs to be equal. However, if there are some transaction costs of using the external market, then the gap between these two Rs will emerge. The conclusion in this paper can potentially solve two issues in the literature. First, Piketty (2013) argues that the rate of return does not fall sufficiently fast with capital deepening. However, many empirical papers have documented the secular decline in the real interest rate. In contrast, this paper here points out that the return rate should be measured with MPK as many productive firms do not finance their investment through the external financial market, which makes the real interest rate less informative on the rate of return. Second, Mankiw (2015) and many other scholars argue that the condition $r > g$ is not surprising under the neoclassical framework. Therefore, there is no apparent reason why we should be concerned about the rising inequality in wealth. However, our paper contends that some economic fundamental changes such as technical innovation can generate both income and risk redistribution. This risk redistribution is likely to shrink the boundary of the price mechanism and decrease capital allocation efficiency. Therefore, we should pay serious attention to the increasing inequality and its related consequences.

Moving firm-market boundary

Another interesting result from our quantitative exercise is that it enables us to observe the secular changes in the area disciplined by the price mechanism. Empirically, the boundary of the invisible hand is difficult to measure as it is invisible by nature. However, after building up a general equilibrium model, we can directly observe how the firm-market boundary moves over time. We present the result in the last part of Table 2.4, which is obtained by calculating the difference in the steady-state values of Ψ between these two subsamples. Ψ is measured as the wealth-weighted share of firms using external financial market, and also represents the effectiveness of the market system. As is shown in Table 2.4, the model implies that the area controlled by the price mechanism has declined by about 10.88% over the past forty years. The value of Ψ in the traditional economy subsample was 0.834. In contrast, in today's economy, that number has decreased to a value of 0.725. Therefore, with the help of the model introduced in the previous section, we can directly observe how the area controlled by the price mechanism shrinks over time.

In addition, the number estimated in this paper is also quantitatively consistent with Bils, Klenow and Ruane (2021)'s finding that there is a modest downward trend in the aggregate allocation efficiency. After correcting the possible measurement errors, they find that capital allocation efficiency has declined by 15% over the past 35 years in the U.S.. They argue that the increasing misallocation could come from specific government policies or capital/labor market frictions. In contrast, this paper rationalizes this trend from a new endogenous firm-market boundary perspective.

2.3.5 Decomposition Exercises

In order to show the relative contributions of different parameters, we conduct some counterfactual exercises. The corresponding quantitative results are provided in Table 2.5. Three conclusions are worth noting. First, the most important driver behind these trends is the change in production technology. In the counterfactual exercises, if the production technology were unchanged across the past several decades, then the dispersion of *mpk* would have only increased by 0.08, the correlation between firm-level TFP and net finance would have actually increased by 0.023, and the MPK & r gap would have only increased by 1.04%. At the same time, within these production technology parameters, the most important factor is the marginal supply cost ξ_0. When we fix ξ_0 to be constant, the model can only explain approximately 50% of these trends. The quantitative exercise here suggests that the emergence of digitization and other technical changes really bring us a society of nearly zero marginal cost, which fundamentally changes the relationship between earnings and product quality and eventually affects the long-run trends in some important macro-finance indicators.

Second, the change in taste for quality parameter ϕ is the second-most important driver behind these aggregate trends in the data. The changes in consumers' preference contributes to roughly one third of these macro-finance trends. This result suggests that the increasing preference of consumers over high-quality products also contributes

Table 2.5: Decomposition exercises

Macro-finance trends	Data	Model						
		Fix ϕ	Fix η	Fix f_0	Fix ξ_0	Fix η, f_0, & ξ_0	Fix ϕ, η, f_0, & ξ_0	Fix β
degree of "misallocation"	+0.22	+0.18	+0.25	+0.26	+0.14	+0.10	+0.08	+0.28
(% of the full model)	-	(58.06%)	(80.65%)	(83.87%)	(45.16%)	(32.26%)	(25.81%)	(90.32%)
correlation between TFP and net finance	-0.164	-0.0713	-0.116	-0.113	-0.0868	-0.0404	+0.023	-0.129
(% of the full model)	-	(54.43%)	(88.55%)	(86.26%)	(66.26%)	(30.84%)	(-17.56%)	(98.47%)
MPK - r	+5.00%	+3.33%	+3.91%	+3.85%	+2.80%	+1.60%	+1.04%	+4.72%
(% of the full model)	-	(64.66%)	(75.92%)	(74.76%)	(54.37%)	(31.07%)	(20.19%)	(91.65%)
area disciplined by the price mechanism	N/A	-7.28%	-9.34%	-9.26%	-5.50%	-3.25%	-3.17%	-10.68%
(% of the full model)	-	(66.91%)	(85.85%)	(85.11%)	(50.55%)	(29.87%)	(29.14%)	(98.17%)

significantly to the rise of superstar firms and their cash holding behaviors. When fixing this demand parameter, our model can only explain 60% of these trends. Therefore, we argue that changes in both the demand and supply sides are important for the understanding of these medium-run trends.

Third, based on the quantitative exercise in Table 2.5, changes in the degree of financial frictions are not an important factor behind these macro-finance trends. When we fix the value of β to be constant, the model still can explain more than 90% of the trends. This result is not surprising as this paper focuses on investigating the impacts of superstar firms. It is unlikely that these superstar firms have become more financially constrained over the past several decades. One caveat is that we could potentially underestimate the relative importance of financial friction in the whole economy because we use a dataset containing public firms only.

2.4 Related Literature and Additional Discussions

2.4.1 A Brief Review of Literature

This paper relates to several strands of literature. First, it is closely related to the growing literature on superstar firms. Rosen (1981) is the first to bring our attention to the economics of superstars. He points out that technical change would allow the most talented entrepreneur to serve a larger group of people and dominate this economy. Many following works attempt to use this idea to explain why earnings distribution is more right-skewed compared to the underlying talent distribution (e.g., Gabaix and Landier, 2008; Tervio, 2008). Recent studies can be classified into three categories. In the first category, people are interested in investigating the origins of superstars. Possible explanations include asset prices (Gomez, 2019), low risk-free rate (Liu, Mian and Sufi, 2019; Kroen et al., 2021), random growth (Luttmer, 2012), and so on. The second group of works focuses on the macro-finance implications of superstar firms. For instance, Autor et al. (2020) and Kehrig and Vincent (2020) discuss how these superstar firms contribute to the decline of labor share; De Ridder (2019) shows that the increasing market power of large firms discourages innovation and leads to the decline in business dynamism. The third category focuses on how different are today's firms compared to their counterparts several decades ago. For example, Hoberg and Phillips (2021) find that U.S. firms have expanded their scope and scale of operations in the past several decades, and this scope expansion significantly increases corporate valuation.

Second, this paper also connects to finance and misallocation literature.[20] One common perspective in this literature is that misallocation could arise from financial frictions such as borrowing constraints (Buera, Kaboski and Shin, 2011; Buera and Shin, 2013; Midrigan and Xu, 2014; Moll, 2014; Gopinath et al., 2017). In other words, these works argue that inefficiency within the market system leads to misallocation. In contrast, this paper shows that the movement in the firm-market boundary is another important origin of capital misallocation. In this way, although this paper still belongs to the finance view of misallocation literature, the underlying mechanism here is more close to the risk and adjustment cost story as in Bloom (2009) and Asker, Collard-Wexler and De Loecker (2014), where the authors argue that uncertainty and adjustment costs in capital accumulation or labor hiring generate an inaction region. Similarly, any dispersion of the marginal product of factors within this region will not be equalized and thus create misallocation.

Third, our paper also relates to the capital structure and product market competition literature. The theoretical works on this topic can date back to Brander and Lewis (1986) and Maksimovic (1988), which point out the role of capital structure in committing to certain product market strategy. In terms of empirical studies, MacKay and Phillips (2005) show that leverage is higher for industries with higher degrees of concentration. Gao (2021) finds that input-output production network also affects firm's optimal choice of internal financing. Recently, Jung, Kadyrzhanova and Subramanian (2021) provide both a theory and some empirical evidence to show that different types of competition can have contrasting implications on optimal leverage. In addition, Dou and Ji (Forthcoming) use a monopolistic competition framework with customer capital

[20] There is an extensive literature related to the topic of misallocation since Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). Please refer to Syverson (2011), Hopenhayn (2014), and Restuccia and Rogerson (2017) for detailed surveys.

to provide an interesting interpretation on the financial origin of markup.

Forth, this paper is also connected to the large literature on dynamic risk management. Examples include Gomes (2001); Hennessy, Levy and Whited (2007); Riddick and Whited (2009); Bolton, Chen and Wang (2011) and many others, where the authors use different dynamic models to jointly investigate the optimal investment, financing, and risk management decisions of firms with financial constraints. Most works in this branch of literature adopt a framework of one representative firm and partial equilibrium analysis. In this paper, in order to investigate the macroeconomic implications of corporate risk management, we use the heterogeneous agents and general equilibrium framework. In this way, we can obtain the endogenous firm-market boundary and investigate how changes in economic fundamentals will affect this boundary. What is interesting is that the endogenous *boundary* of the invisible hand is exactly the Neumann *boundary conditions* of certain HJB equations arising from the optimal decisions made by heterogeneous entrepreneurs. Therefore, from the individual perspective, corporate cash hoarding behavior is simply an optimal decision for firms to switch from external to internal financing. However, from the macro perspective, the expansion of the internal financing region means that the individual-firm-led credit allocation system is substituting the market mechanism, which will affect the allocation efficiency at the aggregate level.

Fifth, this paper is closely related to a voluminous literature on heterogeneous agent incomplete market model that goes back to Imrohoroglu (1989) and Aiyagari (1994). The focus of this branch of literature is to investigate the consumption and/or investment dynamics of heterogeneous agents when faced with uninsurable idiosyncratic shocks. For surveys on papers using discrete-time approach, please refer to Heathcote, Storesletten and Violante (2009), Guvenen (2020) and many others. Recently, there is an increasing number of papers using the continuous-time approach (e.g., Benhabib, Perla and Tonetti, 2019; Luttmer, 2007, 2011; Moll, 2014; Kaplan, Moll and Violante, 2018*a*). This paper also adopts the continuous-time approach mainly due to its computational advantage in solving both stationary equilibria and transition dynamics (Achdou et al., Forthcoming).

Lastly, this paper also speaks to a vast literature on transaction cost economics. Most studies after the pioneering work of Coase (1937) are focused on corporate governance and organizational structure. For a complete review of the recent development, please refer to Williamson (2010). Generally speaking, this paper differs significantly from this branch of literature in two ways. First, the firm-market boundary in the existing literature is on the production side. In contrast, this paper is focused on the financing side: firms can actively choose their risk management policies and determine how much they should rely on external finance, which eventually pins down the endogenous firm-financial market boundary. Second, this transaction cost literature is mainly focused on how institutional quality affects firm-market boundaries. In contrast, here we want to investigate how shifts in some demand and supply factors affect the effectiveness of the market system. One interesting conclusion from this paper is that although the technical change might be beneficial on the production side as it allows the most productive producers to serve more customers, it harms the price mechanism on the financing side as it makes companies save more internally and less disciplined by the market system.

2.4.2 Discussion on the Key Assumptions

Through verbal reasoning, Coase (1937) made two conjectures with the assumption of transaction costs: one is the existence of the firm-market boundary, and the other one is that firm is an allocation system different from the market. Similarly, when investigating the firm-market boundary on the financing side, in order to get these two outcomes, we explicitly assume that using external financial market incurs transaction costs. However, it turns out that this transaction cost assumption alone is not sufficient to get either of these two conclusions. In order to get the existence of firm-market boundary, one additional implicit assumption is **incomplete market**. In that sense, the agents in this economy cannot fully hedge their idiosyncratic risk or perfectly share their risks with other agents in this economy, thus giving them incentives to save internally. In other words, there is limited risk-sharing among heterogeneous agents through the financial market (either exogenously or endogenously). This assumption follows the long tradition of an extensive incomplete market literature (e.g., Aiyagari, 1994; Achdou et al., Forthcoming), where agents are subject to uninsurable idiosyncratic shocks by default. If market is complete, there will be no such boundary between the firm and financial market. The incomplete market assumption here is similar to the incomplete contract assumption used by the property right literature (e.g., Grossman and Hart, 1986; Hart and Moore, 1990, 1994) to derive the firm-market boundary on the production side. One crucial assumption used in these papers is that contracts cannot specify all the possible contingencies.

In order to obtain Coase (1937)'s second conjecture, we need to make a second implicit assumption: these internal resources are saved through some **safe assets**. Although there are different definitions of safe assets in the existing literature, one common view is that a safe asset is an instrument that is expected to preserve its value under *any* circumstances. In other words, a safe asset is fundamentally different from a financial asset. The value of a financial asset fluctuates according to shocks to economic fundamentals, investors demand, and so on. In contrast, the benefits and costs of carrying safe assets are certain and not linked to any shocks to aggregate or individual demand. As a result, when heterogeneous agents accumulate safe assets, the marginal value of holding safe assets will not be equalized. In the model, we assume corporate cash are all saved through safe assets with an inventory-like saving technology. In reality, companies hold their cash through dollars, the U.S. Treasury bills, and other highly liquid assets.

Given the importance of these two assumptions, in the following part of this section, we will provide some additional discussions on their validity.

Incomplete market

As mentioned before, to obtain firm-market boundaries, one crucial assumption is that the market is not (dynamically) complete. Here we want to argue that due to the following three reasons, this incomplete-market assumption remains a valid one even for the mature financial markets in advanced economies.

First, idiosyncratic risks in the systematically important firms might be the origins of aggregate risk, making these firm-level risks neither insurable nor diversifiable. The production network literature has already shown that with the input-output linkages, micro-level shocks might be the origins of aggregate fluctuations in the whole economy (e.g., Acemoglu et al., 2012). Therefore, these firms have to bear idiosyncratic shocks as their idiosyncratic risk is precisely the uninsurable aggregate risk in the economy. In a recent work done by Gao (2021), the author investigates the cash holding behaviors of firms that lie at the center of the U.S. input-output production network. In theory, these firms face undiversifiable shocks and should hold liquidity assets as precautionary savings. Consistent with the theory, Gao (2021) finds that compared to non-central firms, central firms have higher exposures to aggregate shocks and hold more cash reserves.

Second, according to the pecking order theory of capital structure, equity is the most expensive financing tool for companies. One of the leading explanations is that compared to debt financing, equity financing generates substantially higher value dilutions to the existing shareholders (Myers and Majluf, 1984). The intuition is that equity is more sensitive to private information and hence generates higher mispricing arising from asymmetric information. In this perspective, even firms are allowed to finance their investment with 100% equity to perfectly diversify away their idiosyncratic risk, they optimally choose not to as it will be considerably expensive.

Besides, equity finance is likely to become more and more expensive in this new Superstar Economy. Here we want to use the results in Fulghieri, Garcia and Hackbarth (2020) to help explain the underlying mechanism. According to Fulghieri, Garcia and Hackbarth (2020) (Proposition 3), whether equity is more diluting than debt depends on whether the information costs are concentrated on the right tail of the outcome distribution. As shown in Figure A13 in the appendix, due to the difference in their payoff structure, debt creates more dilution to firm value on the left-hand side while equity generates more on the right-hand side. Therefore, whether the optimal security should be more debt-like or equity-like depends on the distribution of the asymmetric information. As shown before, in a Superstar Economy, earnings become a convex function of capital quality. Therefore, information costs are severer in the right tail. In this way, the optimal security is more likely to be debt or have a more prominent debt component. In fact, Doidge, Karolyi and Stulz (2017) find that many companies start to buy back their shares. For example, in the first quarter of 2020, 58 of the 70 S&P 500 companies providing information about buybacks have reported that they repurchased shares from the market.

Third, in reality, some non-economic considerations or real frictions might prevent firms from achieving perfect risk-sharing, even in a mature financial market like the U.S.. To begin with, entrepreneurs have to hold a substantial amount of equity at hand to keep control of their companies. Furthermore, tax benefits and dead-weight costs from bankruptcy give entrepreneurs more incentives to use debt instead of equity to finance their investment.

Given the three reasons above, this uninsurable idiosyncratic risk assumption should be a valid one. It could come exogenously from the input-output production structure or endogenously from the entrepreneur's optimal choices.

Safe assets

At the same time, in order to obtain Coase (1937)'s second conjecture that a firm is a substitution of the market, we need an additional assumption: firms accumulate their internal savings through some safe assets. In order to validate this assumption, first, we want to explain what is the fundamental difference between cash and any other type of financial asset. In the model, we explicitly assume that cash is saved with an inventory technology while debt is modeled as a publicly-traded financial asset with zero net supply. As cash is not publicly traded, its marginal value will not be equalized across different firms. However, this setup is used purely for the simplicity and tractability. The key characteristic of a safe asset lies in the fact that the costs and benefits of holding it are pre-determined and do not fluctuate with the varying demand for that asset.

Here we want to use a simple example with the classic quantity theory of money to illustrate the essential difference between a safe asset and a normal financial asset. Assume that firms accumulate cash through holding money, and money only works as a medium of exchange in this economy. The benefits of holding money are set to be zero, and for simplicity, we assume that the costs of cash holdings come from inflation costs only. According to the quantity theory of money, price level P is affected by monetary policy through this well-known equation $vM = PY$, where v is the velocity of money for all transactions, Y is the aggregate real value of transactions, and M represents the total nominal amount of money in circulation. If the central bank does follow a pre-determined inflation target (as we do observe from historical inflation data in Figure A12 in the appendix), then the government will change money supply endogenously according to changes in demand for money, such that the price level will grow at a roughly constant rate. In this simple example, price stability makes money a safe asset, and the critical difference between a safe asset and a normal financial asset lies in whether the net supply of that asset endogenously reacts to the changes in asset demand. Therefore, here the origin of a safe asset comes from the unintended consequence of inflation targeting strategy of the central bank.[21] In addition, many recent studies have

[21] A similar conclusion can be drawn with the fiscal theory of price level. At the heart of this theory is that given the present value of all the future primary fiscal surpluses, the price level is the inverse of the value of government debt and fiat money. Again, the inflation targeting strategy leads to the fact that the value of government liabilities also stays stable over time, making these money and short-term T-bills safe assets. Besides, we can also assume that the cash market clearance condition obeys $\int \omega_{i,t} di = \mathcal{B}_t$. It means that different from the bond market, the aggregate net cash supply $\mathcal{B}_t$ is not zero. More importantly, the government chooses a path of aggregate cash supply $\mathcal{B}_t$ such that the price of cash remains stable over time.

shown that governments have strong incentives to make their debt safe assets, possibly due to the substantial convenience yields and the reduction in debt interest rates (e.g., Krishnamurthy and Vissing-Jorgensen, 2012; Jiang et al., 2021).

Of course, the precise reason why the government wants to create safe assets is way beyond the scope of this paper. Instead, our main purpose here is to argue that these safe assets are valuable to firms as internal resources because their value can be preserved under any circumstances. More importantly, these safe assets make it possible for firms to become a substitution of the market system for resource allocation. In addition, this safe asset characteristic is also one of the key differences between this paper and the so-called HANK (Heterogeneous Agent New Keynesian) model à la Kaplan, Moll and Violante (2018*a*). In HANK models, there are also two types of assets: liquid and illiquid. However, one crucial difference is that the liquid asset is also a safe asset in this paper, as the costs and benefits of holding it are immune to fluctuations in the economy and financial markets.

2.5 Conclusion

Over the past several decades, we have observed three puzzling macro-finance trends in the data: increasing corporate market power, increasing corporate internal financing, and deteriorating capital allocation efficiency. In this paper, we argue that these three phenomena are deeply connected, and we provide a theoretical and quantitative framework to explain them jointly. The underlying mechanism comes from that two economic fundamental changes from both demand and supply sides can directly impact the level and volatility of corporate earnings. In addition, they lead to both micro-indirect impacts on risk management and macro-indirect impact misallocation. To formally establish this idea, we introduce product market competition and corporate risk management into a standard continuous-time heterogeneous agent model with incomplete markets. In this way, Coase (1937)'s firm-market boundary exists in general equilibrium, and it is endogenously determined by a set of Neumann boundary conditions of certain partial differential equations originating from the entrepreneur's optimal choice. The changes in consumers' taste for quality and producers' marginal supply cost increase the earnings-quality gradient sharply in the right tail, which generates a "winners-take-most" phenomenon and makes current winners inherently riskier. This income and risk redistribution generates a positive correlation between markup and cash value, prompting superstar firms to rely more on internal financing. At the same time, an expansion of the internal financing region weakens the role of the price mechanism in allocating resources, thus leading to an increase in misallocation. After that, we implement several reduced-form empirical investigations to show that our model predictions actually hold in the data. Finally, the model can quantitatively match some important macro-finance trends when taken to the data. It shows that the area disciplined by the market system has declined by about 11% during the past forty years.

The punchline in this paper is that misallocation, narrowly defined as the economy's inability to allocate resources across different agents, increases in the new economy with superstar firms. In terms of the policy implications, this paper points out two types of market failures in the upcoming Superstar Economy. First, the effectiveness of the price mechanism is shrinking as massive corporate internal financing behaviors are limiting its role. Second, as both risk and profits are redistributed to productive firms, these superstar firms have less incentive to invest, leading to a secular decline in business dynamism. The government's role as the visible hand in the new economy and other normative works exploring the optimal policies are left for future research.

Chapter 3

The Rise of (Mega-)Firms with Negative Net Earnings

3.1 Introduction

Conventional wisdom has it that negative or abnormally low net earnings are bad signals for companies. It indicates that these firms could suffer from at least some cyclical issues or even deeper and long-term problems. However, this argument seems no longer true for the new economy. In the past several decades, we have seen many billion-dollar companies with negative net earnings. We provide some examples in Table 3.1. Far from being in trouble, these mega-companies are the most popular and highly-owned public companies in the U.S. It seems that they will thrive for years to come.

Table 3.1: Examples of billion-dollar companies with negative net earnings in 2019

Company name	Market capitalization	Net earnings
Zillow	9.59	-0.31
Pinterest	10.62	-1.36
Lyft	13.02	-2.60
Snap	23.12	-1.03
Spotify	27.57	-0.19
Uber	51.05	-8.51
Tesla	75.72	-0.86

Data source: Yahoo Finance. All numbers are in billion U.S. dollars. We choose 2019 to avoid the impacts of the pandemic outbreak.

This paper argues that those mega-companies listed in Table 3.1 are by no means outliers. As we will see in Section 3.2, the fraction of unprofitable firms has increased substantially in the past several decades. Based on the public-firm-level data for the U.S. economy, we show that the share of firms with negative *net earnings* has risen from 18.3% in 1970 to 54.4% in 2019. However, most of them are still profitable in terms of *gross income*. In addition, when investigating the underlying distributional changes, we find that the mean has considerably shifted to the right over time, which indicates a growing number of mega-firms with huge losses. We document a similar upward trend based on the Initial Public Offerings (IPO) dataset. In 1980, only 24% companies were not making money when they went public. However, this number increased to 77% in 2019. Finally, in terms of international evidence, we show that this growing fraction of unprofitable firms is not a unique phenomenon in the U.S. only. Additionally, on average, the percentages of firms with negative net earnings are higher in rich countries.

After that, we examine the possible explanation behind this long-run trend. We hypothesize that the increasing returns-to-scale on the production side is why companies today have substantially stronger incentives to spend resources on building up their customer bases. Our argument is based on the existing customer capital literature(e.g., Gourio and Rudanko, 2014; Dinlersoz and Yorukoglu, 2012). The key friction in this branch of literature lies in the search cost: firms need to conduct some market activities to sell their products to potential buyers. In this way, the existing customer base limits how much each firm can serve the whole market. With modern technology such as ICT-based technologies, the marginal value of an additional customer has increased substantially due to the considerable reduction in the marginal production cost. Therefore, firms have stronger incentives to spend resources on acquiring new customers. In other words, we interpret those mega-firms arising from natural monopolists. Their natural monopoly power comes from increased scale economies instead of falling competition or regulation. Nevertheless, firms need to pay substantial costs upfront to build their customer base first to exploit the natural monopoly power. Hence, the reported net earnings are extremely low or even negative before their customer base has reached a certain level.

We provide three sets of empirical evidence to support our previous hypothesis. First, we discover that companies have changed their business model substantially in the past several decades. The ratios to sales of capital investment and production-related costs have declined sharply over time. Meanwhile, those of customer capital and

R&D expenses have grown substantially. Besides, there is a crucial difference between the latter two. Left-tail firms with the lowest gross profitability mainly drive the upward trend in R&D expenditures. Meanwhile, increasing customer capital expenses mostly comes from right-tail firms with the highest levels of gross profitability. Second, we empirically show that earning losses are closely related to measured corporate market power. We follow De Loecker, Eeckhout and Unger (2020)'s methodology to provide firm-level markup measures. We find that cross-sectionally, firms with higher markups tend to have higher customer capital expenses and lower net earnings. This negative association implies that the origin of corporate market power may be closely related to their customer base. Third, we document that an industry's marginal cost of production is significantly and negatively correlated with the fraction of firms with negative earnings in that industry. We obtain industry-level measures on the marginal cost of production by following De Ridder (2019). We document a significant and negative association between these two indicators in the data. These three sets of findings provide empirical support to our customer capital story. In this way, we rationalize the rising popularity of mega-firms with negative net earnings instead of treating it as the same speculative behavior as in the late 1990s.

Related literature and contributions Our paper is closely related to three branches of literature. First, our work builds on the literature highlighting the importance of customer capital for industry dynamics.[1] This discussion can date back to the classic *Price & Advertising model* developed by Phelps and Winter (1970), Luptacik (1982), and Feichtinger (1982). After that, many researchers have started to investigate different aggregate implications related to this customer capital. For example, Rotemberg and Woodford (1992) argue that a dynamic general equilibrium model with oligopolistic competition can generate substantial aggregate demand shocks, which turns out to be important for matching the empirical responses estimated with postwar U.S. data. In addition, Ravn, Schmitt-Grohe and Uribe (2006) investigate how endogenous customer capital choice leads to countercyclical markups. Meanwhile, Dinlersoz and Yorukoglu (2012) show that allowing firms to build up their customer base can substantially shape the equilibrium firm size distribution. In recent years, there has been a growing macro-finance literature on this topic. For instance, Dou and Ji (Forthcoming) use a monopolistic competition with customer capital framework to provide an interesting result that

[1] Bagwell (2007) provides an excellent review of the existing literature.

the optimal markup is pinned down by the trade-off between profiting from current customers and developing potential buyers. Besides, Morlacco and Zeke (2021) investigate the interaction between corporate customer capital expenses and monetary policy. They find that advertisement expenses of large firms are more sensitive to monetary policy shocks than those of small firms. Our contribution is to show that customer capital might be the origin of corporate market power in the new economy. A larger customer base allows firms to benefit more from the economies of scale. In this perspective, we claim that corporate market power comes from a larger customer base instead of reduced regulation or other political forces. Our conclusion has completely different policy implications. For instance, we argue that regulators should pay serious attention to industries with higher user-switching costs. In these industries, corporate market power could be sticky once built up. It deters small and innovative firms from entering the market unless the innovation is substantially large.

Second, this paper also connects to the changing business dynamism literature. Jones and Philippon (2016) and Gutierrez and Philippon (2017) document this secular stagnation of corporate investment in the U.S.. The major finding in these papers is that from the early 2000s, corporate investment incentives become weaker, despite the increasing profitability and valuation. Specifically, Gutierrez and Philippon (2017) argue that this pattern could come from three different reasons: intangible capital, market concentration, and corporate governance. Some other related studies have documented a similar pattern with data from some European or developing countries (e.g., Lewisi et al., 2014; Bussiere, Ferrara and Milovich, 2015; Kose et al., 2017). Besides, Kilic, Yang and Zhang (2019) discover that the cross-sectional relation between investment and profitability among U.S. public firms has changed from positive to negative in the past several decades. Olmstead-Rumsey (2021) and De Ridder (2019) are the two papers arguing that declining innovation incentives could be the reason behind the declining business dynamism. Specifically, Olmstead-Rumsey (2021) suggests that the declining innovativeness of market laggards can account for about 40 percent of the rise in market concentration and the 100% productivity slowdown in the past several decades. In contrast, De Ridder (2019) argues that declining marginal costs and rising fixed costs associated with intangible capital contribute to the rise in market concentration. Our contribution is to provide a different explanation for all these trends. We contend that these changes mainly come from companies switching their business model from capital-investment-driven growth to money-burning expansion.

Third, this work contributes to the growing literature on superstar firms. The existing works can be classified into two categories, one focusing on the consequences while the other on the origins of this new superstar economy. For the first category, Autor et al. (2020) and Kehrig and Vincent (2020) argue that the rise of superstar firms is the primary driver of the declining labor share. Similarly, De Loecker, Eeckhout and Unger (2020) claim that the rising markup of large firms could contribute to the declining labor and capital shares and the decrease in labor market dynamism. Besides, Su (2021) investigate how the rise of a risky superstar economy could lead to more internal financing and thus declining capital allocation efficiency. In terms of the second category of the literature, De Loecker, Eeckhout and Mongey (2021) demonstrate that technological innovation and market structure changes contribute to the rise in market power. Compared to the existing literature, this paper focuses on the earnings dynamics in a winner-take-all economy. More importantly, we establish a new fact in addition to the growing literature on changing firms' behaviors. Most of the existing studies are focused on changes in corporate internal financing (e.g., Bates, Kahle and Stulz, 2009*b*), investment (e.g., Gutierrez and Philippon, 2017), or profitability (e.g., Davis, Sollaci and Traina, 2021). In contrast, we investigate the trends in the fraction of firms with negative net earnings.

Layout The rest of the paper is organized as follows. Section 3.2 provides the empirical evidence on secular trends in the fraction of firms with negative net earnings. In Section 3.3, we first conjecture that the increasing economies of scale might be the explanation behind this trend. Then we provide three sets of empirical findings to support our hypothesis. Finally, Section 3.4 concludes.

3.2 Long-term Trends in the Fraction of Unprofitable Public Firms

3.2.1 Data and Variable Construction

Data for the empirical analysis in this section is obtained from *Compustat*, which contains balance-sheet information for publicly listed U.S. companies. We keep all the entries with a foreign incorporation code of "USA", exclude financial firms (SIC 6000-6999) and regulated utilities (SIC 4900-4999), and drop firms with missing/negative values on assets or sales.[2] For global firms, we obtain the dataset from *Global Compustat* dataset, and we conduct similar data cleansing processes. Likewise, *Global Compustat* dataset also provides firm-level balance-sheet information, and it covers publicly traded companies in more than 80 countries. In addition, it represents over 90% of the world's market capitalization.

All variables are constructed by following some recent research studies or the standard practice in the empirical corporate finance literature. We obtain a firm's net earnings from *Compustat* data item *NI*. This item reports the income or loss of a certain company after subtracting *all* expenses and losses from all revenues and gains. In contrast, a company's gross profit (*Compustat* data item *GP*) only subtracts cost of goods sold (*Compustat* data item *COGS*) from total revenue (*Compustat* data item *REVT*). Following the work of Morlacco and Zeke (2021) and Peters and Taylor (2017), we measure firms' expenses on customer capital by computing the net selling, general, and administrative expenses (net XSGA), which is the difference between *Compustat* data item *XSGA* and data item *XRD*. We adopt this approach because expenses on salespeople, marketing, and advertising are usually reported directly in the "Selling, General and Administrative Expenses" (*Compustat* data item *XSGA*). However, in *Compustat* dataset, this item also contains R&D expenditures (*Compustat* data item *XRD*). Therefore, following the existing studies, we use the difference between these two as a proxy for customer capital expenses. To measure firm-level markup, we use the methodology proposed by De Loecker, Eeckhout and Unger (2020). Generally speaking, a firm's markup is estimated as the product between the elasticity of output concerning variable inputs and the revenue share of each variable input.

In addition, we also include some other firm-level characteristics when conducting our empirical analysis. A firm's output is defined as the net sale or turnover (*Compustat* data item *SALE*) and firm size as the natural logarithm of total assets (*Compustat* data item *AT*). Firm age is computed as the year difference from its first appearance in *Compustat*. The book leverage is computed as the ratio of total debts to the sum of total debts and common equity. We measure a firm's return of asset as income before extraordinary items (*Compustat* data item *IB*) scaled by total assets. Asset tangibility

[2] One exception is that we do not exclude financial firms when we explore the industry heterogeneity in these trends.

is the fraction of physical assets in total assets. Investment is obtained as the capital expenditures (*Compustat* data item *CAPX*) scaled by total assets. R&D activities are measured as research and development expenses divided by total assets. Dividend payouts of different firms are captured by dividends (*Compustat* data item *DVC*) scaled by total assets. For all nominal variables, we deflate them by using the annual national consumer price index (CPI) obtained from the U.S. Bureau of Labor Statistics (BLS).

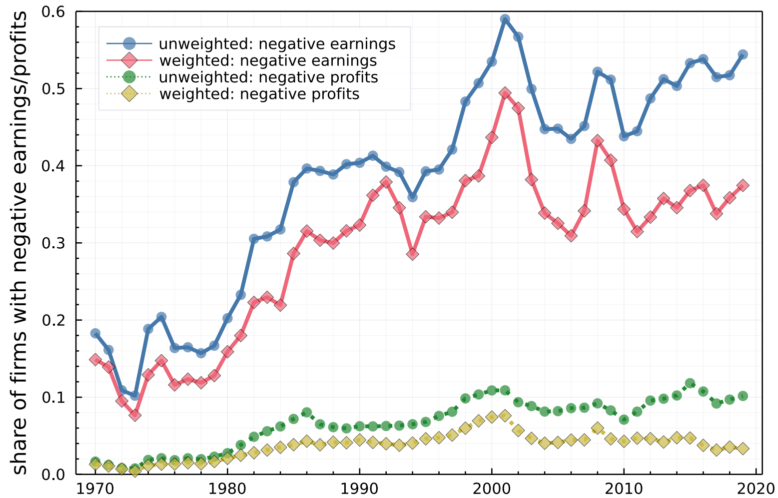
3.2.2 Evidence from the U.S. Public Firms

Aggregate trends

Figure 3.1 presents our baseline result on the time series of the fraction of firms with negative net income. More specifically, in each year, we count the number of firms with negative net incomes and divide it by the total number of firms. We provide two different indicators: one is weighted by the relative output share of the industry that a firm belongs to, and the other unweighted. As we can see from Figure 3.1, there is a steady increase in the share of firms with negative earnings in both measures. For the unweighted indicator, only a fraction of 18.3% firms had negative net income in 1970. However, this number increased to 54.4% in 2019. As for the weighted indicator, this number has changed from 14.8% in 1970 to 37.4% in 2019. Although there is a significant drop around 2000, this upward trend has picked up in recent years. Therefore, based on this simple exercise, we document a secular upward in the fraction of unprofitable public firms in the U.S.

Robustness checks In addition, we have conducted several different robustness checks. To begin with, we show that this upward trend is not limited to one specific industry. In Figure B1 in the appendix, we plot the share of unprofitable firms for each of the following ten industries: Agriculture, Forestry, & Fishing (SIC 01-09); Mining (SIC 10-14); Construction (SIC 15-17); Manufacturing (SIC 20-39); Transportation & Public Utilities (SIC 40-49); Wholesale Trade (SIC 50-51); Retail Trade (SIC 52-59); Finance, Insurance, & Real Estate (SIC 60-67); Services (SIC 70-89); and Public Administration (SIC 90-99). As we can see from Figure B1, the share of unprofitable firms has been increasing steadily in most of these ten industries. In addition, the most pronounced pattern happens in the manufacturing sector, services sector, and public administration sector. In contrast, this pattern is less apparent in industries like finance and insurance. However, our exercise still implies that this upward trend is important to the

Figure 3.1: The rise of firms with negative earnings



Notes: This figure presents the time-series plot of the fraction of unprofitable public firms. In each year, we count the number of firms with negative profits and divide it by the total number of firms. We use two different profitability measures (gross profits and net earnings) and two different aggregating approaches (weighted and unweighted). The weight is computed as the economy's output share of the industry that a firm belongs to. Data is obtained from *Compustat*.

whole economy because the manufacturing and services industries are essential in any developed country.

Then we test whether this phenomenon is driven by the increasing fraction of young firms in *Compustat* dataset. Nowadays, we may have more young public firms with low net earnings. As a result, the increasing fraction of unprofitable firms could purely come from the age effect. To alleviate such concern, in Figure B2 in the appendix, we provide the time series of two age-related indicators. The first one is average firm age, which is presented as the yellow line in Figure B2. As we can see, the average firm age increases over time, which implies that nowadays, on average, we have more mature public firms. The second proxy is the fraction of young firms. Our definition of young firms is these companies with five years or less. This choice of criterion is *ad hoc*, but our conclusion does not depend on this specific criterion. Based on the green line in Figure B2, we can observe that the fraction of young firms fluctuate around some value over time. There is no clear upward trend associated with this proxy.

Finally, we investigate whether this pattern only shows up in particular stock exchanges. As we all know, different stock exchanges have various listing requirements, especially on the financial criteria. Therefore, in Figure B3 in the appendix, we redo our previous exercise but for companies in different stock exchanges. Specifically, the red line in Figure B3 represents the fraction of firms with negative net earnings in New York Stock Exchange (NYSE), the green line is for companies in National Association of Securities Dealers Automated Quotations (NASDAQ), and the yellow line stands for the rest of stock exchanges in the U.S. As we can see from Figure B3, our previous conclusion on the secular rise of unprofitable firms is not limited to one specific stock exchange. Indeed, there exists some heterogeneity across different exchanges. For NYSE, this fraction increased from 10.5% in 1970 to 31.4% in 2019. Meanwhile, for NASDAQ, this number has changed from 15.5% in 1970 to 63.7% in 2019.

Gross v.s. Net

More interestingly, this upward trend is not striking when it comes to the share of firms with negative gross profits. As we can see from the two dotted lines in Figure 3.1, the percentage of firms with negative profits has also increased in the past fifty years. However, the overall importance of those companies to the whole economy is limited. Specifically, with the unweighted measure, the share of firms with negative gross profits has increased from 1.7% in 1970 to 10.2% in 2019. As for the weighted measure, this number changed from 1.3% in 1970 to 3.3% in 2019. Therefore, most public firms are still profitable in terms of gross profits. However, many of these companies may seem in trouble as they report negative or abnormally low earnings.

This difference turns out to be crucial for understanding the underlying mechanism. As explained in the following section, the difference between these two profitability measures mainly comes from the substantial increase in customer capital expenses, especially for the right-tail firms with the highest gross profitability. Intuitively speaking, if a company reports positive gross profits but earnings losses, it indicates that its core business is still profitable. This firm has a negative earning simply because it has spent many resources in expanding the scale of its core business. As we will see in the theoretical explanation, this behavior is rational as firms can benefit more from increasing operating scale in the new economy. In this perspective, current earnings losses imply that firms are in the middle of building up their future advantages.

In addition, the increasing gap between gross profit and net earnings can also help us reconcile the open debate on measuring firm-level markup in the existing literature. De Loecker, Eeckhout and Unger (2020) document that corporate markup has increased substantially in the past several decades. However, some other studies (e.g., Traina, 2021) provide different conclusions. One of the main reasons they obtain different results is that they use different measures of input costs. Traina (2021) use operating expenses but De Loecker, Eeckhout and Unger (2020) use costs of goods sold. In practice, operating expenses include marketing and management expenses, in addition to production-related costs. In this paper, we argue that companies use those sales and marketing expenses to build up their customer base today, in order to obtain market power in the future. Following this interpretation, we should not include those expenses when measuring the current markup.

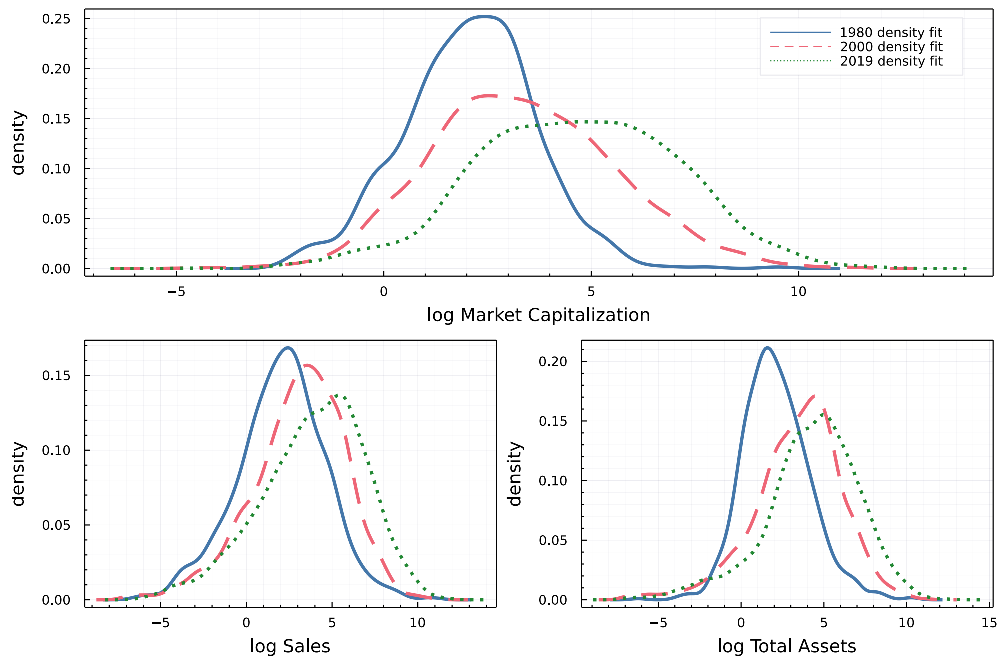
Trends in distribution

Now we investigate the underlying distributional changes in companies with negative earnings. We prepare our empirical results in Figure 3.2 by using the following steps. To begin with, we identify all the firms with negative net earnings each year. Then we plot the size distribution of these unprofitable companies. To show the robustness of our conclusion, we choose three different size-related indicators: market capitalization, total sales, and total assets.

Specifically, to capture the evolution of the entire distributions, we plot the kernel density of each size-related proxy for all the firms with negative earnings in 1980, 2000, and 2019. Likewise, we choose 2019 instead of 2020 to avoid the possible unintended effects of pandemics. The results are presented in Figure 3.2. Based on this figure, we can see that the size distribution of unprofitable firms has changed substantially over time. From unreported results on long-term changes in different data moments, we discover a substantial increase in the mean and standard deviation over time but a considerable decrease in skewness and kurtosis. In addition, the changes in distribution are mainly driven by the right shifts in the mean, which indicates the increasing popularity of large firms with negative net earnings. Compared to the 1980s, nowadays, we have substantially more mega-firms with negative net earnings. In other words, the examples provided in Table 3.1 are by no means some outliers. Instead, compared to forty years ago, there is clearly a growing number of billion-dollar companies with earnings losses in the new economy.

In addition, we can also check how the average size at different percentiles evolves. We calculate the corresponding values at the 90th and 10th percentiles for each year. Then we present the time series plot in Figure B4 in the appendix. As we can see from

Figure 3.2: The rise of firms with negative net earnings: distributional changes



Notes: This figure presents the distributional changes for companies with negative net earnings. In each year, we select all the firms with negative net income and then plot the size distribution. We use three different size-related indicators including market capitalization, total sales, and total assets. We choose 2019 instead of 2020 to avoid the possible unintended effects of pandemics. Data is obtained from *Compustat*.

this figure, the most striking increase comes from firms at the right-tail distribution. The net-income-to-asset ratio at the 90th percentile has increased substantially, while the left-tail has remained roughly unchanged. Again, this conclusion does not depend on which size indicator we choose for our analysis.

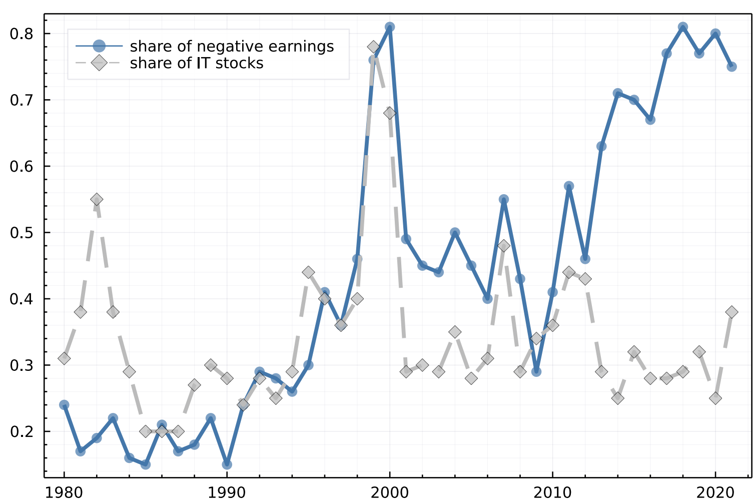
Evidence from Initial Public Offerings

Now we supplement our previous analysis with the IPO dataset provided by Jay Ritter.[3] Figure 3.3 presents the fraction of companies with negative net earnings when they initially went public in the U.S. Following the common practice, the information related to corporate earnings is measured at the most recent twelve months before going public.

[3] We obtain the IPO-related information from Jay Ritter's personal website: https://site.warrington.ufl.edu/ritter/ipo-data/.

Similarly, we estimate the fraction by calculating the ratio of IPO firms with earnings losses to the total number of firms going public in that year. The solid blue line in Figure 3.3 represents the time series plot of this indicator. It clearly shows that the share of IPO firms with negative net earnings has increased steadily in the past several decades. More specifically, in 1980, only 24% of firms did not make money when going public. In contrast, this number rose to 77% in 2019.

Figure 3.3: The rise of IPOs with negative net earnings



Notes: This figure presents the time-series plot of the fraction of unprofitable IPOs. In each year, we count the number of IPOs with negative net earnings and divide it by the total number of IPOs. The information related to corporate earnings is measured at the most recent twelve months before going public. The share of IT stocks is computed as the relative ratio of IT-related IPOs to total IPOs in each year. Data is obtained from Jay Ritter's personal website.

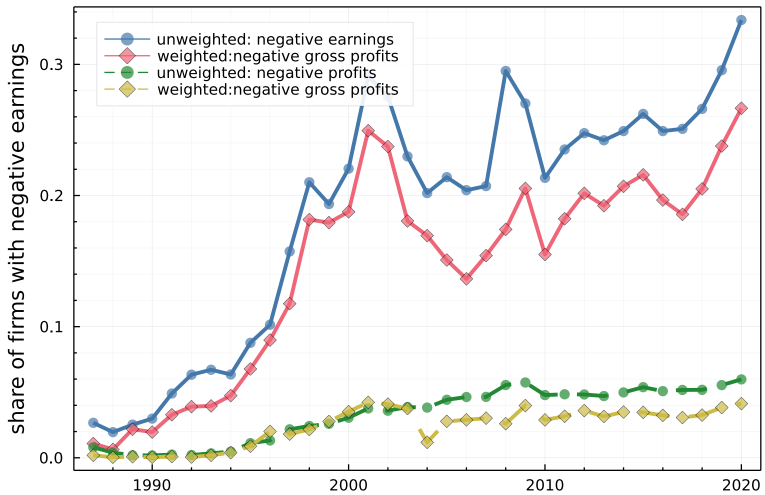
More importantly, this upward trend is not entirely driven by the increasing IPOs for IT firms. The gray dotted line in Figure 3.3 represents how the fraction of IT-related IPOs changes over time. Before 2000, we can observe that the trends in the share of unprofitable IPOs were likely to be driven by the changes in the relative importance of IT firms. However, after 2000, it is no longer the case. Although the share of IPOs with negative income has increased substantially during this period, the relative fraction of IT stocks remain relatively stable. One possible explanation is the emergence of non-traditional IT companies with earnings losses, such as Tesla and Peloton. This finding is also consistent with our previous evidence documented in Figure B1 that this secular upward trend shows up in many different industries.

3.2.3 Global Evidence

The data on the U.S. publicly-traded firms could suffer from some selection bias. In other words, the facts documented before may be simply some unique phenomena that show up in the U.S. only. In order to alleviate such concern, we repeat our former analysis but this time with a global firm-level dataset. Our main results are provided in Figure 3.4. Again, we present the time-series plots on the fraction of firms with negative net earnings or with negative gross profits. Similarly, we provide two different time series for each variable: one is weighted by the relative importance of the industry that a firm belongs to and the other unweighted. As we can see from the two solid lines in Figure 3.4, there exists a global rise in the share of firms with negative earnings. This conclusion does not depend on which measure we use. More specifically, for the unweighted measure, 2.7% firms in 1987 had negative net incomes. However, this number increased to 29.6% in 2019. As for the weighted measure, this fraction has changed from 1.1% in 1987 to 26.4% in 2019. Same as before, this upward trend is less pronounced when we focus on the share of firms with negative gross profits. According to the two dashed lines in Figure 3.4, the percentage of firms with negative profits has increased a bit in the past thirty years. Nevertheless, the numbers are relatively small. For the unweighted measure, the share of unprofitable firms has increased from 0.8% in 1987 to 5.5% in 2019. In contrast, for the weighted measure, it has grown from 0.2% in 1987 to 4.3% in 2019. To sum up, from our exercise here with the global firm-level dataset, we show that our previous findings with the U.S. firms are global phenomena.

Another interesting finding here is that cross-sectionally, countries with higher real GDP per capita are associated with a higher fraction of firms with negative net earnings. Figure 3.5 presents the binned scatter plot between log real GDP per capita and percentage of unprofitable firms in different countries. The blue dash line represents the fitted linear relationship between these two variables. There exists a positive and significant relationship between them in the data: we are more likely to observe firms with earnings losses in rich countries. This positive cross-country relationship is crucial. It indicates that the rise of unprofitable firms may not come from bad institutional quality or poor corporate management. The underlying reason could come from either demand- or supply-side stories. In terms of the supply-side story, there may be more high-tech firms in rich countries. These companies usually employ more intangible capital and deliver some characteristics of increasing returns to scale. However, as we will explain later in the next section, these firms need to spend substantial expenses upfront

Figure 3.4: The global rise of firms with negative net earnings



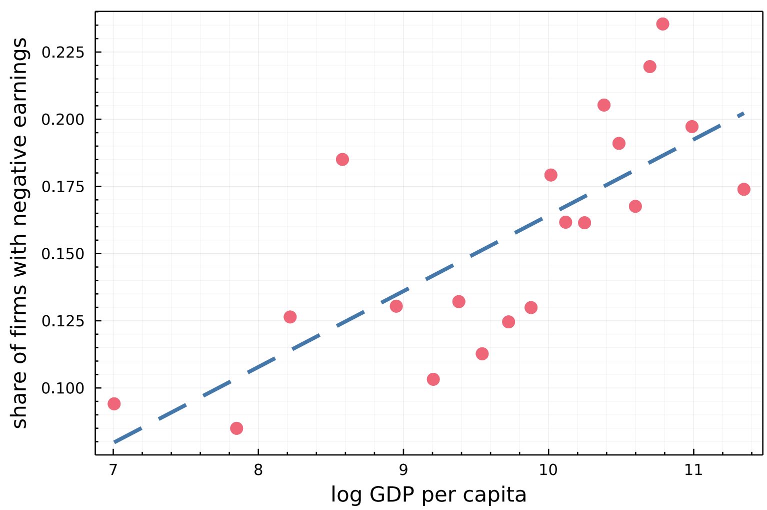
Notes: This figure presents the time-series plot of the fraction of unprofitable public firms. In each year, we count the number of firms with negative profits and divide it by the total number of firms. We use two different profitability measures (gross profits and net earnings) and two different aggregating approaches (weighted and unweighted). The weight is computed as the country's output share of the industry that a firm belongs to. Data is obtained from *Global Compustat*.

on building up their user networks. Therefore, they are more likely to report negative net earnings before becoming superstar firms with dominant market shares. In terms of the demand-side story, developing countries tend to have less mature financial markets. In addition, the IPO regulation requirements are more strict and thus require higher listing standards. Therefore, firms with negative net earnings are less likely to get IPO approval in emerging countries. Generally speaking, we argue that both sides could play an essential role in explaining this positive and significant relationship across different countries.

3.3 Inspecting the Underlying Mechanism

This section explains this long-run upward trend in the fraction of firms with negative net earnings. To begin with, we conjecture that the increasing returns-to-scale, arising from new technology such as digitization, is the underlying reason behind. This argument is based on the existing customer capital literature. After that, we present three sets of empirical findings to support this hypothesis. First, companies, especially

Figure 3.5: The fraction of firms with negative net earnings and real GDP per capita



Notes: This figure presents the binscatter plot between the fraction of firms with negative net earnings and log real GDP per capita across different countries. The fraction of firms with negative net earnings is measured as follows: in each year, we count the number of firms with negative net earnings and divide it by the total number of firms. All the data is obtained from *Global Compustat*. Real GDP per worker is obtained from Penn World Table (PWT) and computed as output-side constant-price real GDP divided by employment.

the highly profitable ones, have changed their business model substantially in the past several decades. Nowadays, profitable firms neither invest much in physical capital nor innovate greatly. Instead, they spend considerable expenses on building their customer base. Second, firms with higher markups tend to have higher customer capital expenses and lower net earnings. In other words, powerful companies tend to spend more on their customer capital and lower net incomes. Third, the fraction of firms with negative net earnings is higher in industries with lower marginal production costs.

3.3.1 Hypothesis and Theoretical Explanation

Here we provide one possible explanation on the rising share of (mega-)firms with negative earnings. We conjecture that this trend is closely related to the increasing returns-to-scale in the new economy. Some recent studies have shown that since the 1980s, companies in advanced economies have seen substaintial reductions in marginal cost of production and hence a major in operating scale (e.g., De Ridder, 2019; Hoberg and Phillips, 2021). Here we argue that the increasing scalability could affect firms' optimal decisions on customer capital and their net earnings dynamics.

The theoretical framework in our mind is mainly based on Gourio and Rudanko (2014). The key assumption in the existing customer capital literature (e.g., Phelps and Winter, 1970; Ravn, Schmitt-Grohe and Uribe, 2006; Dou, Ji and Wu, 2021) is that the product market has search frictions and companies need to conduct advertisement or other marketing activities to sell their products to potential buyers. With a frictional product market with search and matching costs, the total sales units cannot exceed the minimum of customer base and production capacity. In this way, we can easily see that the customer capital becomes more valuable when firms can benefit from increasing their operating scale.

More specifically, based on Gourio and Rudanko (2014)'s framework, the benefits of having one additional customer today come from not only an increase in today's sales revenue by the unit price of products but also the expected increase in the continuation value. The second effect arises because we assume that the new customer will stay with the firm in the next period with some positive probability. In contrast, the cost of one additional customer is from the marginal production cost. Together, these three components determine the marginal value of an additional customer to firms.

With this theoretical framework, we can easily see that the marginal value of an additional customer increases when there is a reduction in the marginal production cost. Meanwhile, companies have stronger incentives to spend many customer capital expenses upfront to build up their customer base due to their continuation value. However, earnings will be relatively low when the existing customer base is still small. The net earnings will eventually turn positive when the firm's customer base has reached a certain level. Before that turning point, these companies continue to report high operating expenses and large losses.[4]

[4]This argument is also consistent with what we observe in reality. A typical example is Amazon and Walmart. These two companies are the two most prominent retailers in the U.S. One crucial difference between them is that e-commerce sales only consist of a small fraction of Walmart's sales. In contrast, it consists of the primary income source of Amazon. In other words, Amazon's business model should have a higher degree of returns-to-scale compared to that of Walmart. We present the time series of their historical net income and revenue information in Figure B7 in the appendix. Both companies have witnessed substantial increases in their total revenues. As we can see from this figure, Walmart's net-income-to-revenue ratio is relatively stable over time. In contrast, Amazon's net income was extremely low and negative when its total revenue was low. However, when it has successfully acquired more customers and increased its total sales, the net-income-to-revenue ratio becomes positive and steadily increases over time. In 2018, it became more profitable than Walmart, and its net-income-to-revenue percentage is still growing. As shown in Figure B8, another similar example is Tesla.

Additionally, the increasing scalability generates asymmetric impacts on the customer and physical capital expenditure. The underlying mechanism is that changes in returns-to-scale will broadly impact the marginal cost of production but not so much on the optimal composition of different productive factors. As a result, the optimal investment-to-capital ratio does not increase in scalability, which generates a declining investment-to-profitability ratio in the new economy. This theoretical prediction is consistent with recent empirical findings that there is a secular stagnation of corporate investment in the U.S., despite the rising profitability and valuation (Jones and Philippon, 2016; Gutierrez and Philippon, 2017).

There are two caveats for our theoretical explanation here. First, we interpret those mega-firms arising from natural monopolists. To exploit this natural monopoly power, companies need to pay substantial costs upfront to build their customer base first. Therefore, negative earnings may not result from companies running into problems. Instead, it indicates that these firms are on their way to becoming superstars. In this perspective, corporate monopoly power comes from increasing economies of scale instead of falling competition or regulation. In addition, our interpretation here is also different from one standard view that increasing market valuation of money-losing enterprises is a classic bubble sign.

Second, we argue that exogenous changes in scalability lead to endogenous transformations in the corporate business model. However, there could be other potential explanations as well. For instance, as shown in Morlacco and Zeke (2021), different interest rate environments could also affect firms' strategic interactions. They use a theoretical model to show that large firms spend disproportionately more on customer capital investment under the low-interest-rate environment. Our technology story is very different from their interest rate story. Besides, in the last part of this section, we provide additional empirical support by exploiting cross-industry differences in the production function.

3.3.2 Fundamental Changes in Corporate Business Model

Many existing studies have documented the secular stagnation of corporate investment in the U.S. (e.g., Hall, 2014; Jones and Philippon, 2016; Gutierrez and Philippon, 2017; Alexander and Eberly, 2018). This weak investment incentive is quite puzzling as profitability or valuation has been relatively stable or even increased. In this section, we demonstrate that the declining investment occurs during the same time when firms vastly increase their expenses on customer capital, especially for firms with the highest gross profitability.

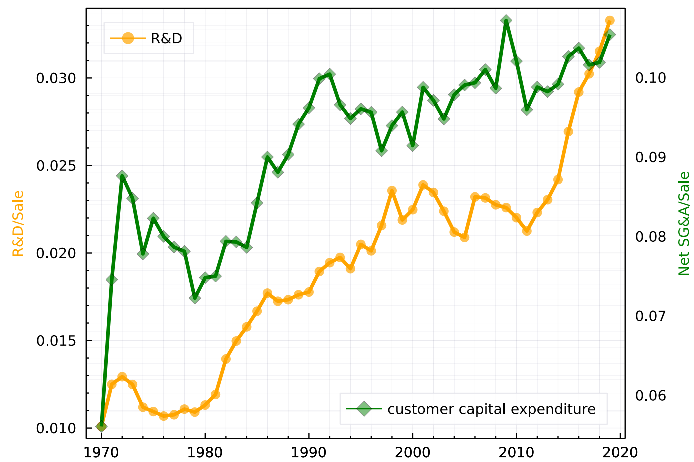
We begin with the average values of all firms and then turn to cross-sectional differences for firms with different profitability. The key message for changes in the average numbers is shown in Figure 3.6. We present the time-series plot of average R&D and customer capital expenses in the top graph. Meanwhile, we plot the time series of average production costs and investment in the bottom chart. All these four indicators are scaled by sales for better comparison. As we can see from Graph (A), the aggregate ratio of customer capital expenses to sales has increased steadily from 5.6% in 1970 to 10.5% in 2019. Simultaneously, the R&D-to-sale ratio has changed from 1.0% to 3.3%. In contrast, Graph (B) shows that the average ratio of production costs to sales has declined from 71.9% in 1970 to 66.0% in 2019. Meanwhile, the investment ratio has decreased from 9.0% to 5.8%. Suppose we simply focus on long-run trends in corporate investment. In that case, the overall business dynamism would appear to decrease as firms nowadays have less incentive to conduct capital investment. However, if we investigate these four trends jointly, one possible explanation for the seemingly declining dynamism could be that companies nowadays have changed their business model from investment-led growth to expenses-driven expansion. Our interpretation here is quite different from the existing literature, which argues that the falling competition mainly contributes to the weak corporate investment (e.g., Jones and Philippon, 2016; Gutierrez and Philippon, 2017).

In addition, our conclusion here is robust to alternative ways of aggregating firm-level information. In Figure 3.6, we calculate the aggregate ratios of each indicator. In addition, we do not weigh each firm according to its relative importance in the economy. However, in Figure B5 and B6 in the appendix, we present the results of using median value and weight each firm according to their relative economic importance, respectively. As we can see from these two figures, our main conclusion does not change with these modifications.

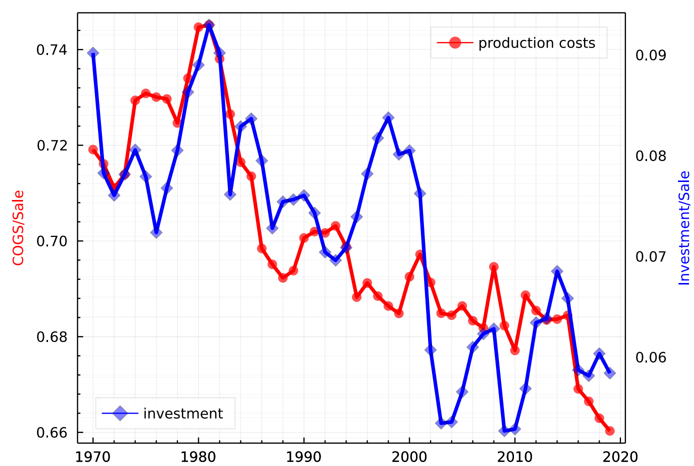
More importantly, right-tail firms, i.e., firms with the highest gross profitability, behave considerably different from the rest. Figure 3.7 presents the same dynamics but for firms with different levels of profitability. More specifically, in each year, we classify firms into five different groups according to their gross profitability. After that, we compute the average values of investment-to-sale, R&D-to-sale, customer-capital-expense-to-sale, and production-cost-to-sale ratios. In Graph (A) of Figure 3.7, we show

Figure 3.6: Changing business model: aggregate levels

(A) increasing expenses on R&D and customer capital



(B) decreasing expenses on production costs and investment



Notes: Graph (A) presents the time series plot of average R&D and customer capital expenses. Graph (B) plots the time series of average production costs and investment. All these four indicators are scaled by firm-level sales for better comparison. Data is obtained from *Compustat*.

the time series of investment-to-sale ratios for firms at different profitability quintiles. As we can see, before 2000, profitable firms, on average, invest substantially more than other groups of firms. For instance, at the peak year of 1981, the investment-to-sale ratio for the top 20% profitable firms has reached 12.8%. In contrast, the investment-to-sale ratio for the bottom 20% profitable firms in the same year was only 3.6%. However, this pattern has changed considerably in the past decades. In 2019, the investment-to-sale ratio for the top 20% profitable firms was 3.5%, and that for the bottom 20% was 3.0%. The difference becomes quite negligible. This broken-link between investment and profitability is similar to Kilic, Yang and Zhang (2019)'s work, but here we represent their key conclusion in a different format.

Meanwhile, those highly profitable firms spend vastly on their customer capital. As shown in Graph (B) in Figure 3.7, there is a significant divergence in customer capital expenses among firms with different gross profitability. Such a gap has been increasing steadily in the past fifty years. More specifically, the average customer-capital-expenditures-to-sale ratios for the top 20% and the bottom 20% profitable firms in the 1970s were 30.3% and 7.9%, respectively. However, in the 2010s, these numbers have changed to 54.4% and 15.1%. In other words, the gap in customer capital expense has risen from 7.2% to 24.1% in the past fifty years. What is surprising is that profitable firms tend not to innovate much more than others. In the bottom left graph in Figure 3.7, we present the time series of R&D-to-sale ratios for firms at different profitability quintiles. As we can see from this graph, R&D expenses among the least profitable firms have increased substantially in the past fifty years. Meanwhile, the innovation incentive of the rest groups has been stable or increasing modestly.

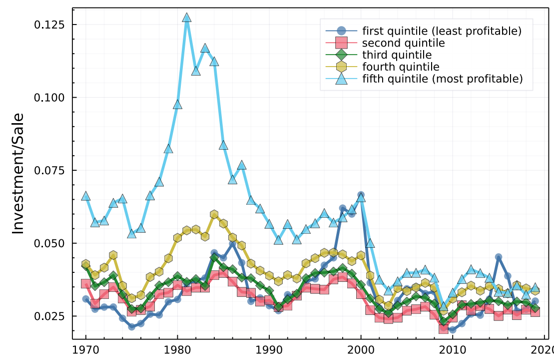
To sum up, compared to their counterparts fifty years ago, firms in the new economy have changed their business model to a large extent. We find that highly profitable firms do not invest or innovate more than others. Nevertheless, they do spend substantially more on customer capital. These findings suggest that the business dynamism may not have declined, as argued in the existing literature. Instead, the stagnation of corporate investment might come from the fact that companies have changed their focus from physical capital to customer capital. As we will see from the following section, such changes is closely related to the rising corporate market power.

3.3.3 Origins of Markup

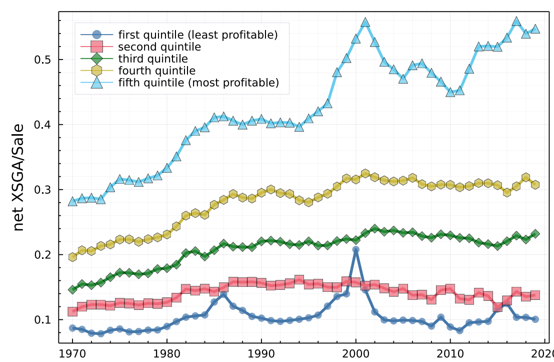
A growing number of studies have documented a substantial increase of average markups in both the U.S. and many other advanced economies (e.g., Nekarda and Ramey, 2013; De Loecker, Eeckhout and Unger, 2020; Eggertsson, Robbins and Wold, 2018; Calligaris, Criscuolo and Marcolin, 2018). These patterns in the data indicate that firms'

Figure 3.7: Changing business model: different percentiles

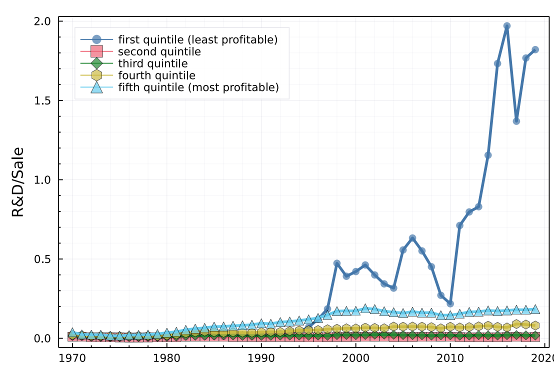
(A) investment



(B) customer capital expenses



(C) R&D expenses



Notes: This figure presents the average investment, customer capital expenses, and R&D expenditures for firms with different gross profitability. In each year, we first classify firms into five different groups according to their gross profitability. After that, we compute the average ratios to sales of investment, customer capital expenses, and R&D expenditures, for each group of companies. Data is obtained from *Compustat*.

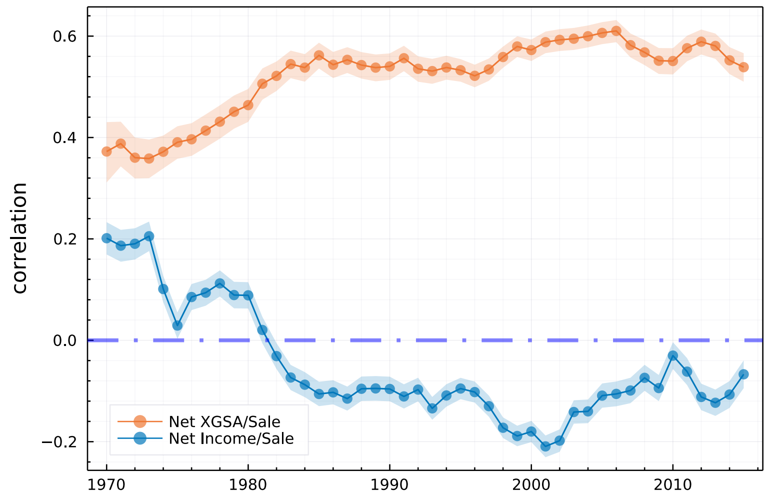
market power has been steadily increasing in today's economy. Meanwhile, many studies attempt to uncover the origins of corporate markup. For instance, Gutierrez and Philippon (2017) focus on the weak competition story, meanwhile Liu, Mian and Sufi (2019) highlight the role of low interest rates in contributing to the rise of market power. Besides, Crouzet and Eberly (2018) and Bessen (2016) focus on the intangible-capital or IT-capital origin of corporate market power, respectively. Finally, Dopper et al. (2021) propose that changing consumer preference could also lead to rising markups because they find that customers have become less sensitive to price over time.

Here we argue that the origins of corporate markup may come from their customer base. Our main idea can be best illustrated in Figure 3.8. Specifically, we compute the firm-level markup and customer capital expenses for all the firms in our sample. Then in each year, we compute the cross-section correlation between these two indicators across different firms. The solid lines are our estimated values, and the shaded areas represent the 95% confidence intervals. The orange line in Figure 3.8 clearly shows that a firm's markup is positively and significantly correlated with its customer capital expenses. In other words, this positive relationship implies that companies with more customer capital expenses have higher markups on average. More importantly, this cross-sectional correlation has been steadily increasing over time, indicating the increasing importance of the customer base in explaining corporate markup.

Similarly, we can also obtain the time-varying correlation between a firm's markup and its net earnings. The blue line in Figure 3.8 shows that the cross-sectional correlation between a firm's net income and its markup has changed from positive to negative. This change in the sign of correlation implies that different from our conventional wisdom, nowadays, firms with more negative net earnings are associated with higher market power. In other words, firms with higher markup are still highly profitable in terms of gross profitability. However, as they have stronger incentives to spend substantial resources on customer capital, their net earnings become negative.

Here we want to give a simple example to explain why we should expect these relationships in the data. Consider a firm that has a new product to sell. Its innovation cost is a fixed cost of f, and the marginal cost of selling it to an additional customer is c. Therefore, if the total number of buyers is q, then given the product price p, the firm's net income is computed as $\pi = pq - f - cq$. Following the standard literature, a firm's markup is defined as its total profits over total costs, which is by definition $\mu \equiv \frac{pq}{f+cq} = \frac{p}{f/q+c}$. Suppose we live in a new economy with a higher fixed cost f and

Figure 3.8: Time-varying correlation between markup and net earnings



Notes: This orange line presents the annual cross-section correlation between markup and customer capital expenses, while the blue line shows the correlation between markup and net earnings. Both customer capital expenses and net earnings are scaled by sales. Firm-level markup is measured by following De Loecker, Eeckhout and Unger (2020)'s approach. Data is obtained from *Compustat*.

nearly zero marginal cost c (De Ridder, 2019). Given the market price p, a firm's markup should be positively related to its customer base q, i.e., μ is increasing in q. In other words, if a firm can increase its customer base by spending more on customer capital, its markup will increase even the price remains unchanged. Meanwhile, its net income will decline due to increased customer capital expenses. As a result, we should observe a positive relationship between a firm's customer capital expenses and its markup in the data, but a negative association between its net income and markup.

Generally speaking, we find that firms with higher markups are more likely to be those with higher customer capital expenses and lower net incomes. One caveat for this conclusion is that the correlation between net income and markup might not be negative forever. This negative sign simply implies that, at this point, many companies are still on their way to becoming superstar firms. Once the industrial concentration has reached certain levels, most firms with large customer bases will have started making positive earnings. In that case, this cross-sectional correlation will likely become positive again.

Finally, we implement some reduced-form fixed-effect regressions to show that our previous conclusion is robust to introducing some additional control variables. The regression results for investigating the association between firm-level markup and customer capital expenses are presented in Table 3.2. The general model specification used in Table 3.2 can be shown as follows:

$$\text{markup}_{i,t} = \alpha + \beta \times \frac{\text{net XSGA}_{i,t}}{\text{sale}_{i,t}} + \Gamma X_{i,t} + \delta_i + \mu_t + \varepsilon_{it}$$

Throughout this part, i and t refer to firm and year, respectively. The variable markup is the firm's estimated markup, and $\frac{\text{net XSGA}_{i,t}}{\text{SALE}_{i,t}}$ here represents our empirical proxy for firm's customer capital expenses. We are primarily interested in the sign and statistical significance of the estimated coefficient β. In addition, X represents a set of firm-level control variables that could affect companies' customer capital expenses. Following the empirical corporate finance literature, we include the return of assets, tangibility, investment, size, profitability, book leverage, dividend payout, cash-to-asset ratio, and Tobin's q. For most columns, we control both firm- and year-fixed effects to account for the unobserved firm and year characteristics, except for the last two columns. All standard errors are clustered at the firm level (or industry level for the last two).

Columns (1) - (9) in Panel A of Table 3.2 present our baseline results using the fixed-effect regression model, with a slight difference in the choices of control variables in each column. In the last three columns, I include all the firm-level control variables. The difference between the last three columns comes from the choices of fixed effects: In column (10), I control for firm and year fixed effects; In column (11), I include 3-digit SIC industry and year fixed effects; Meanwhile, in the last column, I introduce the industry, year, and industry-year fixed effects. Based on the results shown in Table 3.2, we can find that in all specifications, the estimated coefficients of the firm's customer capital expenses are positively significant. In addition, for most of them, the estimated coefficient enters with a positive sign at the 1% significance level. It suggests that companies' markups are significantly and positively associated with their customer capital expenditures. In terms of economic significance, our empirical results in Table 3.2 show that one standard deviation (0.45) increase in customer capital expenditure is associated with a 1.0-2.72 percentage points increase in corporate markup, which is equivalent to an increase by 0.04-0.11 standard deviations. This result implies an economically significant relationship between these two indicators.

Table 3.2: Reduced-form evidence: markup and customer capital expenditure

	markup											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
customer capital expenditure/sale (scaled by 100)	0.047***	0.046***	0.048***	0.048***	0.037***	0.042***	0.047***	0.057***	0.043***	0.037***	0.021**	0.021**
	(5.165)	(5.053)	(5.415)	(5.318)	(4.126)	(4.677)	(5.184)	(5.586)	(4.747)	(3.786)	(2.277)	(2.315)
return of assets		-0.003***								-0.001	-0.000	-0.001
		(-5.104)								(-1.409)	(-0.416)	(-0.700)
tangibility			0.555***							0.712***	0.871***	0.920***
			(59.534)							(61.671)	(77.100)	(81.964)
investment				0.301***						-0.059***	0.047*	0.027
				(20.242)						(-3.272)	(1.957)	(1.122)
size					-0.071***					-0.073***	-0.070***	-0.070***
					(-63.251)					(-54.909)	(-107.122)	(-108.396)
profitability						-0.005***				0.002*	0.002	0.002
						(-6.893)				(1.821)	(1.314)	(1.345)
book leverage							0.000			-0.002***	-0.002***	-0.002***
							(1.374)			(-5.751)	(-3.317)	(-3.758)
payout								0.028***		0.013	0.022*	0.013
								(3.301)		(1.484)	(1.907)	(1.151)
cash/asset									0.126***	0.297***	0.537***	0.526***
									(19.151)	(39.245)	(66.217)	(65.260)
log Tobin's q										-0.012***	0.037***	0.041***
										(-5.994)	(16.535)	(18.489)
	Fixed effects											
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		
Industry (sic3)											Yes	Yes
Industry × Year												Yes
N	126,837	126,832	126,832	125,315	126,832	126,628	125,318	115,334	126,824	97,526	98,823	98,823
Adjusted R^2	0.795	0.796	0.802	0.797	0.802	0.796	0.795	0.798	0.796	0.827	0.508	0.530

Notes: This table presents the association between markup and customer capital expenditure with different fixed-effect model specifications. The dependent variables are corporate markup, and we measure it by following De Loecker, Eeckhout and Unger (2020)'s method. Definitions of customer capital expenditure and all the other control variables are explained in Section 3.2.1. Data used in this table is at firm-year level, and obtained from *Compustat*. In columns (1)-(10), we introduce firm- and year-fixed effects. In column (11), we include industry- and year-fixed effect. In column (12), we use industry-, year-, and industry-year-fixed effects. T-statistics are in parentheses. *, **, and *** represent results significant at the 10%, 5%, and 1% levels, respectively. Standard errors are clustered at the firm level.

3.3.4 Cross-Industry Evidence

Last but not least, we test whether the fraction of firms is higher in industries with higher returns-to-scale. In order to conduct this empirical investigation, first, we need to obtain measures on the industry-level production function. We follow De Ridder (2019)'s methodology to obtain the estimates on fixed cost $\overline{fc}$ and marginal production cost $\overline{mc}$. More specifically, for each year and each industry at the 3-digit SIC level, we use the following two equations for our estimations:

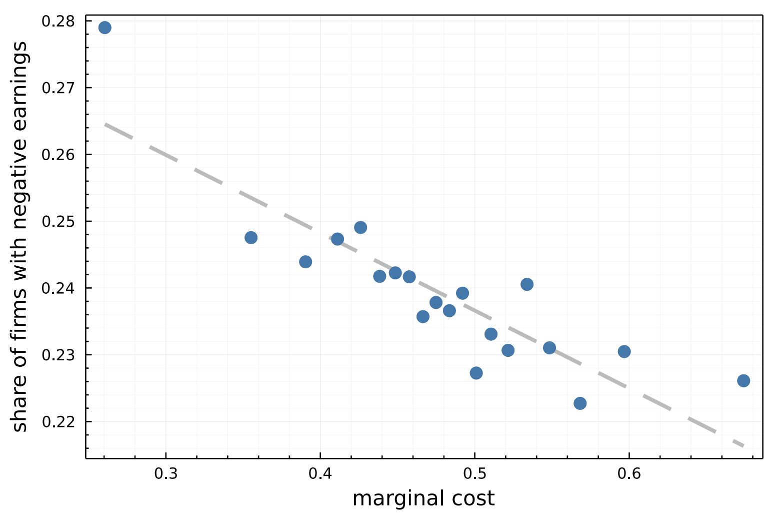
$$\overline{\mathrm{fc}} = \left(1 - \frac{1}{\mathrm{markup}}\right)\mathrm{SALE} - \mathrm{IB} \tag{3.1}$$

$$\overline{\mathrm{mc}} = \frac{\mathrm{COGS} - \overline{\mathrm{fc}}}{\mathrm{SALE}} \tag{3.2}$$

Data source and variable constructions of *SALE*, *IB*, *COGS*, and *markup* are the same as those described in Section 3.2.1. Consistent with other related studies (e.g., De Ridder, 2019; Hoberg and Phillips, 2021; Su, 2021), we also document a secular increase in fixed production cost $\overline{fc}$ and a substantial reduction in the marginal cost of production $\overline{mc}$. In addition, we obtain the industry-year level estimates on the share of firms with negative net earnings as we did in the previous sections.

Figure 3.9 presents the binscatter plot between our industry-level measure of marginal cost $\overline{mc}$ and share of firms with negative net earnings. The gray dash line represents the linear-fit regression. This figure clearly shows a negative and significant relationship between these two variables: industries with lower marginal production costs indeed have a higher fraction of firms with negative net earnings. This significant and negative association in the data supports our previous hypothesis. Companies in industries with relatively low marginal production costs face more economies of scale. As a result, they need to go through a rat race in customer base competition before a small number of them become superstar firms.

Figure 3.9: Binscatter plot between marginal production cost and share of firms with negative net earnings



Notes: This figure presents the binscatter plot between industry-level marginal cost of production and share of firms with negative earnings. The gray dash line represents the linear-fit regression. Specifically, for each year and each industry at the 3-digit SIC level, we obtain the empirical measures on fixed production cost and operating scale by following De Ridder (2019)'s methodology. In addition, for each industry in each year, we count the number of firms with negative profits and divide it by the total number of firms to obtain the industry-level share of firms with negative net earnings. Data is obtained from *Compustat*.

We also report the regression results with different controls and fixed effects in Table 3.3. Based on this table, we can see that this negative and significant association remains robust across various model specifications. In terms of economic significance, our result shows that one standard deviation (0.15) decrease in marginal cost is associated with a 1.47-8.63 percentage points increase in the share of unprofitable firms. The latter is equivalent to an increase by 0.06-0.38 standard deviations. Our empirical finding implies that the close relationship between marginal product cost and the share of unprofitable firms is also economically significant. In other words, the changing economies of scale arising from new technologies such as digitization also transform the corporate business model and the nature of competition between firms.

Table 3.3: Reduced-form evidence: marginal cost and share of firms with negative earnings

	share of firms with negative earnings			
	(1)	(2)	(3)	(4)
marginal cost	-0.304***	-0.098***	-0.575***	-0.115***
	(-23.413)	(-8.619)	(-31.726)	(-6.993)
intercept	0.384***	0.0598***	0.767***	0.357***
	(59.112)	(3.150)	(28.680)	(12.813)
	Fixed effects			
Year	No	Yes	No	Yes
Industry	No	No	Yes	Yes
N	13,769	13,769	13,769	13,769
Adjusted R^2	0.038	0.309	0.215	0.447

Notes: This table presents the association between industry-level marginal cost of production and share of firms with negative earnings with different fixed-effect model specifications. Specifically, for each year and each industry at the 3-digit SIC level, we obtain the empirical measures on fixed production cost and operating scale by following De Ridder (2019)'s methodology. In addition, for each industry in each year, we count the number of firms with negative profits and divide it by the total number of firms to obtain the industry-level share of firms with negative net earnings. Original data used in this table is at firm-year level, and obtained from *Compustat*. T-statistics are in parentheses. *, **, and *** represent results significant at the 10%, 5%, and 1% levels, respectively. Standard errors are clustered at the industry level.

3.4 Conclusion

We document the prevalence of public companies with negative net earnings since the 1970s. We find that the fraction of listed firms with negative net income has increased sharply from 18% in 1970 to 54% in 2019. In addition, such an increase is mainly driven by the right shifts in the mean, i.e., the increasing popularity of sizable firms that are not profitable. We conjecture that the increasing returns-to-scale in the new economy is the main driver behind. We provide three different sets of empirical findings to support our hypothesis. First, earning losses mostly come from the growing customer capital expenses instead of production-related costs, capital investments, or R&D expenditures. Second, cross-sectionally, the net earning to sale ratio is significantly and *negatively* associated with markup: powerful firms tend to have lower net incomes. Third, industries with low marginal production costs have higher percentages of unprofitable companies.

Our paper indicates that the origin of increasing corporate market power documented in De Loecker, Eeckhout and Unger (2020) may come from companies' customer base. In this perspective, regulators should pay serious attention to industries with higher user-switching costs. Once the existing leading firms have successfully built a large customer base, their market power can only be impaired by a sufficiently large innovation.

One caveat is that our main conclusion – the increasing fraction of firms with negative earnings – cannot be generalized to the entire economy with both public and private firms. Based on an unreported exercise, we conducted a similar empirical investigation using the *Orbis* dataset. As widely known, *Orbis* dataset mainly consists of private companies. However, we do not find any apparent trends when investigating the share of firms with negative net earnings. We leave explaining the underlying reason behind this divergence between private and public firms for future research.

Chapter 4

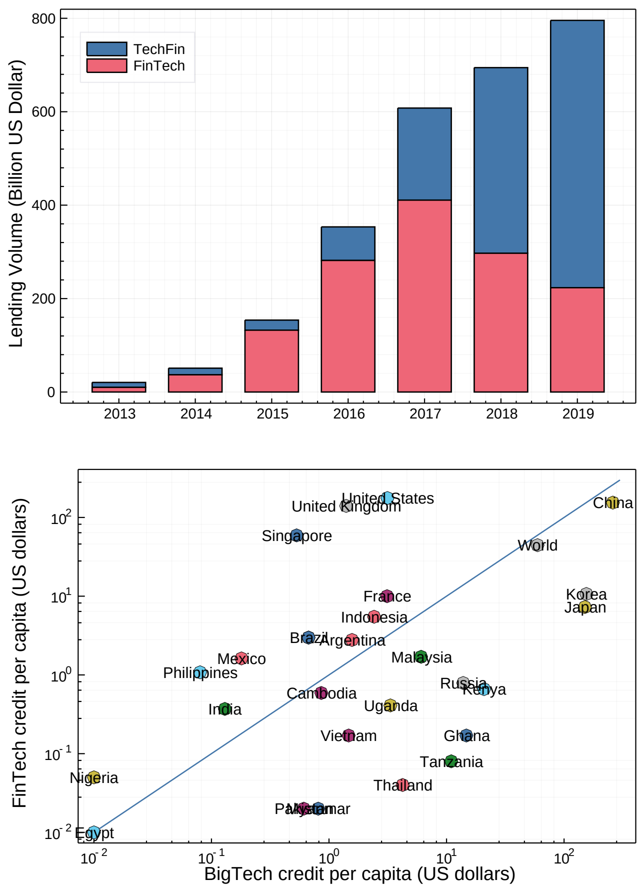
The Macroeconomics of TechFin

Over the past decade, the financial market has seen the arrival of massive new technologies, raising many debates about their consequences. Two most important ones are *FinTech* and *TechFin*.[1] Using the cross-country dataset provided by Cornelli et al. (2020), in the top graph of Figure 4.1, I present the world's total lending volume in billion us dollars for both Fintech and TechFin credits. As we can see, both FinTech and TechFin lendings are becoming increasingly important in the modern financial system. The bottom graph shows the relative importance between them in different countries in 2019. Some countries, such as the United States, the United Kingdom, and Singapore, have more development in FinTech lendings; meanwhile, Asian countries, including China, Korea, and Japan, have better TechFin credit access. Generally speaking, these two new types of financial intermediaries have emerged in a fast pace across different credit markets around the world, which leads to a fast-growing empirical literature on them (e.g., Hau et al., 2018; Tang, 2019).

In this paper, I attempt to investigate, **in theory**, the role of BigTech lending in macroeconomy. More specifically, how should we modify the existing theories of financial intermediation and business cycles so as to accommodate the rise of TechFin? In the existing literature, theories of credit are useful for understanding the mechanism

[1]Throughout this paper, FinTech refers to the situation where financial firms adopt new types of technology, while TechFin means that technology companies provide financial services. Typical examples of FinTech are these digital platforms facilitating peer-to-peer (P2P) lending and borrowing, while examples of TechFin include Ant Group, WeBank, and so on. The broad definition of FinTech provided by the Financial Stability Board is "technologically enabled financial innovation that could result in new business models, applications, processes, or products with an associated material effect on financial markets and institutions, and the provision of financial services."

Figure 4.1: The rise of FinTech and BigTech lending



of business cycles because the credit system acts as a propagation mechanism of first-moment productivity shocks. However, these theories are centered on banks and the key characteristic on bank lending is this collateral-based borrowing constraint. With this financial friction, people find that the aggregate economy has productivity losses in the steady state because the efficient producers cannot borrow enough money (e.g. Moll, 2014). In addition, a financial accelerator mechanism drives the economic fluctuations: small fundamental shocks can be amplified by financial frictions so that they can generate large and persistent fluctuations in aggregate economic activity (Kiyotaki and Moore, 1997; Bernanke and Gertler, 1989).

In this paper, the fundamental difference between a banking sector and a TechFin sector lies in the specific type of borrowing constraints. When BigTech firms lend to the market, they have less demand for collateral and corporate borrowing is subject to an earnings-based borrowing constraint. The reason why I model is related to how BigTech firms reduce asymmetric information or agency problems in the real world. In the traditional banking sector, banks use covenants to mitigate agency frictions. However, for BigTech companies, their technology advantages such as data, algorithms, platform can help reduce these agency frictions. In Section 4.1.4, I will present a simple model to show if the costs of state verification can be significantly reduced with technology, then these BigTech companies will prefer incomplete-collateralized contracts to fully-collateralized ones. In addition, there is indeed empirical support for the assumption of earnings-based borrowing constraint on BigTech lending. For instance, Gambacorta et al. (2020) find that big-tech credit does not correlate with local business conditions and house prices when controlling for demand factors, but that it does react strongly to changes in firm-specific characteristics, such as the transaction volumes and network scores used to calculate firm credit ratings. They argue that the use of technology can allow firms to borrow without any collateral and this new type of borrowing constraint can generate important impacts on macro-finance researches.

After that, I introduce both a banking sector and a TechFin sector into a continuous-time general equilibrium model with heterogeneous entrepreneurs and incomplete markets. These two financial sectors are identical except for the types of borrowing constraints faced by entrepreneurs. Entrepreneurs borrowing from banks are subject to the standard collateral-based borrowing constraints. In contrast, technology advantages allow the big tech companies to resolve agency costs and perform cash flow-based lending. In equilibrium, aggregate productivity is endogenously determined by the net worth share of highly productive firms. Compared to the standard collateral-based borrowing constraint, earnings-based borrowing constraint will amplify the impacts of micro-level uncertainty on net worth inequality by allowing more productive firms to use more leverage and grow faster. As a result, a transitory micro uncertainty shock can lead to persistent changes in finance allocation efficiency and aggregate productivity. This new financial accelerator mechanism, associated with the new TechFin sector, differs from the classic one in three aspects: micro uncertainty instead of aggregate productivity is the primitive shock; financial friction comes from earnings-based borrowing constraints instead of collateral-based ones; and the feedback loops happen between net worth inequality, instead of net worth level, and asset prices. Therefore, the arrival of this new TechFin credit system leads to a higher capital allocative efficiency in the steady state. Moreover, the existence of BigTech lending acts as a propagation mechanism and makes the economy sensitive to both first-moment productivity level shocks and second-moment uncertainty shocks.

Related literature This paper is related to four different branches of literature. First, this paper builds on an extensive literature on financial frictions and business cycles. Two seminal works in this field are Kiyotaki and Moore (1997) and Bernanke and Gertler (1989). Examples of quantitative explorations include Carlstrom and Fuerst (1997), Bernanke, Gertler and Gilchrist (1999), and so on. Recent studies, especially those done after the 2008-2009 global financial crisis, are mainly focused on analyzing the global dynamics and nonlinear effects of shocks with continuoust-time models. Examples include but are not limited to Brunnermeier and Sannikov (2014), Di Tella (2017), He and Krishnamurthy (2013), and Fernandez-Villaverde, Hurtado and Nuno (2019). Besides, Brunnermeier, Eisenbach and Sannikov (2013) provide an excellent and detailed survey on the discrete-time models of macroeconomics and financial frictions, and Brunnermeier and Sannikov (2017) introduce the fundamental tools used in this field.

Second, this paper is closely related to the literature on the macroeconomics of earnings-based borrowing constraint. In terms of empirical evidence, the key finding in Lian and Ma (2021)'s seminal paper is that 80% of the corporate debt value in the US is closely linked to firm's cash flows from their operations instead of the asset liquidation value. Their work intrigues an increasing number of studies that investigate the role of earnings-based borrowing constraint in aggregate fluctuations. For instance, Drechsel (2019) studies the macroeconomic fluctuations through the interaction between earnings-based borrowing constraint and investment shocks. In contrast, Greenwald (2019) mainly focuses on investigating how the transmission of monetary policy shocks differs across firms with different types of covenants.

Third, the basic model framework used in this paper is from the distributional macroeconomics literature, which refers to macroeconomic theories where the relevant state variable is a distribution and the Kolmogorov Forward equation instead of the Euler equation lies at the heart of the analysis. For instance, Moll (2014) studies the impacts of wealth-based borrowing constraints on misallocation and aggregate productivity. Kaplan, Moll and Violante (2018*b*) investigate monetary policy transmission mechanism in a Heterogeneous Agent New Keynesian (HANK) framework. In addition, Fernandez-Villaverde, Hurtado and Nuno (2019) extend the Krusell and Smith (1998) method and explore the relationship between financial frictions and wealth distributions with aggregate shocks. For a concrete introduction to the tools used in this literature, please refer to Achdou et al. (Forthcoming) for details.

Finally, this paper closely relates to the growing literature on the fundamental difference between FinTech and traditional banking sector. Some of the key assumptions used in this paper relies on the empirical findings in the growing FinTech literature. For instance, Gambacorta et al. (2020) empirically show that the lending behaviors of big tech and bank credit are different in terms of their link to collateral value, local economic conditions and firm-specific characteristics. By using the US Peer-to-Peer (P2P) lending data, Tang (2019) finds that FinTech lending works as an complements to bank lending for small-scale loans. Similarly, Cornelli et al. (2020) find that BigTech lendings are complements rather than substitutes to other forms of lending with a cross-country panel dataset for 79 countries during 2013-2019. In addition, Hau et al. (2018) show that the existence of FinTech credit in China improves the credit access condition for firms with lower credit scores.

4.1 Model

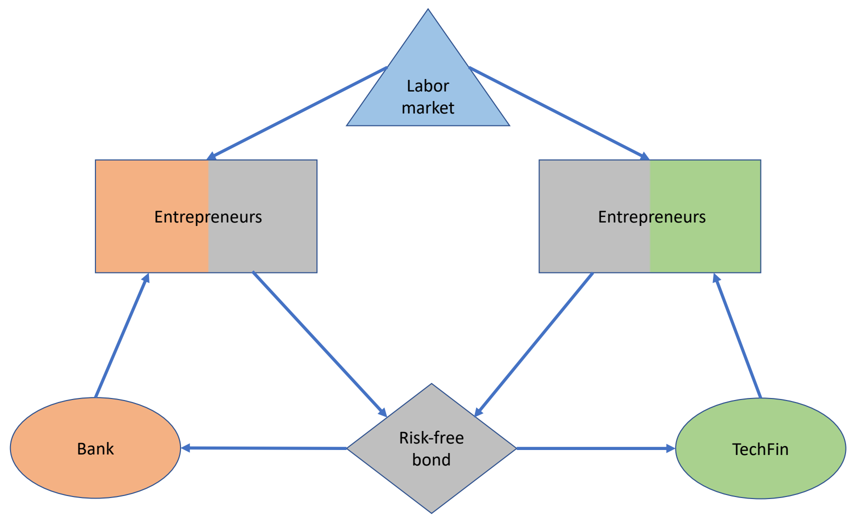
4.1.1 Preference

The model is built upon a standard distributional macro model (e.g., Moll, 2014) but with two different types of financial sectors. Consider a continuous-time infinite horizon economy shown in Figure 4.2 with three different agents: a continuum of entrepreneurs who borrow from the banking sector $\mathcal{B}$, a continuum of entrepreneurs borrowing from the TechFin sector $\mathcal{F}$, and a fixed number $\bar{L}$ of homogeneous workers. Each worker supplies one efficiency unit of labor inelastically. For simplicity, all workers are hand-to-mouth consumers so that we only need keep track of the wealth distributions of entrepreneurs.

Entrepreneurs in these sectors are identical expect that they face different types of borrowing constraints. Each entrepreneur within sector $j \in \{\mathcal{B}, \mathcal{F}\}$ is indexed by their productivity z and wealth a. All the entrepreneurs in this economy have the same additive utility function shown as follows:

$$\mathbb{E}_0 \int_0^\infty e^{-\rho t} \log c_t dt \tag{4.1}$$

Figure 4.2: Summary of economic environment



The choice of this logarithmic utility is purely due to simplification. In this way, the state of the economy can be summarized by the joint distributions of wealth and productivity in two sectors $\left\{\omega_t^{\mathcal{F}}(a,z), \omega_t^{\mathcal{B}}(a,z)\right\}$.

4.1.2 Technology

At time t, each entrepreneur $i \in [0,1]$ in sector $j \in \{\mathcal{B}, \mathcal{F}\}$ owns a private firm that uses both capital k and labor l to produce the final consumption goods with the same production function shown as follows:

$$y_{i,j,t} = (z_{i,j,t} k_{i,j,t})^{\alpha} \, l_{i,j,t}^{1-\alpha} \tag{4.2}$$

The equation above shows that the production technology is in the standard Cobb-Douglas form with parameter α, where $\alpha \in (0,1)$. Entrepreneurs accumulate capital k, and hire homogeneous workers in a competitive labor market at a flat wage rate w_t. At the same time, they can trade in a risk-free bond b subject to a certain type of borrowing constraint. More details on borrowing constraints will be discussed in Section 4.1.4. Capital depreciation rate δ is also the same for both sectors. The reason why I assume that entrepreneurs in both sectors have exactly the same production technology and preference is because the goal of this paper is to investigate the heterogeneous impacts of different types of borrowing constraints. In reality, firms relying on bank financing could have different production technologies from firms using TechFin financing. I denote by $\mathcal{K}_t$ the aggregate units of capital in the economy at time t, and by K_t^j the total capital holdings of entrepreneurs in sector $j \in \{\mathcal{B}, \mathcal{F}\}$.

Following the existing literature (e.g. Moll, 2014; Di Tella, 2017), I assume that the idiosyncratic productivity follows a standard Ornstein–Uhlenbeck process

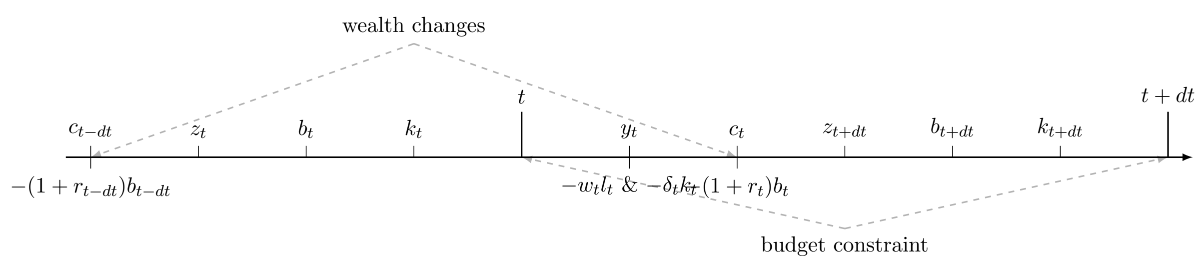
$$dz_{i,j,t} = \frac{1}{\theta}\left(\overline{\mu} - z_{i,j,t}\right) dt + \sigma\sqrt{\frac{1}{\theta}} d\mathcal{W}_{i,j,t} \tag{4.3}$$

In Equation (4.3), $\overline{\mu}$ is the long-run mean level of entrepreneur's productivity. $\mathcal{W}_{i,j,t}$ is the standard exogenous Brownian shock, and it is independent and identically distributed (i.i.d.) across different firms. With the incomplete market assumption, entrepreneurs cannot fully hedge their productivity risk, and σ measures the sensitivity of z to the underlying Brownian shock. Throughout this paper, when investigating how the economy reacts to different types of aggregate shocks, I assume that these economy-wide shocks are "M.I.T. shocks". Following the conventional definition, an "M.I.T. shock" is an unexpected shock that hits an economy at its steady-state, which lead the economy transiting into towards a new one. More specifically, I interpret shocks to $\overline{\mu}$ as shocks to the aggregate productivity, while shocks to σ as shocks to the degree of micro-level uncertainty.

4.1.3 Timeline

This economy's timeline is shown below.

Timeline



This timeline is a little bit different from the conventional one where the borrowing and investment happen before the realization of productivity. In contrast, here I assume that all the people can observe its productivity before they decide their investment and borrowing next period. More specifically, at period t, entrepreneurs can first observe their productivity z_t, then they issue debt b_t to finance their capital investment $x_t = k_t - (1-\delta)\, k_{t-dt}$ to obtain a new capital stock of k_t. After that, the entrepreneurs hire l_t workers and produce y_t. After paying wages to the workers, the entrepreneurs decide how much to consume c_t and how much wealth a_{t+dt} to save to the next period $t+dt$. This way of modelling follows the work of Kiyotaki (1998) and make earnings-based borrowing risk-free so that we can fully investigate the impacts of uncertainty on allocation efficiency and business cycles without taking into consideration corporate default decisions.

According to the timeline of this economy, the budget constraint from time t to time $t+dt$ of any individual entrepreneur can be shown as follows:

$$c_t + k_{t+dt} - (1-\delta)\, k_t + (1+r_t)\, b_t + w_t l_t = y_t + b_{t+dt} \tag{4.4}$$

Entrepreneurs' wealth is defined as the difference between his capital holdings and debt borrowings, i.e., $a_{i,j,t} \equiv k_{i,j,t} - b_{i,j,t}$. Therefore, the changes in wealth can be computed as follows

$$da_t = a_{t+dt} - a_t = [y_t - w_t l_t - \delta k_t - r_t b_t - c_t]\, dt \tag{4.5}$$

According to Equation (4.5), changes in wealth from t to $t+dt$ are from the following items: current output, wage payments, depreciated capital, interest payments of debt, and consumption. More importantly, at time $t+dt$, z_{t+dt} and a_{t+dt} are state variables. Meanwhile b_{t+dt} and k_{t+dt} are endogenously determined by entrepreneurs. In other words, after observing his productivity z_{t+dt}, the entrepreneur will optimally allocate their wealth a_{t+dt} into capital k_{t+dt} and bond b_{t+dt} to maximize the profits at the next period.

From now on, for the simplification of notations, I will suppress the agent and time subscripts unless it is necessary.

4.1.4 Borrowing Constraints

In this section, I explain the difference in borrowing constraint faced by entrepreneurs when they borrow from different financial institutions. Generally speaking, the type of borrowing constraint in the banking sector is simply the standard collateral-based one, while that in the TechFin sector is modeled as the new earnings-based borrowing constraint as documented in some recent empirical works (e.g. Lian and Ma, 2021; Gambacorta et al., 2020). After describing the model setup, I provide its micro-foundation in Section 4.1.4 and discuss the similarity and difference between them in Section 4.1.4.

Banking sector

To begin with, I assume that all the entrepreneurs face the same collateral-based borrowing constraint when they borrowing from the traditional banking sector:

$$(1+r)b \leq \lambda_{\mathcal{B}} k \tag{4.6}$$

where $0 \leq \lambda_{\mathcal{B}} \leq 1$. Equation (4.6) shows that due to issues such as limited enforcement or asymmetric information, only a fraction $\lambda_{\mathcal{B}}$ of entrepreneur's capital stock can be externally financed. The level of $\lambda_{\mathcal{B}}$ represents the severeness of these frictions. More specifically, if $\lambda_{\mathcal{B}} = 0$, Equation (4.6) means the entrepreneurs can only self-financing their capital investment. At the same time, if $\lambda_{\mathcal{B}} = 1$, it means that all the capital stock can be externally financed. Both situations are extreme cases and we will investigate how the magnitudes of $\lambda_{\mathcal{B}}$ affects the role of banking sector in driving the macroeconomic fluctuations.

Rewriting this equation with the definition of wealth a can give us the standard wealth-based borrowing constraint shown as follows:

$$b \leq \frac{\lambda_B}{1+r-\lambda_B} a \tag{4.7}$$

Similarly, Equation (4.7) captures the common intuition that the amount of capital available to an entrepreneur is limited by his personal wealth a and again the magnitude of $\lambda_{\mathcal{B}}$ captures the degree of financial development in the banking system. As Equation (4.7) is more comparable to the borrowing constraint in the TechFin sector, therefore I will use this equation throughout the rest of the paper.

TechFin sector

Of course, the precise modelling of TechFin depends on the interpretation on the fundamental difference between a traditional banking sector and this new financial sector. In this paper, I assume that entrepreneurs can borrow against their future earnings instead of the current collateral values. Therefore, in this paper's perspective, the fundamental difference between banking and the new TechFin sector is that TechFin sector allows the entrepreneurs to borrow against their future earnings.

More specifically, I assume that all the entrepreneurs in the TechFin sector face the same future earnings-based borrowing constraint shown as follows:

$$(1+r)b \leq \lambda_{\mathcal{F}}\pi \tag{4.8}$$

where π is entrepreneur's earnings and $\lambda_{\mathcal{F}} \leq 1$. Whether the earnings are current or in next period depends on how to interpret the model. On one hand, it could be interpreted as future earnings as firms first borrow and then produce to make money. On the other hand, earnings could also be interpreted as current as financing and production decisions happen after firms can observe their productivity. In this paper, I follow Lian and Ma (2021)'s work and assume that only a fraction $\lambda_{\mathcal{F}}$ of *current* earnings can be externally financed. Again, the existence of $\lambda_{\mathcal{F}}$ also comes from the limited enforcement that entrepreneurs might steal a fraction of their companies' earnings. $\lambda_{\mathcal{F}} = 0$ refers to the situation where entrepreneurs can only self-financing, and $\lambda_{\mathcal{F}} = 1$ means that all earnings can be externally financed. Rewriting this equation can give us the wealth-based borrowing constraint in TechFin sector as follows:

$$b \leq \frac{\lambda_F \xi z}{1+r-\lambda_F \xi z} a \tag{4.9}$$

where $\xi = \alpha\left(\frac{1-\alpha}{w}\right)^{\frac{1-\alpha}{\alpha}}$. One technical issue with this earnings-based borrowing constraint is that even for a reasonable level of $\lambda_{\mathcal{F}}$, if the firm's productivity is too large, then the maximum amount of borrowing could be infinite. To solve this technical issue, when numerically solving the model, I will impose an *ad-hoc* up boundary $\bar{\phi}$ for this earnings-based borrowing constraint, i.e.,

$$k \leq \begin{cases} \frac{a}{1-\lambda_{\mathcal{F}}\xi z} & z < \left(1-\frac{1}{\bar{\phi}}\right)\frac{1}{\lambda_{\mathcal{F}}\xi} \\ \bar{\phi}a & z \geq \underline{z} \end{cases} \tag{4.10}$$

For simplicity, from now on, I will still use equation (4.9) for characterizing the equilibrium and the optimal decisions. At the same time, when numerically solving the model, I will impose (4.10) and an ad hoc choice of $\bar{\phi}$.

One caveat is that in the model the only difference between banking sector and FinTech sector lies in the type of borrowing constraints. However, it does not mean that in reality this is the only difference between these two sectors. As pointed out in some recent review papers (e.g. Boot et al., Forthcoming; Thakor, 2020; Huang et al., 2020), what's new about FinTech could be technological innovations in both information processing and communication channels, or some new entrants providing non-intermediated financial services, or some big technology firms provide lending services to small and medium entrepreneurs with no collaterals. For example, with the US loan-level data on mortgage applications and origination, Fuster et al. (2019) show that FinTech lenders originate mortgages faster and screen borrowers more effectively compared to other lenders. Philippon (2016) suggest that FinTech can lower the costs of financial services provided by financial intermediations. Thakor and Merton (2019) have developed a theory of bank and non-bank lending in which banks have an endogenous advantage over non-bank lenders when it comes to being trusted to make good loans because banks possess an advantage in developing investor trust due to their unique access to low-cost deposit funding. However, the reason why I focus on this specific characteristic is because in the existing macro-finance literature, firm's borrowing constraint is essential to the macroeconomic analysis of financial frictions. In addition, there is indeed empirical support for the assumption of earnings-based borrowing constraint on BigTech lending. For instance, Gambacorta et al. (2020) find that big-tech credit does not correlate with local business conditions and house prices when controlling for demand factors, but that it does react strongly to changes in firm-specific characteristics, such as the transaction volumes and network scores used to calculate firm credit ratings. The use of technology can allow firms to borrow without any collateral. The type of borrowing constraints can have important impacts on macro-finance mechanisms.

Microfoundation

There is an extensive number of theoretical papers discussing what determines corporate borrowings. For instance, Stiglitz and Weiss (1981) and Holmstrom and Tirole (1997) point out that corporate debt capacity should be determined by corporate earnings. In contrast, Hart and Moore (1994), Kiyotaki and Moore (1997), and Bernanke and Gertler (1989) argue that corporate borrowing should be tightly linked to the asset liquidation value. Given the fact that there has been a lot of possible explanations in the literature, here I lay out a simple model to explain why we should expect the coexistence of two types of borrowing constraints in reality. Generally speaking, the coexistence could either come from the heterogeneous information technology advantages on the lender's side or the different uses of intangible capital on the borrower's side.

Basic framework Let us consider a revised example in Bernanke and Gertler (1989) or originally from Townsend (1979) to show that some lenders will strictly prefer cash flow-based over asset-based lending while others do the exact opposite.

I assume that entrepreneurs with a capital stock of k need to borrow money from lenders for some investment projects. I denote the amount of money that entrepreneurs can borrow as b. For simplicity, both entrepreneurs and lenders are risk-neutral and entrepreneurs have linear preference in their consumption. The lender's opportunity cost is fixed to be r and the capital's unit liquidation value is assumed to be l. As we will see later, one crucial assumption on intangible capital is that it has a relatively low value of l. For simplicity, here I do not distinguish between the asset's actual resalability and the value damage from agency frictions such as managers' business stealing behaviors. In theory, it is highly possible that firms with more intangibles have lower liquidation values simple because they face severer agency problems. But here I do not distinguish between these two different origins.

There are two possible project outcomes. Entrepreneur's earnings could be $z_G k$ with a probability of p, or they could be $z_B k$ with a probability of $1 - p$. Without loss of generality, we assume that $z_G > z_B > l > 0$.

Investors can choose between two types of lending. The first one is what Bernanke and Gertler (1989) call full-collateralization contract, which means that the entrepreneur's net worth is sufficiently large that he is able to pay lenders their require return even in the worst state, where the lenders seize his capital and resell it in the market:

$$(1 + r)\, b \leq lk \tag{4.11}$$

One advantage of using this type of lending is that lenders do not need to do any earnings verification. Instead, lenders use a contract linked to the liquidation value of capital, which gives us the standard collateral-based borrowing constraint.

The second way of lending is to secure the ownership instead of the collateral. This is the incomplete collateralization case. The focus here is to investigate the characteristics of an optimal contract, which can be stated mathematically as follows:

$$\max_{\{q, c_G, c_B, \tilde{c}_B\}} pc_G + (1 - p)\left[qc_B + (1 - q)\, \tilde{c}_B\right] \tag{4.12}$$

subject to the following constraints

$$
\begin{aligned}
(1+r)\,b &\leq p\,(z_G k - c_G) + (1-p)\left[z_B k - q\,(c_B + f) - (1-q)\,\tilde{c}_B\right] && (4.13)\\
c_G &\geq (1-q)\left[(z_G - z_B)\,k + c_B\right] && (4.14)\\
c_G, c_B, \tilde{c}_B &\geq 0 && (4.15)\\
0 \leq q &\leq 1 && (4.16)
\end{aligned}
$$

In the equations above, c_G means entrepreneur's consumption when he announces the good state. $\tilde{c}_B$ represents entrepreneur's consumption when he announces the bad state and lenders choose not to verify while c_B represents entrepreneur's consumption when he announces the bad state and lenders choose to verify with a cost of f. As we will discuss later, some lenders such as these big tech companies have a relatively low cost of verification, which is crucial for their preferred choice on cash flow-based lending. Equation (4.13) represents the participation constraint and (4.14) is the incentive constraint. The last two equations are feasibility constraints. The optimal contract $\{p, c_G, c_B, \tilde{c}_B\}$ maximizes the entrepreneur's expected consumption in Equation (4.12) subject to constraints (4.13) to (4.16).

Now I provide two different stories on why some lenders prefer cash-flow-based over asset-based lending while others do not. The intuition is the following. From profit-maximization perspective, lenders always prefer earnings-based lending. However, due to the cost of state verification or cash flow pledgeability problems, collateral is useful for resolving these agency costs. At the same time, if lenders can find ways to reduce the cost of state verification or increase cash flow pledgeability, then they will strictly prefer cash flow-based lending.

Information asymmetry story The first result I want to show is that for lenders with low cost of state verification, they strictly prefer using cash flow-based lending. I introduce the following assumption.

Assumption 4.1 *Informational technology advantages allow some lenders such as BigTech firms to reduce the cost of state verification.*

The key assumption in this information asymmetry story is that some lenders has Technology advantages in monitoring and predicting firms future earnings while others do not. One great example are BigTech firms and traditional banks. For instance, Thakor (2020) also point out that the use of the Blockchain technology and many other technological advancement in the TechFin sector can lead to the significant reduction in the cost of verification.[2] In reality, *Ant Group* uses *Alipay* system to help their lending because *Alipay* allows it to easily verify the cash flows of companies with a very low cost.

For these lenders with information technology advantage, they will strictly prefer cash flow-based lending. I summarize the main result in the following lemma.

Lemma 4.1 *If the cost of state verification f is lower than some threshold f^*, then it is more attractive for lenders to implement cash flows-based lending instead of asset-based lending.*

Detailed proof can be found in the appendix. The basic idea here is that, if lenders manage to overcome such problems with big data or platform advantages, they will use cash flow-based lending. As pointed out in some recent review papers (e.g. Boot et al., Forthcoming; Thakor, 2020; Huang et al., 2020), one of the most important new characteristics about FinTech is that some big technology firms provide lending services to small and medium entrepreneurs with no collateral.

The theoretical result in Lemma 4.1 is also consistent with some empirical findings in the existing literature. For instance, Gambacorta et al. (2020) find that big-tech credit does not correlate with local business conditions and house prices when controlling for demand factors, but that it does react strongly to changes in firm-specific characteristics, such as the transaction volumes and network scores used to calculate firm credit ratings. The use of technology can allow firms to borrow without any collateral. They also argue this new type of borrowing constraints can have important impacts on macro-finance mechanisms.

Intangible capital story An alternative way of understanding the different uses of lending is based on the difference between intangible capital and tangible capital. I assume that intangible capital has a low liquidation value.

Assumption 4.2 *Intangible capital has low liquidation value l.*

[2] The other possible changes mentioned in Thakor (2020) are the reduced search costs of matching transacting parties, increasing economics of scale in gathering and using large data, and cheaper and more secure information transmission.

A business is typically liquidated as part of a bankruptcy process. During this process, tangible assets such as machines and plants are sold quickly and their value can be easily evaluated. In practice, when investors evaluate the liquidation value of companies, they often exclude the intangible assets as the value of intangible investments is difficult to assess by outsiders. Our assumption here is consistent with the existing literature that assumes liquidation cost is decreasing in investment tangibility (e.g., Beck et al., 2020).

With Assumption 4.2, we can easily show that if the asset tangibility is lower than some threshold, it is optimal for lenders to impose a cash flow-based lending.

Lemma 4.2 *If the liquidation value l is lower than some threshold l^*, then it is more attractive for lenders to implement cash flow-based lending instead of asset-based lending.*

The detailed proof can be shown in the appendix. The intuition behind Lemma 4.2 is that if the nature of assets is not collaterizable, then it is better for lenders to use cash flow-based lending as it is more profitable to do that. Our result on the relationship between intangible capital and cash flow-based lending is also consistent with the existing literature. For example, Haskel and Westlake (2018) find that the financing of intangibles is less related to traditional forms of credit constraints. Our result in Lemma 4.2 rationalizes this empirical finding in the data.

Similarity and difference between two types of borrowing constraints

debt capacity and net worth In the existing literature, some papers argue that these two types of borrowing constraints are fundamentally different because earning is a concept of flows while collateral is a stock variable. However, here I argue that if we investigate the relationship between debt capacity and net worth, then there is some similarity between these two borrowing constraints. More specifically, based on the previous model setup, by the time entrepreneurs need to repay their debt at $t + dt$, earnings at t have already become part of the net worth at $t + dt$. Therefore, both borrowing constraints link debt capacity to verifiable net worth as follows:

$$\textbf{debt capacity} = \phi \times \textbf{verifiable net worth}$$

Therefore, if entrepreneurs' net worth can be observed, then these two types of borrowing constraints are similar in the sense that they impose a requirement that the maximum amount of debt entrepreneurs can borrow is a fraction of their net worth.

cross-sectional difference Comparing equations (4.7) and (4.9), we can clearly see that the fundamental difference between these two types of borrowing constraints lies in their cross-sectional difference. For the entrepreneurs faced with collateral-based borrowing constraint, given their wealth, productive firms do not face any additional advantages because the tightness of borrowing constraint is the same for all entrepreneurs. However, as for entrepreneurs with earnings-based borrowing constraint, given their wealth, productive firms have additional advantages because the tightness of constraints is decreasing in productivity. As we will see in the following sections, such a difference is crucial for our understanding in the different macroeconomic implications of these two borrowing constraints.

4.1.5 Equilibrium Definition

The equilibrium definition here can be summarized as in Definition 4.1.

Definition 4.1 *A stationary recursive competitive equilibrium consists of prices* $\{r_t, w_t\}_{t=0}^{\infty}$ *and allocations* $\left\{(l_{t,i,j}, k_{t,i,j}, b_{t,i,j}, c_{t,i,j})_{i\in[0,1],j\in\{\mathcal{B},\mathcal{F}\}}\right\}_{t=0}^{\infty}$ *that satisfy the following conditions:*

1. ***Optimization**: given market prices* $\{r_t, w_t\}_{t=0}^{\infty}$*, resource allocations*

$$\left\{(l_{t,i,j}, k_{t,i,j}, b_{t,i,j}, c_{t,i,j})_{i\in[0,1],j\in\{\mathcal{B},\mathcal{F}\}}\right\}_{t=0}^{\infty}$$

   maximize each entrepreneur's life-time utility (4.1) subject to constraints (4.2), (4.4), (4.7), (4.9), and initial endowment $(k_{0,i,j}, a_{0,i,j})$*.*

2. ***Market clearance**: market prices* $\{r_t, w_t\}_{t=0}^{\infty}$ *satisfy the following conditions*

   - *labor market at any time t*

$$\iint l_t^{\mathcal{B}}(a,z)\,\omega_t^{\mathcal{B}}(a,z)dadz + \iint l_t^{\mathcal{F}}(a,z)\,\omega_t^{\mathcal{F}}(a,z)dadz = \bar{L} \tag{4.17}$$

   - *bond market at any time t*

$$\iint b_t^{\mathcal{B}}(a,z)\,\omega_t^{\mathcal{B}}(a,z)dadz + \iint b_t^{\mathcal{F}}(a,z)\,\omega_t^{\mathcal{F}}(a,z)dadz = 0 \tag{4.18}$$

- *final goods market at any time t*

$$\mathcal{C}_t^{\mathcal{F}} + \mathcal{C}_t^{\mathcal{B}} + \mathcal{C}_t^{\mathcal{L}} + \mathcal{X}_t^{\mathcal{F}} + \mathcal{X}_t^{\mathcal{B}} = \mathcal{Y}_t^{\mathcal{F}} + \mathcal{Y}_t^{\mathcal{F}} \tag{4.19}$$

  where

$$\begin{aligned}
\mathcal{C}_t^j &= \iint c_t^j(a,z)\,\omega_t^j(a,z) dadz,\ j \in \{\mathcal{B},\mathcal{F}\} \\
\mathcal{C}_t^{\mathcal{L}} &= w_t \bar{L} \\
\mathcal{X}_t^j &= \iint x_t^j(a,z)\,\omega_t^j(a,z)\,dadz,\ x = k' - (1-\delta)\,k \text{ and } j \in \{\mathcal{B},\mathcal{F}\} \\
\mathcal{Y}_t^j &= \iint y_t^j(a,z)\,\omega_t^j(a,z)\,dadz,\ j \in \{\mathcal{B},\mathcal{F}\}
\end{aligned}$$

3. ***Stationary distribution****: the wealth distributions in two sectors* $\left\{\omega_t^{\mathcal{F}}(a,z), \omega_t^{\mathcal{B}}(a,z)\right\}$ *obey entrepreneur's optimal decision and the exogenous productivity process (4.3), and they are stationary, i.e.,* $\frac{\partial \omega_t^{\mathcal{B}}(a,z)}{\partial t} = 0$ *and* $\frac{\partial \omega_t^{\mathcal{F}}(a,z)}{\partial t} = 0$.

4.1.6 Preliminary Analysis

Individual optimal decisions

To begin with, I first characterize the optimal policy functions for each individual entrepreneur. The optimal decisions of bond holdings and capital investment can be summarized as in Lemma 4.3.

Lemma 4.3 *Given the market prices r and w, there is a same productivity cutoff for being active* $\underline{z}$ *for entrepreneurs in both sectors. The optimal capital and debt holdings for entrepreneurs in the banking sector are*

$$\begin{aligned}
b_{\mathcal{B}}(a,z) &= \begin{cases} \frac{\lambda_{\mathcal{B}} a}{1+r-\lambda_{\mathcal{B}}} & z \geq \underline{z} \\ -a & z < \underline{z} \end{cases} \\
k_{\mathcal{B}}(a,z) &= \begin{cases} \frac{(1+r)a}{1+r-\lambda_{\mathcal{B}}} & z \geq \underline{z} \\ 0 & z < \underline{z} \end{cases}
\end{aligned}$$

Meanwhile, the optimal capital and debt holdings for entrepreneurs in the TechFin sector are

$$
\begin{aligned}
b_{\mathcal{F}}(a,z) &= \begin{cases} \frac{\lambda_{\mathcal{F}}\xi z a}{1+r-\lambda_{\mathcal{F}}\xi z} & z \geq \underline{z} \\ -a & z < \underline{z} \end{cases} \\
k_{\mathcal{F}}(a,z) &= \begin{cases} \frac{(1+r)a}{1+r-\lambda_{\mathcal{F}}\xi z} & z \geq \underline{z} \\ 0 & z < \underline{z} \end{cases}
\end{aligned}
$$

where $\underline{z} = \frac{r+\delta}{\xi}$ *and* $\xi = \alpha\left(\frac{1-\alpha}{w}\right)^{\frac{1-\alpha}{\alpha}}$.

Given our assumption on the constant return-to-scale technology and frictionless labor market, the marginal product of capital for any individual entrepreneur with productivity z is always proportional to $z-\delta$. If the current interest rate on the market is r, then the optimal capital choice is a corner solution: it is zero for entrepreneurs with productivity lower than some threshold $\underline{z}$, and the maximal amount of borrowing for entrepreneurs with productivity higher than $\underline{z}$. At the same time, these inactive entrepreneurs will lend all their wealth to the market so that they can get a constant return r. The cutoff $\underline{z}$ is the same for two sectors simply because we assume the production technology in these two sectors is the same.

Again, one important observation from Lemma 4.3 is that compared to entrepreneurs in the traditional banking sector, productive firms borrowing from BigTech companies have additional advantages in lending and capital accumulation. As shown in the following lemma, these advantages eventually reflect on the wealth evolution dynamics.

Lemma 4.4 *With preference assumption (4.1), entrepreneur's wealth a in both sectors evolves according to the following equations:*

$$
\begin{aligned}
da_{\mathcal{B}} &= \left\{1_{z\geq\underline{z}} \times \left[\frac{(1+r)(\xi z - r - \delta)}{1+r-\lambda_{\mathcal{B}}} + r - \rho\right] + 1_{z<\underline{z}} \times (r-\rho)\right\} a_{\mathcal{B}}dt \equiv \Gamma^{\mathcal{B}}(z)\, a_{\mathcal{B}}dt \\
da_{\mathcal{F}} &= \left\{1_{z\geq\underline{z}} \times \left[\frac{(1+r)(\xi z - r - \delta)}{1+r-\lambda_{\mathcal{F}}\xi z} + r - \rho\right] + 1_{z<\underline{z}} \times (r-\rho)\right\} a_{\mathcal{F}}dt \equiv \Gamma^{\mathcal{F}}(z)\, a_{\mathcal{F}}dt
\end{aligned}
$$

For simplicity, we will write them as

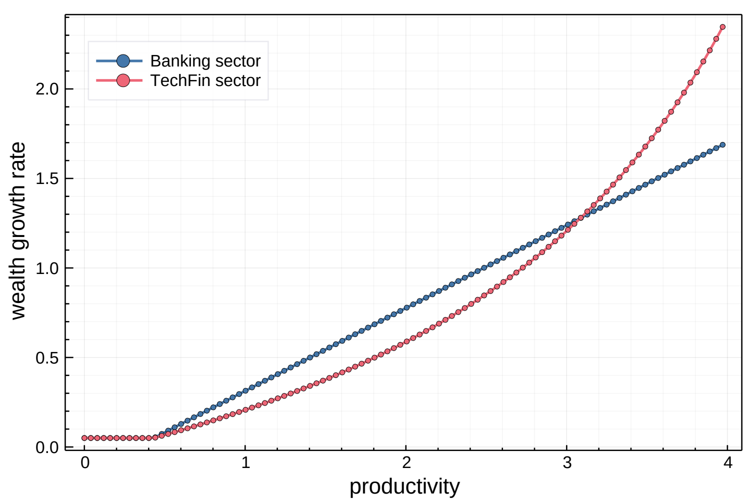
$$
da_j = \Gamma_t^j(z)\, a_j dt \tag{4.20}
$$

where Γ *is the wealth growth rate function, and it depends on entrepreneur's idiosyncratic productivity* z *and the sector* $j \in \{\mathcal{B}, \mathcal{F}\}$ *he belongs to.*

As we can see from the proof in the appendix, with the assumption of log-utility, the optimal consumption choice will always be a constant ρ fraction of wealth, where ρ is the time value. Therefore, wealth growth rate of firms with lower productivity in two sectors is always $r - \rho$ because these firms are not producing anything on the market. Instead they will lend all their wealth to these productive entrepreneurs, as a result, they will get a constant rate of return r before determining their consumption.

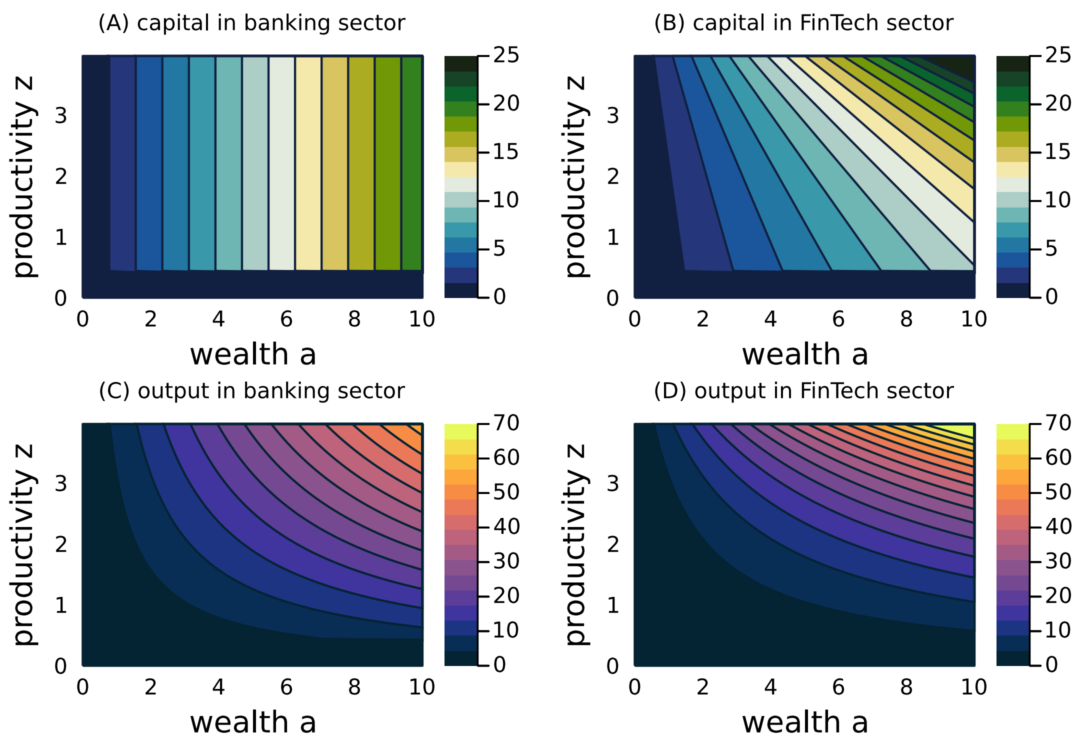
What is interesting here is that the wealth growth rate is different for productive firms in different sectors. For any active producer borrowing from the banking sector, his wealth growth rate is $(\xi z - r - \delta)\,\lambda_{\mathcal{B}} + r - \rho$, which is a linear function in z. In contrast, wealth growth rate of any active firm with productivity z borrowing from the TechFin sector is $\frac{\lambda_{\mathcal{F}}(\xi z - r - \delta)}{\lambda_{\mathcal{F}} - (\lambda_{\mathcal{F}} - 1)\xi z} + r - \rho$, which is a convex function of z. Such difference can be better illustrated in Figure 4.3. As we will see it later, the convexity of wealth growth rate in the TechFin sector is the underlying reason why uncertainty, i.e., the second-moment shocks, matters for the aggregate quantities over the business cycles. The degree of $\lambda_{\mathcal{F}}$ determines the convexity of this relationship, which as a result affects the impacts of uncertainty.

Figure 4.3: Wealth growth rate in Banking and TechFin sector: an example



Besides, the different types of borrowing constraints will also lead to difference in distributions of capital holdings and outputs in these two sectors. In the top two graphs in Figure 4.4, I present the capital holdings for firms with different productivity and wealth. If the entrepreneur's productivity is lower than the threshold, then firms will not produce anything therefore the capital holdings are zeros for these low-productivity firms. Once firms become productive, then the equilibrium capital holdings are different in these two sectors. In the banking sector, entrepreneurs face wealth-based borrowing constraint, therefore the capital holdings only depend on wealth and do not depend on productivity at all. In contrast, in the TechFin sector, capital holdings are increasing in both wealth and productivity, which makes the sectoral capital is concentrated on firms with highest wealth and productivity. Since optimal output is linear in capital, as we can see from the bottom two graphs in Figure 4.4, we can achieve the same conclusion for firms output.

Figure 4.4: Capital and output in Banking and TechFin sector: an example



Dynamics of wealth distributions

After discussing the optimal policy function for any individual entrepreneur, now I turn to characterize how the wealth distributions in both sectors evolve over time. With the exogenous productivity process (4.3) and the endogenous wealth process (4.20), the wealth distribution evolves according to the following equations:

$$\frac{\partial \omega_t^j(a,z)}{\partial t} = -\frac{\partial\left[\Gamma_t^j(z)\, a\omega_t^j(a,z)\right]}{\partial a} - \frac{\partial\left[\frac{1}{\theta}(\overline{\mu}-z)\,\omega_t^j(a,z)\right]}{\partial z} + \frac{1}{2}\frac{\partial^2\left[\frac{1}{\theta}\sigma^2\omega_t^j(a,z)\right]}{\partial z^2} \quad \text{where } j \in \{\mathcal{B},\mathcal{F}\} \tag{4.21}$$

Generally speaking, what determines the evolution of wealth distribution in this economy is a system of high-dimensional partial differential equations (PDEs). Previous works, including Kiyotaki (1998), Caselli and Gennaioli (2013), and Moll (2014), use wealth shares to characterize aggregates so that we can save one state variable. However, this method is not applicable here as we have two different sectors in this economy. Scaling the wealth by using aggregate capital and getting wealth share cannot reduce the number of state variables. Therefore, I follow Raissi, Perdikaris and Karniadakis (2019) and use Physics-informed neural network (PINN) approach to numerically solve the dynamics of ω_j. This deep learning method utilizes the advantages of deep neural networks to solve high-dimensional PDEs and it has a significantly reduced time and memory costs compared to those classical methods such as finite difference and finite element. The advantage of deep learning approach lies in the fact that the algorithm does not require interpolation and coordinate transformation because it is universal nonlinear approximators (Bach, 2017) and thus avoids the curse of dimensionality. In addition, it can overcome the local optimization problem by introducing some penalty factors or stochasticity into the loss function. Other possible solving methods include adaptive sparse grids approach by Brumm and Scheidegger (2017) and some different neural network approaches proposed by Fernandez-Villaverde et al. (2020) and Chen, Didisheim and Scheidegger (2021).

4.2 Two Types of Financial Accelerators

In this section, I turn to investigating model implications with numerical exercises. The focus is to investigate the impacts of different types of borrowing constraints on both steady-state allocative efficiency and business cycles.

4.2.1 Parameterization

The choice on values of different parameters is shown in Table 4.1. Following the existing literature, I set the rate of time preference to be 0.05, capital share to be 0.33, and the labor market size normalized to be 1.0. Using the U.S. Bureau of Economic Analysis (BEA) Fixed Assets Tables dataset, I compute the average capital depreciation rate to be 6%. Based on Moll (2012), the random death rate of entrepreneurs is picked to be 0.05. Following Asker, Collard-Wexler and De Loecker (2014), in the baseline model, I set $\overline{\mu}$ to be zero, the persistence of productivity to be 0.85, and the idiosyncratic productivity to be 0.56. The choices of these parameters related to the underlying productivity are consistent with the actual firm-level TFP measure in the data. In addition, the ad hoc choice on the upper boundary for corporate leverage is assigned to be 10. However, the precise choice on this parameter does not matter significantly for the model outcomes.

Table 4.1: Parameterization

Parameter	**Description**	**Value**	**Source/Reference**
ρ	rate of time preference	0.05	
α	capital share	0.33	Moll (2014)
$\overline{L}$	labor market size	1.0	
δ	capital depreciation rate	0.06	BEA-FAT
χ	death rate	0.05	Moll (2012)
$\overline{\mu}$	log idiosyncratic productivity mean	0.0	
θ	autocorrelation $e^{-\theta}$	0.16 (corr = 0.85)	Asker, Collard-Wexler and De Loecker (2014)
σ	log idiosyncratic productivity s.d.	0.56	
$\bar{\phi}$	upper boundary for corporate leverage	10.0	

Throughout this section, I will discuss how various choices on the tightness of constraints in banking $\lambda_{\mathcal{B}}$, the tightness of constraints in TechFin $\lambda_{\mathcal{F}}$, degrees of first-moment shocks $\Delta\bar{\mu}$, and degrees of second-moment shocks $\Delta\sigma$, affect our final conclusions.

4.2.2 Earnings-based Borrowing Constraint as the Financial Accelerator of Micro-uncertainty

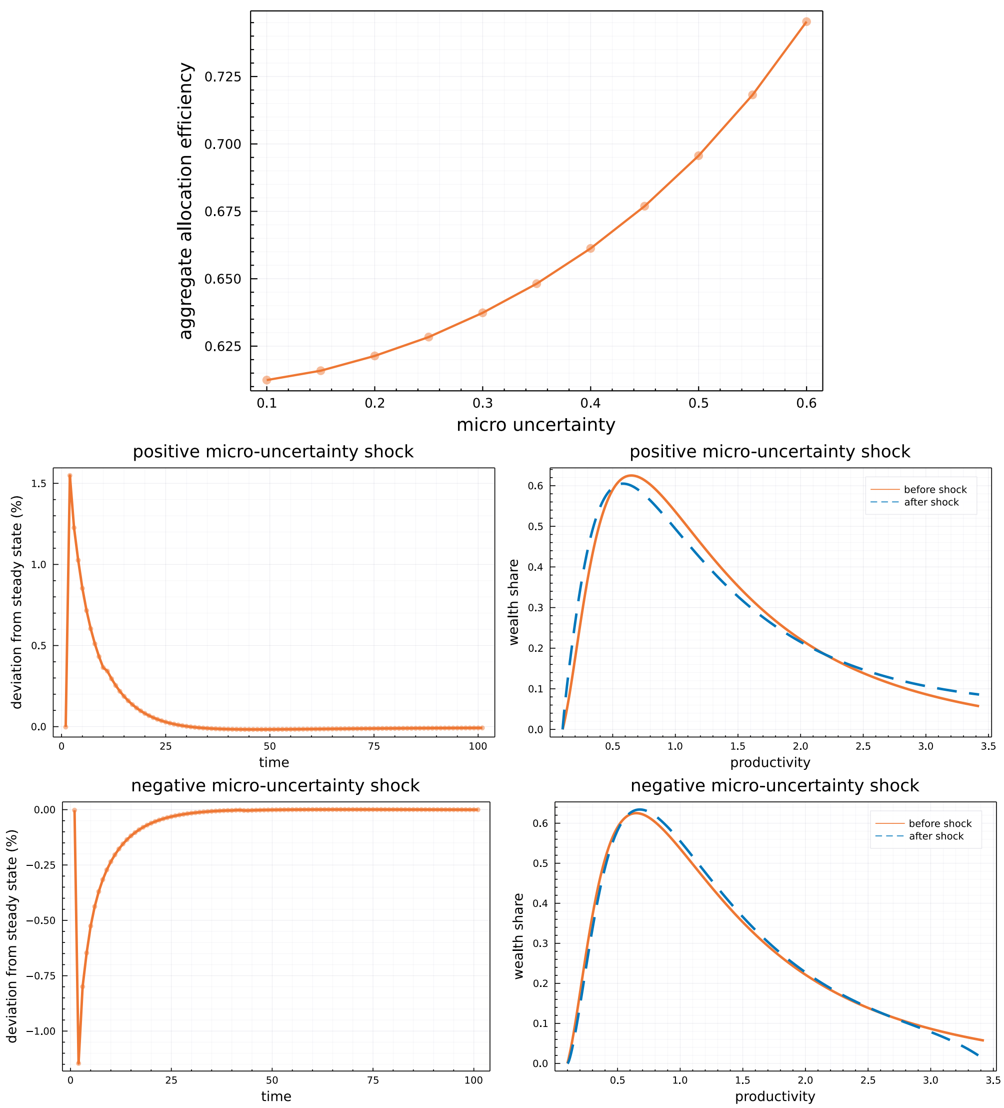
Before investigating the macroeconomic implications of the co-existence of two types of financial sectors, I start with discussing the interesting amplification mechanism between micro-uncertainty and earnings-based borrowing constraint. As mentioned before, one important characteristic of this earnings-based borrowing constraint is that it will generate an asymmetric net worth growth rate: for unproductive firms, their wealth growth rate is always the market interest rate; meanwhile, for productive firms, their wealth rate is increasing in their productivity. As a result, compared to the standard collateral-based borrowing constraint, earnings-based borrowing constraint can help productive firms quickly build up their net worth, and this interesting feature will make the TechFin sector sensitive to the underlying second-moment shocks.

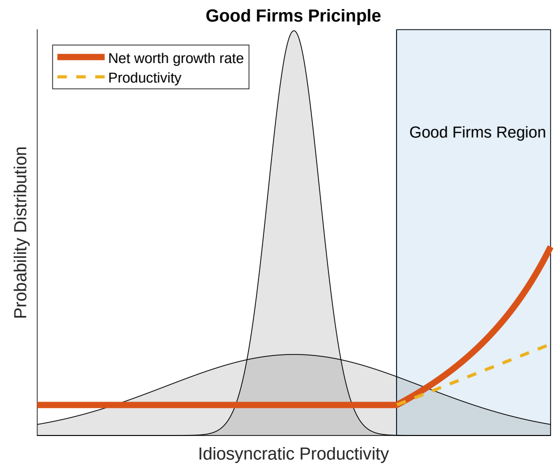
There are three key implications from our numerical exercises in this part. First, within the TechFin sector, aggregate allocative efficiency is increasing in the dispersion of underlying firm-level productivity. As shown in Graph (A) in Figure 4.5, the aggregate allocative efficiency is higher when the micro-level uncertainty is larger. This conclusion seems counter-intuitive at first, but the underlying mechanism can be easily summarized as "good-firms principle" shown as in the following figure:

In an economy with a higher degree of micro-level uncertainty, it means that there exists more firms that are highly unproductive and productive, i.e., the productivity dispersion is significantly wider. However, unproductive entrepreneurs do not matter for the aggregate economy because they choose to not operate and lend their wealth to the productive entrepreneurs. With the help of earnings-based borrowing constraint, the highly productive entrepreneurs can increase their wealth faster because they get to use more leverage. As a result, there will be more active and highly productive entrepreneurs in this economy when dispersion is higher. Therefore, an increase in micro uncertainty leads to higher total outputs and allocative efficiency because the good firms become more important. I name this mechanism "good firms principle" to echo "good news principle" proposed by Bernanke (1983). Good news principle in Bernanke (1983) means that only good news matters in growth options because bad news will not be proceeded further. With earnings-based borrowing constraint, what is essential for an economy is the number of superstar firms and their relative importance. Therefore, an increased uncertainty will generate positive outcomes on the aggregate productivity.

Second, there is a feedback loop between micro-uncertainty and earnings-based borrowing constraint. As shown in Graph (B) in Figure 4.5, when there is a positive or negative shock to micro-level uncertainty, it generates amplified and persistent fluctuations to the aggregate productivity. This outcome is similar to the classical financial accelerator mechanism as shown in Bernanke and Gertler (1989) and Kiyotaki and Moore (1997). The underlying mechanism for this new financial accelerator mechanism comes from the fact that the shocks to micro-level uncertainty also changes the wealth share of firms with different productivity in this economy. For example, a positive shock to micro-level uncertainty, combined with the asymmetric wealth growth rate characteristic of earnings-based borrowing constraint, leads to the increasing importance of

Figure 4.5: Micro-uncertainty and earnings-based borrowing constraint: a new type of financial accelerator





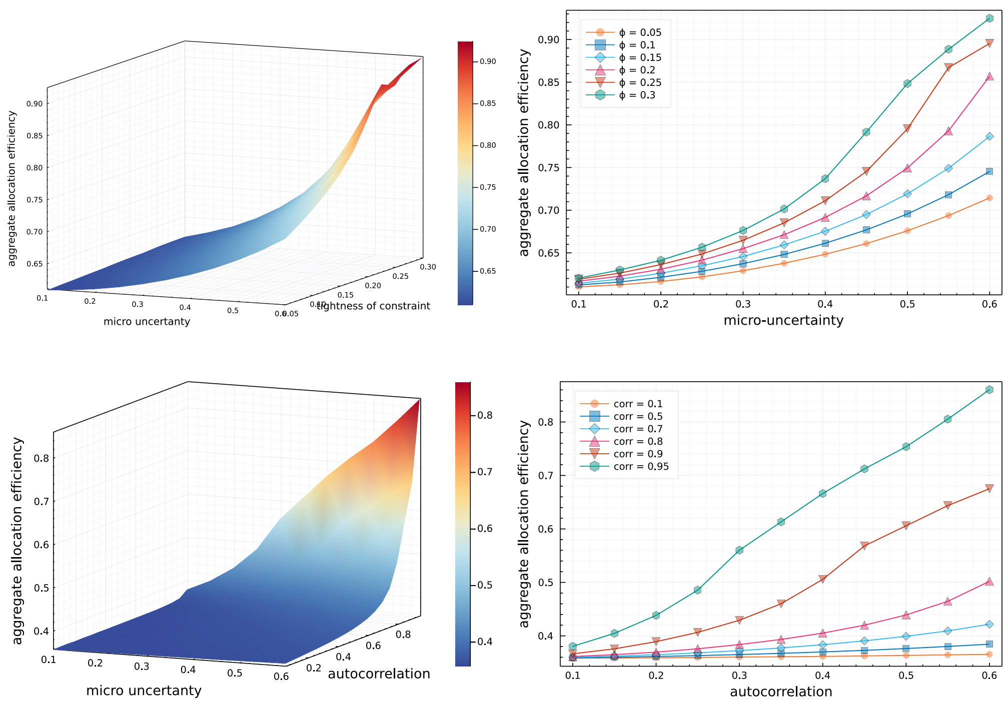
large firms in this economy. In this way, earnings-based borrowing constraint amplifies the impacts of uncertainty on net worth inequality by allowing more productive firms to use more leverage and grow faster. This effect is persistent as these changes on net worth inequality are persistent over time. As a result, a transitory micro-level uncertainty shock can lead to persistent changes in finance allocation efficiency and aggregate productivity. This new financial accelerator mechanism differs from the classic one in three aspects: micro uncertainty instead of aggregate productivity is the primitive shock; financial friction comes from earnings-based borrowing constraint instead of collateral-based borrowing constraints; and the feedback loops happen between net worth inequality, instead of net worth level, and asset prices.

Third, the strength of this new financial accelerator mechanism crucially depends on the tightness of the borrowing constraint and also the persistence of underlying productivity process. As shown in Figure 4.6 and 4.7, the importance of this new type of financial accelerator mechanism in driving the economic fluctuations is larger when the borrowing constraint is more slack and the underlying productivity process is more persistent. This result shows up again because their impacts on the wealth share of the highly productive firms in this economy.

4.2.3 A Macroeconomic Model with Two Financial Sectors

Now I turn to investigate the model implications with two types of financial sectors. For better comparison and discussion, I include one benchmark economy with a banking

Figure 4.6: Determinants of aggregate allocation efficiency

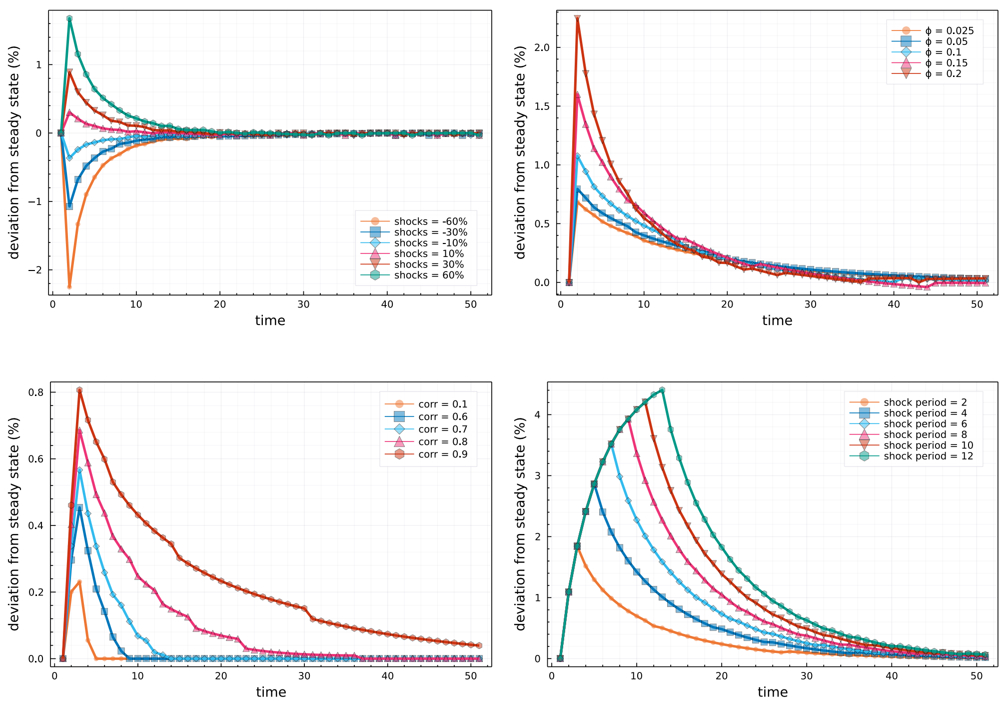


sector only, i.e., $\lambda_{\mathcal{F}} = 0$. In all the following exercises, I report the behavior of all variables relative to their steady-state values.

Impacts on aggregate allocative efficiency

The impacts of the development of TechFin sector on aggregate allocative efficiency can be summarized in Figure 4.8. As the TechFin sector develops, the aggregate allocative efficiency is increasing in the economy. More importantly, compared to the development of traditional banking sector, the availability of this new type of financial intermediaries is more essential to reducing the wedges between marginal products of capital. The underlying mechanism is the following. In an idea world with a perfectly well-functioning credit market, only the entrepreneur with the highest productivity should be operating in equilibrium. With financial frictions, we can no longer achieve this first-best outcome as some productive firms are financially constrained so their marginal products of capital are higher than the market average. However, there is a fundamental difference

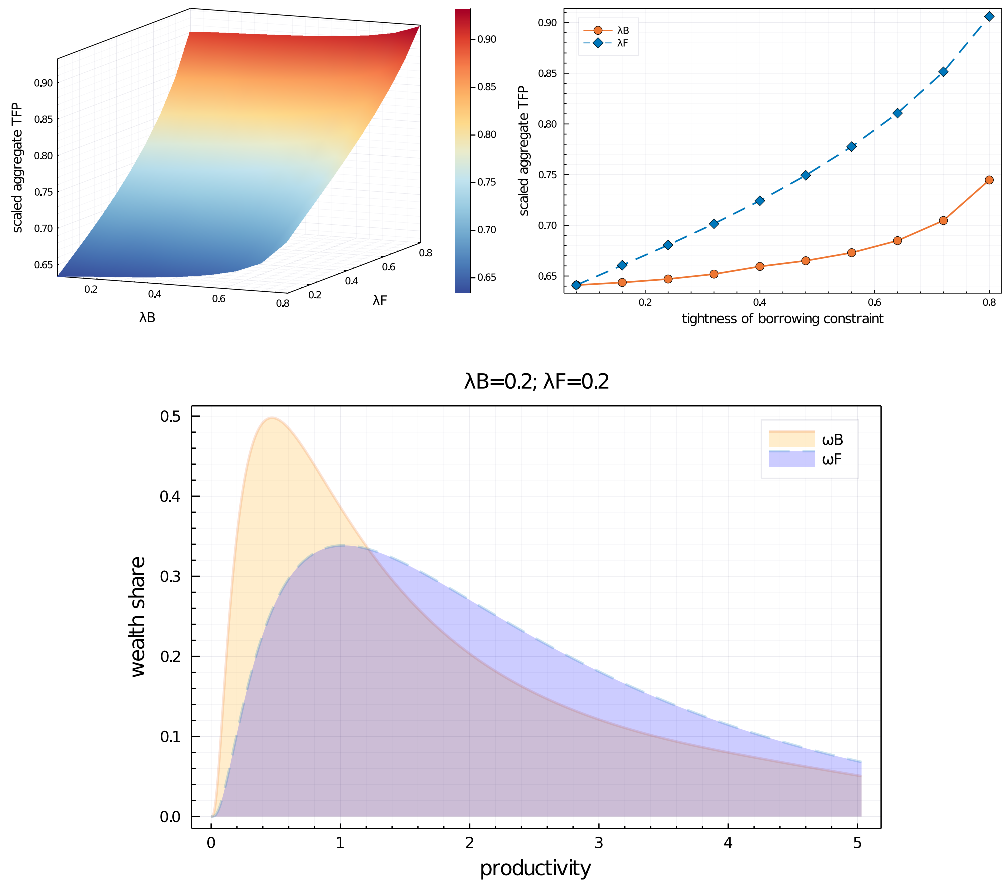
Figure 4.7: Determinants of business cycle fluctuations



between the collateral-based borrowing constraint in banking sector and the earnings-based borrowing constraint in TechFin sector. In banking sector, the maximum amount of debt can be borrowed is linked to capital stock. Meanwhile, in TechFin sector, the upper limit of debt financing is directly related to productivity. Therefore, compared to the traditional banking sector, the *de facto* tightness of borrowing constraints for highly productive firms are looser for BigTech lendings. Thus, the equilibrium wealth is more concentrated towards productive firms, and the degree of capital misallocation is lower in TechFin sector.

The result here has some important policy implications. It is widely common that underdeveloped countries have underdeveloped financial markets. The existence of financial frictions will affect the accumulation of capital and wealth, which eventually slows the economic growth rate in these emerging countries. However, if the superstar firms in these countries, especially those Tech Giants, can provide financial services to other firms, then it has the potential to narrow down the differences in per capita income. Based on our numerical exercises here, the technology advantage of these superstar firms

Figure 4.8: A macroeconomic model with two financial sectors: aggregate allocative efficiency

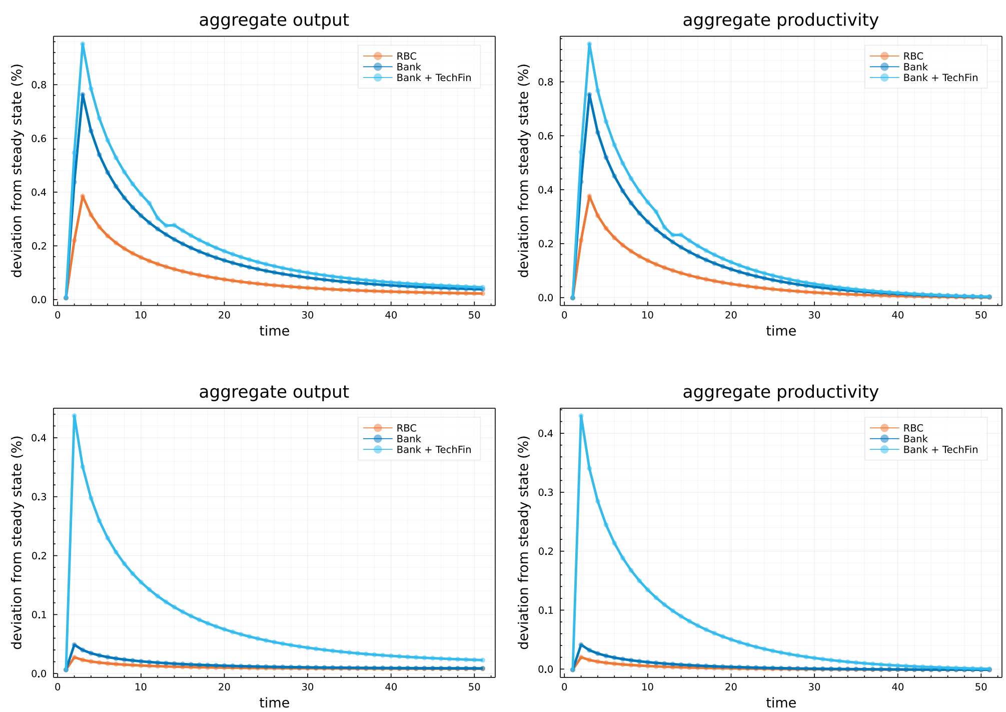


can work better to improve the aggregate capital allocative efficiency, compared to the traditional banking sector.

Impacts on economic fluctuations

For investigating the implications on business cycles, I conduct two different experiments. For both experiments, I compute the impulse responses from three different models: a standard real business cycle model with no financial frictions, a model with banking sector only, and a model with both banking and TechFin sectors. As the steady-state values are also different across these models, I report the behaviors of all variables relative to their steady-state values.

Figure 4.9: A macroeconomic model with two financial sectors: business cycles



The first experiment I consider is a one-time shock to the level of (aggregate) productivity. This shock is the standard one in the traditional real business cycle models. Although this is one-time shock, because technology is autocorrelated, productivity will stay above/below trend for several periods. Graph (A) in Figure 4.9 presents the economy's response to a one-time 1% increase in the productivity level. As we can see from this graph, in the frictionless real business cycle model, capital structure is indeterminate and irrelevant to real economic outcomes. Therefore, the decline in the need for external finance has no negligible amplified impacts on the aggregate economy. In contrast, in a model with collateral-based borrowing constraint, increases in aggregate productivity level. It is consistent with the classical financial accelerator literature: when credit markets have asymmetric information and agency problems, the role of credit-market frictions are important drivers of cyclical economic fluctuations. What's interesting here is that the magnitude of these effects is about the same for an economy with two different types of financial sectors. A different type of borrowing constraint will not weaken this mechanism as the link between external financing cost and the net worth of borrowers still exists for earnings-based borrowing constraint. A higher borrower net worth reduces the agency costs of financing, and this mechanism holds for both types of borrowing constraints. Therefore, if the underlying fundamental shock is from the productivity level, then the earnings-based borrowing constraint magnifies and propagates the technology shocks in a similar way as the collateral-based one. Compared to the frictionless benchmark economy, Graph (A) in Figure 4.9 shows that the dynamics in the early periods are quite different because of the role of net worth (inequality). In addition, the hump shape in investment and the reverse hump shape in consumption indicate the unique role of borrowing constraints in driving business cycles.

The real difference between these two types of borrowing constraints shows up when the fundamental shock is from the second-moment uncertainty. The second experiment I consider is a one-time shock to micro-level uncertainty. This experiment is useful for considering the impacts of various shocks to the economy that only affect the degree of uncertainty. As we can see from Graph (B) in Figure 4.9, a positive shock to micro-level uncertainty has essentially no effect in the frictionless benchmark and the economy with a banking sector only, but has both significant impact and propagation effects when there exists a TechFin sector with earnings-based borrowing constraint. The transfer of wealth inequality drives up the demand for investment, initiating a positive feedback loop. The substantial persistence comes from the slow decay of entrepreneurial net worth inequality. The introduction of a TechFin sector raises the possibility that relatively small changes in micro-level uncertainty could generate significant economic fluctuations. In contrast, such characteristic is negligible in the frictionless benchmark economy, as well as the economy with only a banking sector, as this feature of asymmetric wealth growth rate is a special characteristic of TechFin.

The results here are closely related to several papers documenting the real effects of second-moment shocks (e.g. Bloom, 2009; Bloom et al., 2018; Alfaro, Bloom and Lin, 2019). However, there exists a fundamental difference. Indeed, as the TechFin sector develops, the whole economy is becoming more sensitive to uncertainty shocks, but in a positive relationship. In contrast, the standard uncertainty literature documents the negative impacts of uncertainty on the real economy. The reason why such a difference exists comes from the different underlying mechanisms. In this paper, the mechanism is feedback loop between earnings-based borrowing constraint and micro-level uncertainty. In contrast, in their works, the mechanism comes from the fact that is higher uncertainty causes firms to temporarily pause their investment and hiring, which leads to a decline in productivity growth.

4.3 Conclusion

This paper investigates the role of a TechFin sector in driving macroeconomic fluctuations. I introduce both a traditional financial sector and a TechFin sector into a general equilibrium model with heterogeneous entrepreneurs and incomplete markets. These two financial sectors are identical expect for the types of borrowing constraints faced by entrepreneurs: entrepreneurs borrowing from the banking sector are subject to the standard collateral-based borrowing constraints, while those borrowing from the TechFin sector are subject to the earnings-based borrowing constraints. I use a deep learning neural network approach to obtain global solutions, and the main conclusions are twofold. First, this new TechFin credit system leads to a higher capital allocative efficiency in the steady state. Second, the existence of BigTech lending acts as a propagation mechanism and makes the economy sensitive to both first-moment productivity level shocks and second-moment uncertainty shocks.

References

Acemoglu, Daron, Vasco Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. "The Network Origins of Aggregate Fluctuations." *Econometrica*, 80(5): 1977–2016.

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. Forthcoming. "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach." Review of Economic Studies.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica*, 83(6): 2411–2451.

Aiyagari, S. Rao. 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving." *Quarterly Journal of Economics*, 109(3): 659–684.

Alexander, Lewis, and Janice Eberly. 2018. "Investment Hollowing Out." *IMF Economic Review*, 66(1): 5–30.

Alfaro, Iván, Nick Bloom, and Xiaoji Lin. 2019. "The Finance Uncertainty Multiplier." Unpublished Working Paper.

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Dynamic Inputs and Resource (Mis)Allocation." *Journal of Political Economy*, 122(5): 1013–1063.

Atkeson, Andrew, and Ariel Burstein. 2019. "Aggregate Implications of Innovation Policy." *Journal of Political Economy*, 127(6): 2625–2683.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics*, 135(2): 645–709.

Bach, Francis. 2017. "Breaking the Curse of Dimensionality with Convex Neural Networks." *Journal of Machine Learning Research*, 18: 1–53.

Bagwell, Kyle. 2007. "The Economic Analysis of Advertising." In *Handbook of Industrial Organization.* , ed. Mark Armstrong and Robert Porter, 1701–1844. Amsterdam:North-Holland.

Bates, Thomas W., Kathleen M. Kahle, and Rene M. Stulz. 2009*a*. "Why Do U.S. Firms Hold So Much More Cash than They Used To?" *Journal of Finance*, 64(5): 1985–2021.

Bates, Thomas W., Kathleen M. Kahle, and Rene M. Stulz. 2009*b*. "Why Do U.S. Firms Hold So Much More Cash Than They Used To?" *Journal of Finance*, 64(5): 1985–2021.

Beck, Thorsten, Robin Döttling, Thomas Lambert, and Mathijs A. Van Dijk. 2020. "Liquidity Creation, Investment, and Growth." Unpublished working paper.

Benhabib, Jess, Jesse Perla, and Christopher Tonetti. 2019. "Reconciling Models of Diffusion and Innovation: A Theory of the Productivity Distribution and Technology Frontier." NBER Working Paper No. 23095.

Bernanke, Ben, and Mark Gertler. 1989. "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review*, 79(1): 14–31.

Bernanke, Ben S. 1983. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression." *American Economic Review*, 73(3): 257–276.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist. 1999. "The financial accelerator in a quantitative business cycle framework." In *Handbook of Macroeconomics.* Vol. 1, , ed. John B. Taylor and Michael Woodford, Chapter 21, 1341–1393. Elsevier.

Bessen, James E. 2016. "Accounting for Rising Corporate Profits: Intangibles or Regulatory Rents?" Boston University School of Law, Law and Economics Research Paper No. 16-18.

Bils, Mark, Peter J. Klenow, and Cian Ruane. 2021. "Misallocation or Mismeasurement?"

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica*, 77(3): 623–685.

Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J. Terry. 2018. "Really Uncertain Business Cycles." *Econometrica*, 86(3): 1031–1065.

Bolton, Patrick, Hui Chen, and Neng Wang. 2011. "A Unified Theory of Tobin's q, Corporate Investment, Financing, and Risk Management." *Journal of Finance*, 66(5): 1545–1578.

Boot, Arnoud, Peter Hoffmann, Luc Laeven, and Lev Ratnovski. Forthcoming. "Fintech: what's old, what's new?" *Journal of Financial Stability.*

Brander, James A., and Tracy R. Lewis. 1986. "Oligopoly and Financial Structure: The Limited Liability Effect." *American Economic Review*, 76(5): 956–970.

Brumm, Johannes, and Simon Scheidegger. 2017. "Using Adaptive Sparse Grids to Solve High-Dimensional Dynamic Models." *Journal of Machine Learning Research*, 85(5): 1575–1612.

Brunnermeier, Markus K., and Yuliy Sannikov. 2014. "A Macroeconomic Model with a Financial Sector." *American Economic Review*, 104(2): 379–421.

Brunnermeier, Markus K., and Yuliy Sannikov. 2017. "Macro, Money and Finance: A Continuous-Time Approach." Handbook of Macroeconomics.

Brunnermeier, Markus, Thomas Eisenbach, and Yuliy Sannikov. 2013. "Macroeconomics with financial frictions: A survey." Advances in Economics and Econometrics.

Buera, Francisco J., and Yongseok Shin. 2013. "Financial Frictions and the Persistence of History: A Quantitative Exploration." *Journal of Political Economy*, 121(2): 221–272.

Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin. 2011. "Finance and Development: A Tale of Two Sectors." *American Economic Review*, 101(5): 1964–2002.

Bussiere, Matthieu, Laurent Ferrara, and Juliana Yael Milovich. 2015. "Explaining the Recent Slump in Investment: the Role of Expected Demand and Uncertainty." Banque de France Working Paper 571.

Caballero, Ricardo J., Emmanuel Farhi, and Pierre-Olivier Gourinchas. 2017. "Rents, Technical Change, and Risk Premia: Accounting for Secular Trends in Interest Rates, Returns on Capital, Earning Yields, and Factor Shares." *American Economic Review*, 107(5): 614–620.

Calligaris, Sara, Chiara Criscuolo, and Luca Marcolin. 2018. "Mark-ups in the Digital Era." OECD Working Paper 2018-10.

Carlstrom, Charles T., and Timothy S. Fuerst. 1997. "Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis." *American Economic Review*, 87(5): 893–910.

Caselli, Francesco, and James Feyrer. 2007. "The Marginal Product of Capital." *Quarterly Journal of Economics*, 122(2): 535–568.

Caselli, Francesco, and Nicola Gennaioli. 2013. "Dynastic Management." *Economic Inquiry*, 51(1): 971–996.

Cavenaile, Laurent, and Pau Roldan-Blanco. 2021. "Advertising, Innovation, and Economic Growth." *American Economic Journal: Macroeconomics*, 13(3): 251–303.

Chen, Hui, Antoine Didisheim, and Simon Scheidegger. 2021. "Deep Structural Estimation: With an Application to Option Pricing."

Coase, Ronald. 1937. "The Nature of the Firm." *Economica*, 4(16): 386–405.

Cornelli, Giulio, Jon Frost, Leonardo Gambacorta, Raghavendra Rau, Robert Wardrop, and Tania Ziegler. 2020. "Fintech and big tech credit: a new database." BIS Working Papers No. 887.

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross. 1985. "A Theory of the Term Structure of Interest Rates." *Econometrica*, 53(2): 385–407.

Crouzet, Nicolas, and Janice C. Eberly. 2018. "Understanding Weak Capital Investment: the Role of Market Concentration and Intangibles." NBER Working Paper 25869.

Crouzet, Nicolas, and Janice Eberly. 2020. "Rents and Intangible Capital: A Q+ Framework."

Dambra, Michael, Laura Casares Field, and Matthew T. Gustafson. 2015. "The JOBS Act and IPO volume: Evidence that disclosure costs affect the IPO decision." *Journal of Financial Economics*, 116(1): 121–143.

David, Joel, Lukas Schmid, and David Zeke. 2019. "Risk-Adjusted Capital Allocation and Misallocation."

Davis, Carter, Alexandre Sollaci, and James Traina. 2021. "Profit Puzzles or: Public Firm Profits Have Fallen." Tuck School of Business Working Paper No. 3746660.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics*, 135: 561–644.

De Loecker, Jan, Jan Eeckhout, and Simon Mongey. 2021. "Quantifying Market Power and Business Dynamism."

De Ridder, Maarten. 2019. "Market Power and Innovation in the Intangible Economy."

Dinlersoz, Emin M., and Mehmet Yorukoglu. 2012. "Information and Industry Dynamics." *American Economic Review*, 102(2): 884–913.

Di Tella, Sebastian. 2017. "Uncertainty Shocks and Balance Sheet Recessions." *Journal of Political Economy*, 125(6): 2038–2081.

Dixit, Avinash. 1993. *The Art of Smooth Pasting.* Routledge.

Doidge, Craig, G. Andrew Karolyi, and Rene M. Stulz. 2017. "The U.S. listing gap." *Journal of Financial Economics*, 123(3): 464–487.

Doidge, Craig, Kathleen M. Kahle, G. Andrew Karolyi, and Rene M. Stulz. 2018. "Eclipse of the Public Corporation or Eclipse of the Public Markets?" *Journal of Applied Corporate Finance*, 30(1): 8–16.

Dopper, Hendrik, Alexander MacKay, Nathan H. Miller, and Joel Stiebale. 2021. "Rising Markups and the Role of Consumer Preferences."

Dou, Winston Wei, and Yan Ji. Forthcoming. "External Financing and Customer Capital: A Financial Theory of Markups." *Management Science*.

Dou, Winston Wei, Yan Ji, and Wei Wu. 2021. "Competition, Profitability, and Discount Rates." *Journal of Financial Economics*, 140: 582–620.

Dou, Winston Wei, Yan Ji, and Wei Wu. Forthcoming. "The Oligopoly Lucas Tree." *Review of Financial Studies*.

Drechsel, Thomas. 2019. "Earnings-Based Borrowing Constraints and Macroeconomic Fluctuations." Unpublished working paper.

Duffie, Darrell, and Larry G. Epstein. 1992. "Stochastic Differential Utility." *Econometrica*, 60(2): 353–394.

Dumas, Bernard. 1991. "Super contact and related optimality conditions." *Journal of Economic Dynamics and Control*, 15: 675–685.

Eckbo, B. Espen, and Markus Lithell. 2021. "Merger-Driven Listing Dynamics."

Eggertsson, Gauti B., Jacob A. Robbins, and Ella Getz Wold. 2018. "Kaldor and Piketty's Facts: The Rise of Monopoly Power in the United States." NBER Working Paper No. 24287.

Ewens, Michael, and Joan Farre-Mensa. 2020. "The Deregulation of the Private Equity Markets and the Decline in IPOs." *Review of Financial Studies*, 33(12): 5463–5509.

Ewens, Michael, Kairong Xiao, and Ting Xu. 2021. "Regulatory Costs of Being Public: Evidence from Bunching Estimation."

Farhi, Emmanuel, and Francois Gourio. 2018. "Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia." *Brookings Papers on Economic Activity*, Fall.

Feenstra, Robert, Robert Inklaar, and Marcel Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review*, 105(10): 3150–3182.

Feichtinger, Gustav. 1982. "Saddle Point Analysis in a Price-Advertising Model." *Journal of Economic Dynamics and Control*, 4(1): 319–340.

Fernandez-Villaverde, Jesus, Galo Nuno, George Sorg-Langhans, and Maximilian Vogler. 2020. “Solving High-Dimensional Dynamic Programming Problems using Deep Learning.” Unpublished working paper.

Fernandez-Villaverde, Jesus, Samuel Hurtado, and Galo Nuno. 2019. “Financial Frictions and the Wealth Distribution.” NBER Working Paper No. 26302.

Frank, Murray Z., and Keer Yang. 2019. “Does Finance Flow to High Productivity Firms?”

Froot, Kenneth A., David S. Scharfstein, and Jeremy C. Stein. 1993. “Risk Management: Coordinating Corporate Investment and Financing Policies.” *Journal of Finance*, 48(5): 1629–1658.

Fulghieri, Paolo, Diego Garcia, and Dirk Hackbarth. 2020. “Asymmetric Information and the Pecking (Dis)order.”

Fuster, Andreas, Matthew Plosser, Philipp Schnabl, and James Vickery. 2019. “The Role of Technology in Mortgage Lending.” *Review of Financial Studies*, 32(5): 1854–1899.

Gabaix, Xavier, and Augustin Landier. 2008. “Why has CEO Pay Increased So Much?” *Quarterly Journal of Economics*, 123(1): 49–100.

Gambacorta, Leonardo, Yiping Huang, Zhenhua Li, Han Qiu, and Shu Chen. 2020. “Data vs collateral.” BIS Working Papers No. 881.

Gao, Janet. 2021. “Managing Liquidity in Production Networks: The Role of Central Firms.” *Review of Finance*, 25(3): 819–861.

Gao, Xiaodan, Toni M. Whited, and Na Zhang. Forthcoming. “Corporate Money Demand.” *Review of Financial Studies.*

Gao, Xiaohui, Jay R. Ritter, and Zhongyan Zhu. 2013. “Where Have All the IPOs Gone?” *Journal of Financial and Quantitative Analysis*, 48(6): 1663–1692.

Gomes, Joao F. 2001. “Financing Investment.” *American Economic Review*, 91(5): 1263–1285.

Gomez, Matthieu. 2019. “Asset Prices and Wealth Inequality.”

Gomme, Paul, B. Ravikumar, and Peter Rupert. 2011. "The Return to Capital and the Business Cycle." *Review of Economic Dynamics*, 14(2): 262–278.

Gopinath, Gita, Sebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. 2017. "Capital Allocation and Productivity in South Europe." *Quarterly Journal of Economics*, 132(4): 1915–1967.

Gourio, Francois, and Leena Rudanko. 2014. "Customer Capital." *Review of Economic Studies*, 81: 1102–1136.

Graham, John R. 2000. "How Big Are the Tax Benefits of Debt?" *Journal of Finance*, 55(5): 1901–1941.

Greenwald, Daniel. 2019. "Firm Debt Covenants and the Macroeconomy: The Interest Coverage Channel." MIT Sloan Research Paper No. 5909-19.

Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94(4): 691–719.

Grullon, Gustavo, Yelena Larkin, and Roni Michaely. 2019. "Are US Industries Becoming More Concentrated?" *Review of Finance*, 23: 697–743.

Gutierrez, German, and Thomas Philippon. 2017. "Investmentless Growth: An Empirical Investigation." *Brookings Papers on Economic Activity*, 48: 89–190.

Gutiérrez, Germán, and Thomas Philippon. 2019. "Fading Stars." NBER Working Paper No. 25529.

Guvenen, Fatih. 2020. "Macroeconomics With Heterogeneity: A Practical Guide." NBER Working Paper No. 17622.

Hall, Robert E. 2014. "Quantifying the Lasting Harm to the U.S. Economy from the Financial Crisis." NBER Working Paper No. 20183.

Hart, Oliver, and John Moore. 1994. "A Theory of Debt Based on the Inalienability of Human Capital." *Quarterly Journal of Economics*, 109(4): 841–879.

Hart, Oliver D., and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98(6): 1119–1158.

Haskel, Jonathan, and Stian Westlake. 2018. *Capitalism without Capital: The Rise of the Intangible Economy.* Princeton University Press.

Hau, Harald, Yi Huang, Hongzhe Shan, and Zixia Sheng. 2018. "FinTech Credit and Entrepreneurial Growth." Unpublished Working Paper.

Hayashi, Fumio. 1982. "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica*, 50(1): 213–224.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2009. "Quantitative Macroeconomics with Heterogeneous Households." *Annual Review of Economics*, 1: 319–354.

Hennessy, Christopher A., Amnon Levy, and Toni M. Whited. 2007. "Testing Q theory with financing frictions." *Journal of Financial Economics*, 83(3): 691–717.

Herskovic, Bernard, Bryan Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh. 2016. "The common factor in idiosyncratic volatility: Quantitative asset pricing implications." *Journal of Financial Economics*, 119(2): 249–283.

He, Zhiguo, and Arvind Krishnamurthy. 2013. "Intermediary Asset Pricing." *American Economic Review*, 103(2): 732–770.

Hoberg, Gerard, and Gordon M. Phillips. 2021. "Scope, Scale and Competition: The 21st Century Firm."

Holmstrom, Bengt, and Jean Tirole. 1997. "Financial Intermediation, Loanable Funds, and The Real Sector." *Quarterly Journal of Economics*, 112(3): 663–691.

Hopenhayn, Hugo A. 2014. "Firms, Misallocation, and Aggregate Productivity: A Review." *Annual Review of Economics*, 6(1): 735–770.

Hsieh, Chang-Tai, and Esteban Rossi-Hansberg. 2019. "The Industrial Revolution in Services." NBER Working Paper w25968.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403–1448.

Hsieh, Chang-Tai, and Peter J. Klenow. 2018. "The Reallocation Myth."

Huang, Yiping, Longmei Zhang, Zhenhua Li, Han Qiu, Tao Sun, and Xue Wang. 2020. "Fintech Credit Risk Assessment for SMEs: Evidence from China." IMF Working Paper No. 20-193.

Imrohoroglu, Ayse. 1989. "Cost of Business Cycles with Indivisibilities and Liquidity Constraints." *Journal of Political Economy*, 97(6): 1364–1383.

Imrohoroglu, Ayse, and Selale Tuzel. 2014. "Firm-Level Productivity, Risk, and Return." *Management Science*, 60(8): 1861–2109.

Jensen, Michael C. 1986. "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers." *American Economic Review*, 76(2): 323–329.

Jensen, Michael C. 1989. "Eclipse of the Public Corporation." Havard Business Review.

Jiang, Zhengyang, Hanno Lustig, Stijn Van Nieuwerburgh, and Mindy Z. Xiaolan. 2021. "Manufacturing Risk-free Government Debt." NBER Working Paper w27786.

Jones, Callum, and Thomas Philippon. 2016. "The Secular Stagnation of Investment."

Jung, Hae Won (Henny), Dalida Kadyrzhanova, and Ajay Subramanian. 2021. "Capital Structure under Imperfect Product Market Competition: Theory and Evidence."

Kahle, Kathleen M., and René M. Stulz. 2017. "Is the US Public Corporation in Trouble?" *Journal of Economic Perspectives*, 31(3): 67–88.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018*a*. "Monetary Policy According to HANK." *American Economic Review*, 108(3): 697–743.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018*b*. "Monetary Policy According to HANK." *American Economic Review*, 108(3): 697–743.

Kaplan, Steven N., and Luigi Zingales. 1997. "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?" *Quarterly Journal of Economics*, 112(1): 169–215.

Karabarbounis, Loukas, and Brent Neiman. 2018. “Accounting for Factorless Income.” NBER Working Paper No. 24404.

Kehrig, Matthias, and Nicolas Vincent. 2020. “The Micro-Level Anatomy of the Labor Share Decline.” NBER Working Paper No. 25275.

Kelley, C. T., and David E. Keyes. 1998. “Convergence Analysis of Pseudo-Transient Continuation.” *SIAM Journal on Numerical Analysis*, 35(2): 508–523.

Kelley, C. T., Li-Zhi Liao, Liqun Qi, Moody T. Chu, J. P. Reese, and C. Winton. 2008. “Projected Pseudotransient Continuation.” *SIAM Journal on Numerical Analysis*, 46(6): 3071–3083.

Keyes, David E., and Mitchell D. Smooke. 1987. “A parallelized elliptic solver for reacting flows.” *Parallel computations and their impact on mechanics*, 375–402.

Kilic, Mete, Louis Yang, and Miao Ben Zhang. 2019. “The Great Divorce Between Investment and Profitability.”

Kiyotaki, Nobuhiro. 1998. “Credit and Business Cycles.” *Japanese Economic Review*, 49(1): 18–35.

Kiyotaki, Nobuhiro, and John Moore. 1997. “Credit Cycles.” *Journal of Political Economy*, 105(2): 211–248.

Kose, M. Ayhan, Franziska Ohnsorge, Lei Sandy Ye, and Ergys Islamaj. 2017. “Weakness in Investment Growth : Causes, Implications and Policy Responses.” World Bank Policy Research Working Paper No.7990.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2012. “The Network Origins of Aggregate Fluctuations.” *Journal of Political Economy*, 120(2): 233–267.

Kroen, Thomas, Ernest Liu, Atif R. Mian, and Amir Sufi. 2021. “Falling Rates and Rising Superstars.”

Krusell, Per, and Anthony A. Smith. 1998. “Income and Wealth Heterogeneity in the Macroeconomy.” *Journal of Political Economy*, 106(5): 867–896.

Lee, Dong Wook, Hyun-Han Shin, and Rene M. Stulz. 2020. “Why Does Equity Capital Flow Out of High Tobin’s q Industries?”

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, 70(2): 317–341.

Lewisi, Christine, Nigel Paini, Jan Stráskýi, and Fusako Menkynai. 2014. "Investment Gaps after the Crisis." OECD Economics Department Working Papers No. 1168.

Lian, Chen, and Yueran Ma. 2021. "Anatomy of Corporate Borrowing Constraints." *Quarterly Journal of Economics*, 136(1): 229–291.

Liu, Ernest, Atif Mian, and Amir Sufi. 2019. "Low Interest Rates, Market Power, and Productivity Growth."

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics*, 135: 561–644.

Luptacik, Mikulas. 1982. "Optimal Price and Advertising Policy under Atomistic Competition." *Journal of Economic Dynamics and Control*, 4(1): 57–71.

Luttmer, Erzo G. J. 2007. "Selection, Growth, and the Size Distribution of Firms." *Quarterly Journal of Economics*, 122(3): 1103–1144.

Luttmer, Erzo G.J. 2011. "On the Mechanics of Firm Growth." *Review of Economic Studies*, 78(3): 1042–1068.

Luttmer, Erzo G.J. 2012. "Technology diffusion and growth." *Journal of Economic Theory*, 147(2): 602–622.

MacKay, Peter, and Gordon Phillips. 2005. "How Does Industry Affect Firm Financial Structure?" *Review of Financial Studies*, 18(4): 1433–1466.

Maksimovic, Vojislav. 1988. "Capital Structure in Repeated Oligopolies." *RAND Journal of Economics*, 19(3): 389–407.

Mankiw, Gregory. 2015. "Yes, r ¿ g. So What?" *American Economic Review: Papers & Proceedings*, 105(5): 43–47.

McFadden, Daniel. 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." *Econometrica*, 57(5): 995–1026.

Midrigan, Virgiliu, and Daniel Yi Xu. 2014. "Finance and Misallocation: Evidence from Plant-Level Data." *American Economic Review*, 104(2): 422–458.

Miller, Merton H., and Daniel Orr. 1966. "A Model of the Demand for Money by Firms." *Quarterly Journal of Economics*, 80(3): 413–435.

Moll, Benjamin. 2012. "Inequality and Financial Development: A Power-Law Kuznets Cruve."

Moll, Benjamin. 2014. "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?" *American Economic Review*, 104(10): 3186–3221.

Morlacco, Monica, and David Zeke. 2021. "Monetary Policy, Customer Capital, and Market Power." *Journal of Monetary Economics*, 121: 116–134.

Mulder, Wim A., and Bram Van Leer. 1985. "Experiments with implicit upwind methods for the Euler equations." *Journal of Computational Physics*, 59(2): 232–246.

Myers, Stewart C., and Nicholas S. Majluf. 1984. "Corporate financing and investment decisions when firms have information that investors do not have." *Journal of Financial Economics*, 13(2): 187–221.

Nekarda, Christopher J., and Valerie A. Ramey. 2013. "The Cyclical Behavior of the Price-Cost Markup." NBER Working Paper No. 19099.

Nikolov, Boris, and Toni M. Whited. 2014. "Agency Conflicts and Cash: Estimates from a Dynamic Model." *Journal of Finance*, 69(5): 1883–1921.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Economics Letters*, 64(6): 1263–1297.

Olmstead-Rumsey, Jane. 2021. "Market Concentration and the Productivity Slowdown."

Peters, Ryan, and Lucian Taylor. 2017. "Intangible capital and the investment-q relation." *Journal of Financial Economics*, 123(2): 251–272.

Phelps, Edmund S., and Sidney G. Winter. 1970. "Optimal Price Policy Under Atomistic Competition." In *Microeconomic Foundations of Employment and Inflation Theory.* , ed. G. C. Archibald, Armen A. Alchian and Edmund S. Phelps, 309–337. Macmillan, New York:W. W. Norton & Company.

Philippon, Thomas. 2016. "The FinTech Opportunity." NBER Working Paper 22476.

Piketty, Thomas. 2013. *Capital in the Twenty-First Century.* Harvard University Press.

Raissi, M., P. Perdikaris, and G. E. Karniadakis. 2019. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations." *Journal of Computational Physics*, 378(1): 686–707.

Ravn, Morten, Stephanie Schmitt-Grohe, and Martin Uribe. 2006. "Deep Habits." *Review of Economic Studies*, 73(1): 195–218.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments." *Review of Economic Dynamics*, 11(4): 707–720.

Restuccia, Diego, and Richard Rogerson. 2017. "The Causes and Costs of Misallocation." *Journal of Economic Perspectives*, 31(3): 151–174.

Riddick, Leigh A., and Toni Whited. 2009. "The Corporate Propensity to Save." *Journal of Finance*, 64(4): 1729–1766.

Rosen, Sherwin. 1981. "The Economics of Superstars." *American Economic Review*, 71(5): 845–858.

Rotemberg, Julio, and Michael Woodford. 1992. "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity." *Journal of Political Economy*, 100(6): 1153–1207.

Sattinger, Michael. 1993. "Assignment Models of the Distribution of Earnings." *Journal of Economic Literature*, 31(2): 831–880.

Scheuer, Florian, and Iván Werning. 2017. "The Taxation of Superstars." *Quarterly Journal of Economics*, 132(1): 211–270.

Smith, Adam. 1759. *The Theory of Moral Sentiments.* Oxford University Press.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations.* Oxford University Press.

Stiglitz, Joseph E., and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review*, 71(3): 393–410.

Stokey, Nancy. 2008. *The Economics of Inaction: Stochastic Control Models with Fixed Costs.* Princenton University Press.

Su, Dan. 2021. "Rise of Superstar Firms and Fall of the Price Mechanism."

Summers, Lawrence H. 2013. "Remarks at IMF Fourteenth Annual Research Conference in Honor of Stanley Fischer."

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Perspectives*, 49(2): 326–365.

Tang, Huan. 2019. "Peer-to-peer lenders versus banks: substitutes or complements?" *Review of Financial Studies*, 32(5): 1900–1938.

Tervio, Marko. 2008. "The Difference That CEOs Make: An Assignment Model Approach." *American Economic Review*, 98(3): 642–668.

Thakor, Anjan V. 2020. "Fintech and banking: What do we know?" *Journal of Financial Intermediation*, 41: 1–13.

Thakor, Richard, and Robert Merton. 2019. "Trust in Lending." MIT Sloan Research Paper No. 5524-18.

Townsend, Robert. 1979. "Optimal contracts and competitive markets with costly state verification." *Journal of Economic Theory*, 21(2): 265–293.

Traina, James. 2021. "Is Aggregate Market Power Increasing? Production Trends Using Financial Statements." Unpublished Working Paper.

Wang, Chong, Neng Wang, and Jinqiang Yang. 2012. "A Unified Model of Entreprenuership Dynamics." *Journal of Financial Economics*, 106: 1–23.

Williamson, Oliver. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications.* The Free Press.

Williamson, Oliver. 1979. "Transaction-Cost Economics: The Governance of Contractual Relations." *Journal of Law and Economics*, 22(2): 233–271.

Williamson, Oliver. 1981. "The Economics of Organization: The Transaction Cost Approach." *American Journal of Sociology*, 87(3): 548–577.

Williamson, Oliver. 2010. "Transaction Cost Economics: The Natural Progression." *American Economic Review*, 100(3): 673–690.

Wooldridge, Jeffrey M. 2009. "On estimating firm-level production functions using proxy variables to control for unobservables." *Economics Letters*, 104(3): 112–114.

Appendix A

Appendix for Chapter 2

A.1 Proof

Proof of Lemma 2.1

The optimization problem for the entrepreneur can be rewritten as follows:

$$\pi \equiv \max_{y} \zeta^{\phi} y - f_0 - \xi_0 y^{\frac{1}{\eta}} \tag{A1}$$

Therefore, first order condition gives us the optimal choice of product quantity y as follows:

$$y = \left(\frac{\eta \zeta^{\phi}}{\xi_0} \right)^{\frac{\eta}{1-\eta}} \tag{A2}$$

Therefore, the markup for the entrepreneur with capital quality ζ can be computed as follows:

$$\mu = \frac{py}{f_0 + \xi_0 y^{\frac{1}{\eta}}} = \frac{1}{\eta + f_0 \left[\zeta^{\frac{\phi}{1-\eta}} \left(\frac{\eta}{\xi_0} \right)^{\frac{\eta}{1-\eta}} \right]^{-1}}$$

The entrepreneur's earnings can be calculated as follows:

$$\pi = py - f_0 - \xi_0 y^{\frac{1}{\eta}} = (1 - \eta) \left(\frac{\eta}{\xi_0} \right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi}{1-\eta}} - f_0$$

Proof of Lemma 2.2

The capital quality follows the following process

$$d\zeta_t = \left(\bar{\mu} + \iota_t^{\zeta} - \delta\zeta_t\right) dt + \sigma\sqrt{\zeta_t} d\mathcal{Z}_t$$

From Lemma 2.1, we know that π is a function of ζ, i.e., $\pi = (1-\eta)\left(\frac{\eta}{\xi_0}\right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi}{1-\eta}} - f_0$. With Ito's lemma, we can obtain the earnings process as follows:

$$d\pi_t = \left[\pi'(\zeta_t)\left(\bar{\mu} + \iota_t^{\zeta} - \delta\zeta_t\right) + \frac{\sigma^2\zeta_t}{2}\pi''(\zeta_t)\right] dt + \pi'(\zeta_t)\sigma\sqrt{\zeta_t} d\mathcal{Z}_t$$

where we have

$$\begin{aligned} \pi' &= \phi\left(\frac{\eta}{\xi_0}\right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi}{1-\eta}-1} \\ \pi'' &= \phi\left(\frac{\phi}{1-\eta} - 1\right)\left(\frac{\eta}{\xi_0}\right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi}{1-\eta}-2} \end{aligned}$$

Proof of Proposition 2.1

If $\phi+\eta > 1$, then we have $\pi'(\zeta) > 0$ and $\pi''(\zeta) > 0$ for all possible values of ζ. Therefore, π is a convex function of ζ.

At the same time, output y is not necessarily a convex function of ζ if $\phi + \eta > 1$. As $y = \left(\frac{\eta\zeta^{\phi}}{\xi_0}\right)^{\frac{\eta}{1-\eta}}$, therefore we have

$$\begin{aligned} y'(\zeta) &= \frac{\phi\eta}{1-\eta}\left(\frac{\eta}{\xi_0}\right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi\eta-1+\eta}{1-\eta}} \\ y''(\zeta) &= \frac{\phi\eta}{1-\eta}\frac{\phi\eta-1+\eta}{1-\eta}\left(\frac{\eta}{\xi_0}\right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi\eta-2+2\eta}{1-\eta}} \end{aligned}$$

Output y is not necessarily a convex function of ζ if and only $> \frac{1-\eta}{\eta}$, which is a stricter condition than $\phi > 1-\eta$.

The drift part of $d\pi$ is $\pi'(\zeta_t)\left(\bar{\mu} + \iota_t^{\zeta} - \delta\zeta_t\right) + \frac{\sigma^2\zeta_t}{2}\pi''(\zeta_t)$ and its volatility part is $\pi'(\zeta_t)\sigma\sqrt{\zeta_t}$. In steady state, we have $\iota^{\zeta} = \delta\zeta$. With $\pi'(\zeta) > 0$ and $\pi''(\zeta) > 0$, we can easily check that these two components are increasing in ζ in the steady state.

Proof of Proposition 2.2

Proof of Proposition 2.2 has mainly three parts. First, we will show the existence of this double-barrier cash accumulation policy. Second, we will derive the Hamilton–Jacobi–Bellman (HJB) equation that characterizing firm's optimal choice between internal and external financing. Third, we will list the corresponding boundary conditions for the HJB equation.

Existence First, we need to prove the existence of downward control boundary $\overline{\Omega}$ and upward control boundary $\underline{\Omega}$.

Suppose there is no such downward control boundary $\overline{\Omega}$, which means that

$$lim_{\omega \to +\infty} \left[\mathcal{J}\left(\zeta, \omega, b\right) - \mathcal{J}^0\left(\zeta, \omega, b\right) \right] = 0 \tag{A3}$$

where $\mathcal{J}^0\left(\zeta, \omega, b\right)$ represents the value with no boundary controls.

Now let's consider an downward cash adjustment from z to $z - a$, where $a > 0$ and z is any finite number. As we assume that there is no such downward control boundary $\overline{\Omega}$, thus we have

$$\mathcal{J}\left(\zeta, z, b\right) + \chi_0 + \chi_1 a \geq \mathcal{J}\left(\zeta, z - a, b\right) \tag{A4}$$

In addition, entrepreneur's utility function u is concave function, so there exists $\delta > 0, \varepsilon > 0$ such that

$$\left| \mathcal{J}\left(\zeta, x, b\right) - \mathcal{J}^0\left(\zeta, x, b\right) \right| < \delta \tag{A5}$$

$$-\lambda u'\left(c\left(x\right)\right) + \rho\chi_1 < -\varepsilon \tag{A6}$$

for all $z - a \leq x \leq z$.

Therefore, Equation (A4) can be rewritten as

$$0 \leq \chi_0 + \int_{z-a}^{z} \left[-\lambda u'(c\left(\omega\right)) + \rho\chi_1 \right] d\omega + 2\delta < \chi_0 - a\varepsilon + 2\delta \tag{A7}$$

For any ε and δ, there exists a sufficiently large a such that the equation above fails. Therefore, the original assumption does not hold, which means there is a downward control boundary $\overline{\Omega}$.

A similar argument holds for the existence of an upward control boundary $\underline{\Omega}$.

After showing the existence, now we need to characterize the corresponding HJB equations and the corresponding boundary conditions in each region.

HJB equation *Internal financing region.* Within this region, we know that $b = 0$ and $\iota^b = 0$ because entrepreneurs only use internal cash to finance their investment. The budget constraint here can be simplified to

$$c + \iota^\zeta + \iota^\omega = \pi - \frac{\kappa_0 \left(\iota^\zeta\right)^2}{2\zeta} \tag{A8}$$

Therefore, the value function $\mathcal{J}(\zeta, \omega)$ satisfies the following HJB equation:

$$0 = \max_{\iota^\zeta, c} \left\{ \begin{array}{l} f(c, \mathcal{J}) + \left(\bar{\mu} + \iota^\zeta - \delta\zeta\right) \mathcal{J}_\zeta + \frac{\zeta\sigma^2}{2} \mathcal{J}_{\zeta\zeta} + \zeta\sigma^2 \pi'(\zeta) \mathcal{J}_{\zeta\omega} + \frac{\zeta\sigma^2}{2} (\pi'(\zeta))^2 \mathcal{J}_{\omega\omega} \\ + \left[\pi'(\zeta)\left(\bar{\mu} + \iota^\zeta - \delta\zeta\right) + \frac{1}{2}\pi''(\zeta)\zeta\sigma^2 - c - \iota^\zeta - \lambda\omega - \frac{\kappa_0\left(\iota^\zeta\right)^2}{2\zeta}\right] \mathcal{J}_\omega \end{array} \right\} \tag{A9}$$

First order conditions give us

$$\begin{aligned} f_c(c, \mathcal{J}) &= \mathcal{J}_\omega & \text{(A10)} \\ \frac{\mathcal{J}_\zeta}{\mathcal{J}_\omega} + \pi'(k) &= 1 + \frac{\kappa_0 \left(\iota^\zeta\right)^2}{2\zeta} & \text{(A11)} \end{aligned}$$

External lending region. Within this region, we know that $b < 0$ and $\iota^\omega = \lambda\overline{\Omega}$ because entrepreneurs have accumulated enough cash so any additional "savings" will be lent in the financial market. As lending doesn't incur any transaction costs, the budget constraint here can be simplified to

$$c + \iota^\zeta - \iota^b = \pi - rb - \frac{\kappa_0 \left(\iota^\zeta\right)^2}{2\zeta} - \lambda\overline{\Omega} \tag{A12}$$

Therefore, the value function $\mathcal{J}(\zeta, b)$ satisfies the following HJB equation:

$$0 = \max_{\iota^\zeta, c} \left\{ \begin{array}{l} f(c, \mathcal{J}) + \left(\bar{\mu} + \iota^\zeta - \delta\zeta\right) \mathcal{J}_\zeta + \frac{\zeta\sigma^2}{2} \mathcal{J}_{\zeta\zeta} - \zeta\sigma^2 \pi'(\zeta) \mathcal{J}_{\zeta b} + \frac{\zeta\sigma^2}{2} (\pi'(\zeta))^2 \mathcal{J}_{bb} \\ - \left[\pi'(\zeta)\left(\bar{\mu} + \iota^\zeta - \delta\zeta\right) + \frac{1}{2}\pi''(\zeta)\zeta\sigma^2 - c - \iota^\zeta - rb - \chi_0 + \chi_1 b - \lambda\overline{\Omega} - \frac{\kappa_0\left(\iota^\zeta\right)^2}{2\zeta}\right] \mathcal{J}_b \end{array} \right\} \tag{A13}$$

First order conditions give us

$$\begin{aligned} f_c(c,\mathcal{J}) &= -\mathcal{J}_b && \text{(A14)} \\ -\frac{\mathcal{J}_\zeta}{\mathcal{J}_b} + \pi'(\zeta) &= 1 + \kappa_0 \frac{\iota^\zeta}{\zeta} && \text{(A15)} \end{aligned}$$

External borrowing region. Within this region, we know that $b > 0$ and $\iota^\omega = \lambda\underline{\Omega}$ because entrepreneurs have run short of cash so he will borrow from the financial market despite the transaction costs. The budget constraint here can be written as

$$c + \iota^k - \iota^b = \pi - rb - \mathbf{1}_{b_{i,t}>0}(\chi_0 + \chi_1 b_{i,t}) - \frac{\kappa_0 \left(\iota^\zeta\right)^2}{2\zeta} - \lambda\underline{\Omega} \tag{A16}$$

Therefore, the value function $\mathcal{J}(\zeta, b)$ satisfies the following HJB equation:

$$0 = \max_{\iota^\zeta, c} \left\{ \begin{array}{l} f(c,\mathcal{J}) + \left(\bar{\mu} + \iota^\zeta - \delta\zeta\right)\mathcal{J}_\zeta + \frac{\zeta\sigma^2}{2}\mathcal{J}_{\zeta\zeta} - \zeta\sigma^2\pi'(\zeta)\mathcal{J}_{\zeta b} + \frac{\zeta\sigma^2}{2}(\pi'(\zeta))^2\mathcal{J}_{bb} \\ -\left[\pi'(\zeta)\left(\bar{\mu} + \iota^\zeta - \delta\zeta\right) + \frac{1}{2}\pi''(\zeta)\zeta\sigma^2 - c - \iota^\zeta - rb - \chi_0 - \chi_1 b - \lambda\underline{\Omega} - \frac{\kappa_0\left(\iota^\zeta\right)^2}{2\zeta}\right]\mathcal{J}_b \end{array} \right\} \tag{A17}$$

First order conditions give us

$$\begin{aligned} f_c(c,\mathcal{J}) &= -\mathcal{J}_b && \text{(A18)} \\ -\frac{\mathcal{J}_\zeta}{\mathcal{J}_b} + \pi'(\zeta) &= 1 + \kappa_0 \frac{\iota^\zeta}{\zeta} && \text{(A19)} \end{aligned}$$

In addition, we have the following borrowing constraint:

$$b \leq \eta \frac{\pi - \chi_0}{1 + r + \chi_1} \tag{A20}$$

Boundary conditions To completely characterize the economy, we also need to determine the boundary $\underline{\Omega}$ at which the entrepreneurs raise new external funds, and the boundary $\overline{\Omega}$ at which the entrepreneurs start to lend in the capital market.

External borrowing region. Since external financing is costly, therefore entrepreneurs will only issue debt when their cash holdings have been below some level $\underline{\Omega}$. As entrepreneur's continuation value is continuous before and after debt issuance, we have the following requirement for marginal value of cash $\mathcal{J}_\omega(\zeta,\omega)$ at the upward control boundary $\underline{\Omega}$:

$$\mathcal{J}_\omega\left(\zeta,\underline{\Omega}\right)=1+r+\chi_1$$

The intuition is that entrepreneurs will not borrow from the financial market unless the marginal value of internal financing has already reached the level of marginal cost of external financing. As marginal borrowing cost contains both the interest rate r and marginal issuance cost χ_1, we have the equation above for $\underline{\Omega}$.

External lending region. Similarly, entrepreneur's continuation value must be continuous before and after lending in the financial market. Therefore, for $\omega>\overline{\Omega}$, we have the following equation for $\mathcal{J}$:

$$\mathcal{J}\left(\zeta,\omega\right)=\mathcal{J}\left(\zeta,\overline{\Omega}\right)+\left(1+r\right)\left(\omega-\overline{\Omega}\right)$$

Since the above equation also holds for ω close to $\overline{\Omega}$, we may take the limit and obtain the following condition for the endogenous upper boundary $\overline{\Omega}$:

$$\mathcal{J}_\omega\left(\zeta,\overline{\Omega}\right)=1+r-\chi_1$$

The intuition behind the equation above is that entrepreneurs have accumulated enough cash such that cash becomes equivalent to the negative debt. This condition is different from Bolton, Chen and Wang (2011) as in their paper, entrepreneurs pay out cash as dividends at the downward control boundary. Therefore, in Bolton, Chen and Wang (2011), we have $\mathcal{J}_\omega\left(\zeta,\overline{\Omega}\right)=1$. However, in this paper, entrepreneurs have options to lend their cash to others in the financial market, so the marginal value of cash cannot go below marginal value of debt. Here Since the external lending boundary $\overline{\Omega}$ is optimally chosen, we also have the following "super contract" condition (Dumas, 1991):

$$\mathcal{J}_{\omega\omega}\left(\zeta,\overline{\Omega}\right)=0$$

Reflecting barriers. We also need the boundary conditions in ζ-dimension, which correspond to "reflecting barriers" at lower and upper bounds for capital quality, ζ_{min} and ζ_{max} (Dixit, 1993).

$$\mathcal{J}_{\zeta}\left(\zeta_{min},\omega\right)=0,\forall\omega$$
$$\mathcal{J}_{\zeta}\left(\zeta_{max},\omega\right)=0,\forall\omega$$

A.2 Data and Variable Construction

The data used in this paper mainly comes from two sources. First, firm-level balanced sheet data is obtained from the WRDS Compustat North America Fundamentals data file from 1980 to 2019 with consolidation level "C", industry format "INDL", data format "STD", and population source "D". The dataset contains both surviving and non-surviving firms. We keep all the entries with a foreign incorporation code of "USA", exclude financial firms (SIC 6000-6999) and regulated utilities (SIC 4900-4999), and drop firms with missing or negative values on assets or sales. As for the international evidence section, we have adopted similar data cleaning process but with WRDS Compustat Global Fundamentals data file. Second, historical gross output price and capital price at the sectoral level are obtained from Integrated Industry-Level Production Account (KLEMS). We use the gross output price indices to deflate firms' sales and cost of goods sold, and use capital price indices to deflate firms' capital investment and physical asset stocks.

- **asset tangibility**: ratio of physical assets (Compustat series $PPENT$) to total assets (Compustat series AT)
- **book leverage**: the ratio of the amount of debts to the sum of total debts and common equity, i.e., book leverage = $\frac{\text{DLTT+DLC}}{\text{DLTT+DLC+CEQ}}$, where $DLTT$, DLC, and CEQ represent the long-term debt, current liabilities, and common equity, respectively.
- **dividend payout**: dividends (Compustat series DVC) scaled by total assets
- **investment**: capital expenditures (Compustat series $CAPX$) scaled by total assets (Compustat series AT)
- **markup**: In the existing literature, there are three different approaches to measuring the corporate markup: the accounting profits approach, the user cost approach, and the production function estimation approach. As discussed in Loecker,

Eeckhout and Unger (2020), the first two approaches have some serious issues, so we prefer using the production cost function approach as they do in their paper. Generally speaking, we compute the firm-level markup in three steps. First, we estimate the elasticity of output with respect to variable inputs. Second, we compute the revenue share of each variable input. Finally, we obtain an estimate of markup by calculating the product of these two key ingredients. This approach is advantageous as it does not require specifying a particular demand system. More specifically, we use Olley and Pakes (1996) methodology but with Ackerberg, Caves and Frazer (2015)'s correction to estimate the output elasticities. In addition, 3-digit industry-level sales shares are included to control for the estimation of markups.

- **net finance measures**:
  - ratio of financing activities net cash flow (Compustat series *FINCF*) to total assets (Compustat series *AT*)
  - ratio of financing activities net cash flow with dividend adjustments (Compustat series *DLTIS* - *DLTR* + *DLCCH* + *SSTK* - *PRSTKC* - *DV*) to total assets
  - ratio of financing activities net issuance (Compustat series *DLTIS* - *DLTR* + *DLCCH* + *SSTK* - *PRSTKC*) to total assets
- **operating profits**: Compustat series *IB*
- **operating expenses**: Compustat series *XOPR*
- **Peters and Taylor (2017)'s Total** q: One issue with the standard q measure is that we do not consider the role of intangible capital. Peters and Taylor (2017) provide a new dataset on Total q and consider the intangible capital's replacement costs. Here we simply follow their approach and estimate the replacement cost of firms' intangible capital by accumulating past investments in Research and Development(Compustat series *XRD*) and Selling, General, and Administrative Expenses (Compustat series *XSGA*). Then we can estimate the Peters and Taylor (2017) 's Total q as an improved proxy for Tobin's q.
- **research and development expenditures**: Compustat series *XRD*

- **return of asset**: income before extraordinary items (Compustat series IB) scaled by total assets (Compustat series AT)
- **revenue**: Compustat series $SALE$
- **size**: natural logarithm of total assets (Compustat series AT)
- **selling, general and administrative expense**: Compustat series $XSGA$
- **Tobin's** q: Tobin's q is the ratio between a firm's market value over the replacement cost of its assets. Following the existing empirical works (e.g., Kaplan and Zingales, 1997), we measure firm-level Tobin's q as the market value of the firm's total assets divided by the book value of its assets. In terms of the market value of assets, we compute it as book value of assets (Compustat series AT) plus the market value of common stock (Compustat series $PRCC\_C$ times Compustat series $CSHO$) minus the book value of equity, where the equity book value is estimated as the sum of shareholder equity (Compustat series SEQ), deferred taxes (Compustat series $TXDB$), and investment tax credit (Compustat series $ITCB$), minus the value of preferred stocks (coalesce outcomes of Compustat series $PSTKRV$, $PSTKL$, and $PSTK$).

A.3 Computational Details

The Hamilton–Jacobi–Bellman (HJB) equation in this paper is highly non-linear while the Kolmogorov forward equation (KFE) is still a linear partial differential equation (PDE). Therefore, we use implicit finite difference scheme for solving the HJB equation. After finding the solutions to HJB, KFE can be easily solved with linear operator. A sketch of computational alogorithm is listed as follows:

1. guess the market prices and distribution
2. global search on boundary conditions
3. solve non-linear PDEs given boundary conditions
4. obtain stationary distribution
5. check whether market clears and distribution is the guessed one

A.3.1 Boundary Condition

One of the crucial steps is to find $N_\zeta \times 1$ vector $\mathcal{BC}_{\omega_{min}}$, $N_\zeta \times 1$ vector $\mathcal{BC}_{\omega_{max}}$, $N_\omega \times 1$ vector $\mathcal{BC}_{\zeta_{min}}$, and $N_\omega \times 1$ vector $\mathcal{BC}_{\zeta_{max}}$ for the following boundary conditions

$$\tilde{\mathcal{J}}_\omega(\omega_{min}) = \begin{bmatrix} \mathcal{J}_\omega(\zeta_{min}, \omega_{min}) \\ \ldots\ldots \\ \mathcal{J}_\omega(\zeta_{max}, \omega_{min}) \end{bmatrix} = \mathcal{BC}_{\omega_{min}} \tag{A21}$$

$$\tilde{\mathcal{J}}_\omega(\omega_{max}) = \begin{bmatrix} \mathcal{J}_\omega(\zeta_{min}, \omega_{max}) \\ \ldots\ldots \\ \mathcal{J}_\omega(\zeta_{max}, \omega_{max}) \end{bmatrix} = \mathcal{BC}_{\omega_{max}} \tag{A22}$$

$$\tilde{\mathcal{J}}_\zeta(\zeta_{min}) = \begin{bmatrix} \mathcal{J}_\zeta(\zeta_{min}, \omega_{min}) \\ \ldots\ldots \\ \mathcal{J}_\zeta(\zeta_{min}, \omega_{max}) \end{bmatrix} = \mathcal{BC}_{\zeta_{min}} \tag{A23}$$

$$\tilde{\mathcal{J}}_\zeta(\zeta_{max}) = \begin{bmatrix} \mathcal{J}_\zeta(\zeta_{max}, \omega_{min}) \\ \ldots\ldots \\ \mathcal{J}_\zeta(\zeta_{max}, \omega_{max}) \end{bmatrix} = \mathcal{BC}_{\zeta max} \tag{A24}$$

for the HJB equation for the internal financing region, such that

1. solution to the HJB equation (2.27) converges
2. Neumann boundary conditions in the ζ-dimension (2.30) and (2.31) hold
3. there exist a downward control boundary $\overline{W}^\zeta$, an upward control boundary $\underline{W}^\zeta$, and an exit boundary W^ζ that satisfy the Neumann boundary conditions (2.32)-(2.34)

We search the solutions of $\mathcal{BC}_{\omega_{min}}$, $\mathcal{BC}_{\omega_{max}}$, $\mathcal{BC}_{\zeta_{min}}$, and $\mathcal{BC}_{\zeta_{max}}$ with a global solution algorithm *Simulated Annealing.*

A.3.2 HJB Equation

Once we have the information on the boundary conditions, we can solve the HJB equation. Due to the non-linearity of Equation (2.27), we cannot use the linear operator approach in Achdou et al. (Forthcoming) or the usual Newton-Armijo method. Here we use the pseudo-transient continuation method with switched evolution relaxation method (Ψtc-SER), which is one of the widely used strategy for finding the steady-state global solution to a nonlinear PDE. The basic idea of Ψtc is the following (Keyes and Smooke, 1987; Kelley et al., 2008). As we gradually include the boundary condition information across different stages of calculations, and the amount of information incorporated has strong propagation effects on transients, therefore we need to be careful about choosing the time steps. Otherwise, the results could easily become non-robust and/or not converging to the global solution. The idea of switched evolution relation is set the time step based on a measure of convergence inferred from reduction in a residual norm between consecutive iterations.

More specifically, the original problem can be stated as follows:

$$\nabla \mathcal{J} = \mathcal{F}(\mathcal{J}); \mathcal{J}(0) = \mathcal{J}_0 \tag{A25}$$

Assuming that a stable steady-state solution exists, global convergence and local superlinear convergence to that are proved in Kelley and Keyes (1998) for a class of methods that integrate

$$\nabla \mathcal{J} = -\mathcal{H}^{-1}\mathcal{F}(\mathcal{J}); \mathcal{J}(0) = \mathcal{J}_0 \tag{A26}$$

by a variable time step method that attempts to increase the time step as the integration progresses and steady state is approached. One method is

$$\mathcal{J}_{n+1} = \mathcal{J}_n - \left(\delta_n^{-1}\mathcal{H} + \nabla\mathcal{F}(\mathcal{J}_n)\right)^{-1}\mathcal{F}(\mathcal{J}_n) \tag{A27}$$

where δ_n is the time step. With the switched evolution relaxation method (Mulder and Leer, 1985), we have

$$\delta_n = \delta_{n-1}\frac{\mathcal{F}(\mathcal{J}_{n-1})}{\mathcal{F}(\mathcal{J}_n)} \tag{A28}$$

A.3.3 KFE and Stationary Distribution

After solving the HJB equation, essentially we get the transition matrix of all the state variables in this economy. The Kolmogorov forward equation in the baseline model of paper is the following:

$$0 = -\frac{\partial}{\partial\zeta}\left[\mu^{\zeta,*}(\zeta)\,\Upsilon_t(\zeta,\omega,b)\right] - \frac{\partial}{\partial\omega}\left[\mu^{\omega,*}(\omega)\,\Upsilon_t(\zeta,\omega,b)\right] - \frac{\partial}{\partial b}\left[\mu^{b,*}(b)\,\Upsilon_t(\zeta,\omega,b)\right] + \frac{1}{2}\frac{\partial^2}{\partial\zeta^2}\left[\sigma^{\zeta,*}(\zeta)^2\,\Upsilon_t(\zeta,\omega, \tag{A29}$$

Here we will use this equation as an example. As for all the model extensions, we have solved the corresponding KFE in a similar way. We will start with the steady-state equilibrium and then turn to the time-varying case.

In order to solve this equation, we need to specify a linear differential operator $\mathcal{L}$, a boundary condition operator $\mathcal{T}$, and the affine terms $\mathcal{C}(\cdot)$ and d. In this way, the problem can be written as follows:

$$\mathcal{L}\Upsilon(\zeta,\omega,b) = \mathcal{C}(\zeta,\omega,b) \tag{A30}$$

$$\mathcal{T}\Upsilon(\zeta,\omega,b) = d \tag{A31}$$

Comparing to Equation (A29), in stationary equilibrium, we have

$$\mathcal{L} = -\mu_\zeta^{\zeta,*} - \mu_\omega^{\omega,*} - \mu_b^{b,*} + \frac{1}{2}\left(\sigma^{\zeta,*}\right)^2_{\zeta\zeta} - \mu^{\zeta,*}\partial_\zeta - \mu^{\omega,*}\partial_\omega - \mu^{b,*}\partial_b + \frac{\left(\sigma^{\zeta,*}\right)^2}{2}\partial_{\zeta\zeta} \tag{A32}$$

$$\mathcal{C}(\zeta,\omega,b) = 0 \tag{A33}$$

$$\mathcal{T} = \begin{bmatrix} \partial_\zeta|_{\zeta=\zeta_{min},\forall\omega,\forall b} \\ \partial_\zeta|_{\zeta=\zeta_{max},\forall\omega,\forall b} \end{bmatrix} \tag{A34}$$

$$d = \begin{bmatrix} \mathbf{0}_{\omega\times b} \\ \mathbf{0}_{\omega\times b} \end{bmatrix} \tag{A35}$$

The choice of the boundary condition is the reflecting barrier, i.e., homogenous Neumann boundary conditions.

A.3.4 SMM

Generally speaking, the idea of SMM is to simulate the model S times and use the average values of the moments from the simulated data as the estimator for the model moments. More specifically, let m be a vector of moments estimated from the data, and $\hat{m}^s(\Theta)$ be the corresponding vector of moments estimated from the s-th sample simulated using parameters where $s = 1, \cdots, S$. Θ consists of the set of parameters we are interested in. The SMM estimator $\hat{\Theta}$ is to choose Θ to minimize the distance between the data moments and the simulated model moments

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \left(\frac{m - \frac{1}{S}\sum_{s=1}^{S} \hat{m}^s(\Theta)}{m} \right)' W \left(\frac{m - \frac{1}{S}\sum_{s=1}^{S} \hat{m}^s(\Theta)}{m} \right) \tag{A36}$$

Here the moment error function e is chosen to be the percent difference instead of level difference. In this way, we can put all the moments in the same units.

W is a weighting matrix. Following the literature, here we use the two-step variance-covariance estimator of W. The first step is to estimate Θ using the simple identity matrix, i.e., $W = I$, and we can easily get the first-step estimator $\hat{\Theta}_1$:

$$\hat{\Theta}_1 = \underset{\Theta}{\operatorname{argmin}} \left(\frac{m - \frac{1}{S}\sum_{s=1}^{S} \hat{m}^s(\Theta)}{m} \right)' \left(\frac{m - \frac{1}{S}\sum_{s=1}^{S} \hat{m}^s(\Theta)}{m} \right) \tag{A37}$$

Secondly, we can re-estimate Θ using the optimal two-step weighting matrix

$$\hat{\Theta}_2 = \underset{\Theta}{\operatorname{argmin}} \left(\frac{m - \frac{1}{S}\sum_{s=1}^{S} \hat{m}^s(\Theta)}{m} \right)' \hat{\Psi}^{-1} \left(\frac{m - \frac{1}{S}\sum_{s=1}^{S} \hat{m}^s(\Theta)}{m} \right) \tag{A38}$$

A.4 Additional Figures and Tables

Figure A1: Increasing misallocation: scaled and unscaled

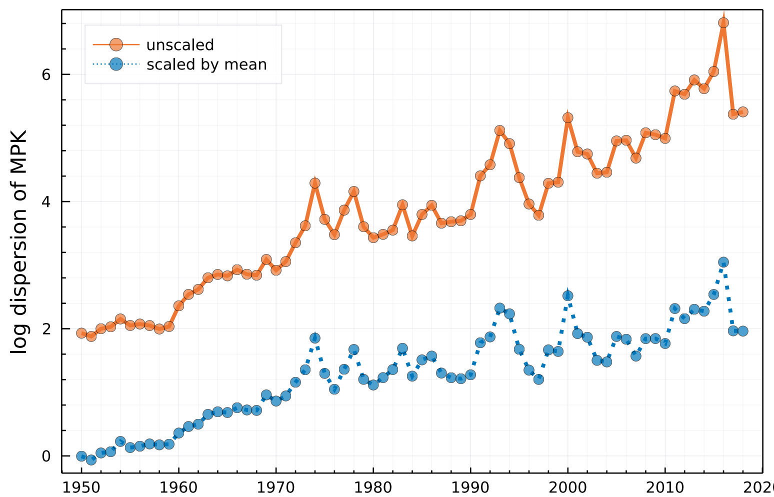


Figure A2: Increasing misallocation: robustness checks

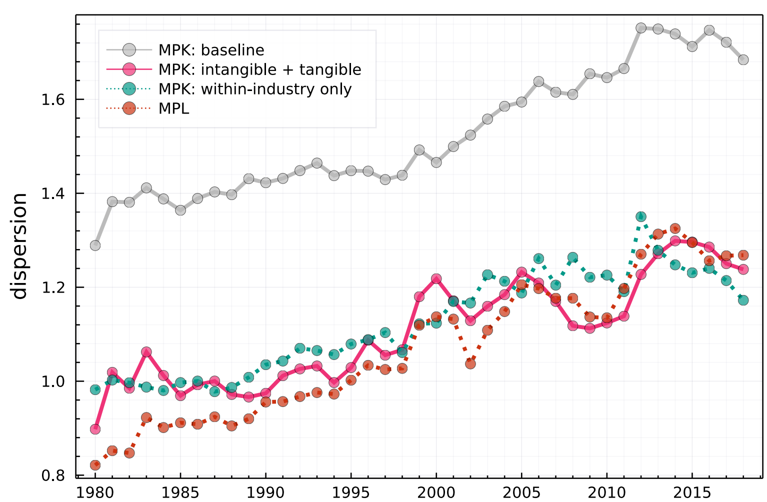


Figure A3: Declining reallocation efficiency: MPK v.s. TFP

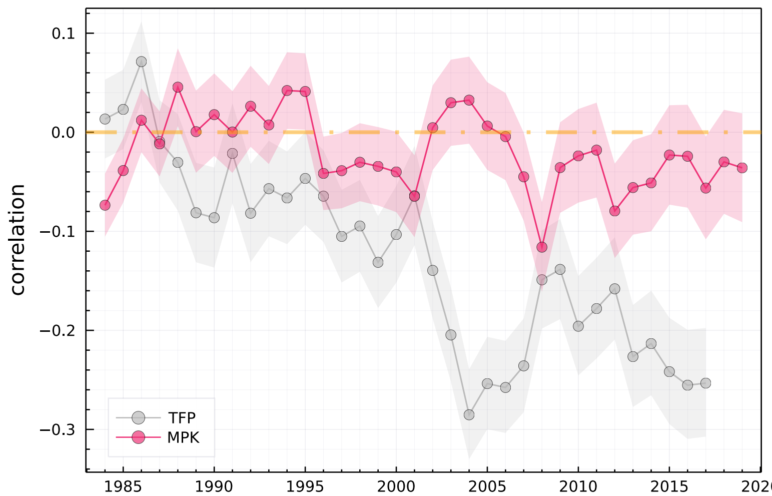


Figure A4: Changing correlation between industry q and funding rate (Lee, Shin and Stulz, 2020)

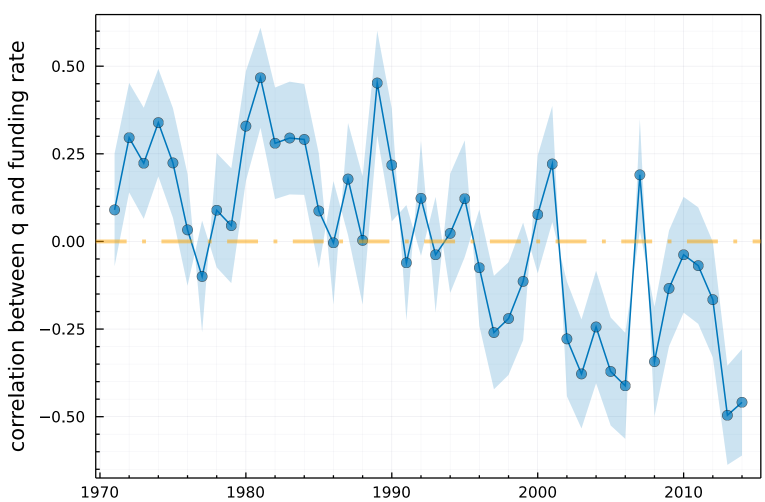


Figure A5: Employment share of large firms

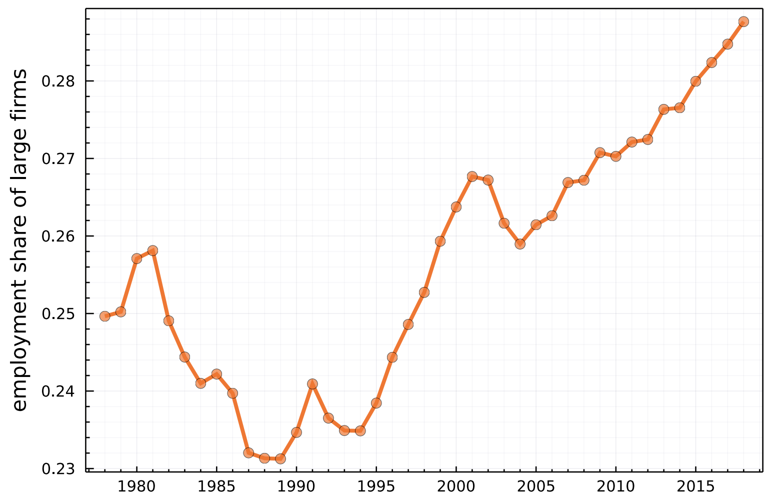


Figure A6: HHI in employment

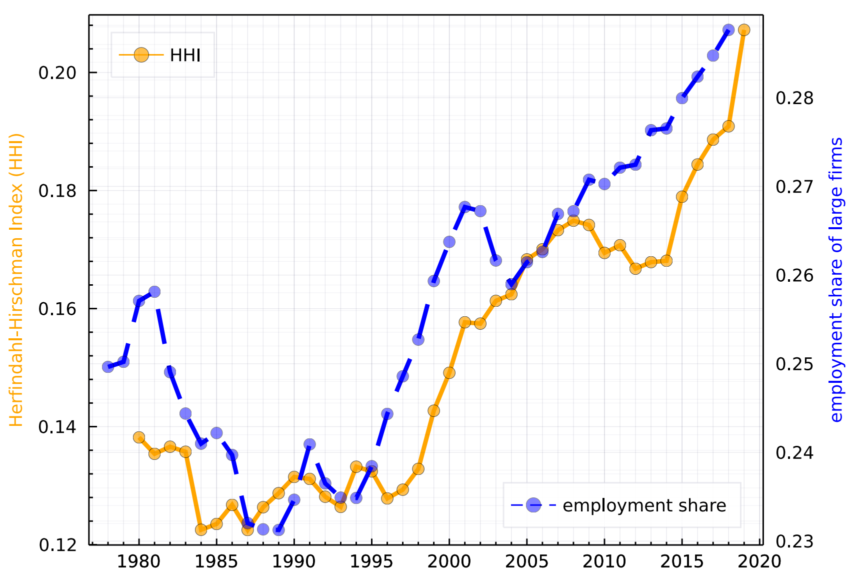


Figure A7: Markup distribution (De Loecker, Eeckhout and Unger, 2020)

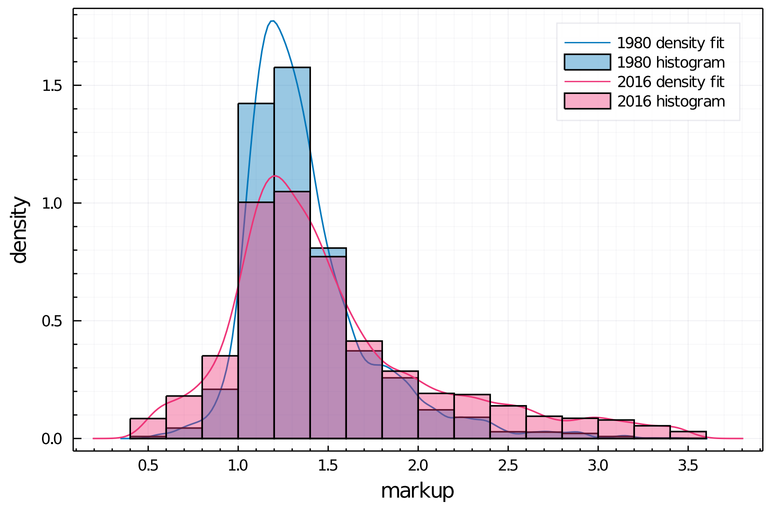


Figure A8: MPK minus r with depreciation and risk premium adjustments

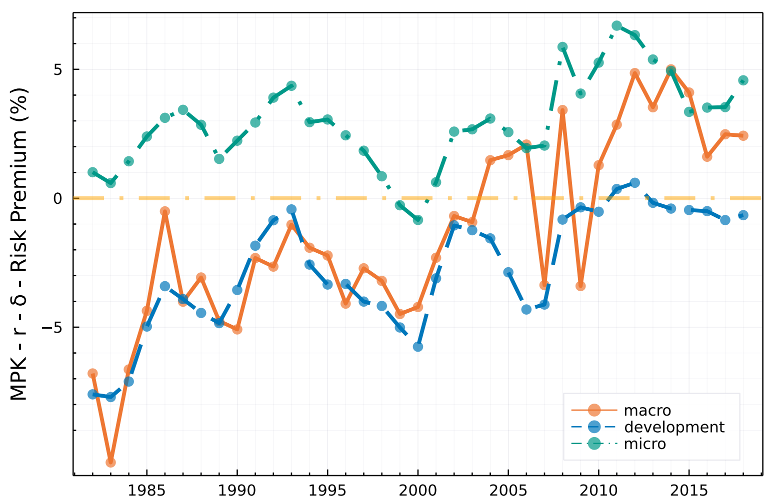
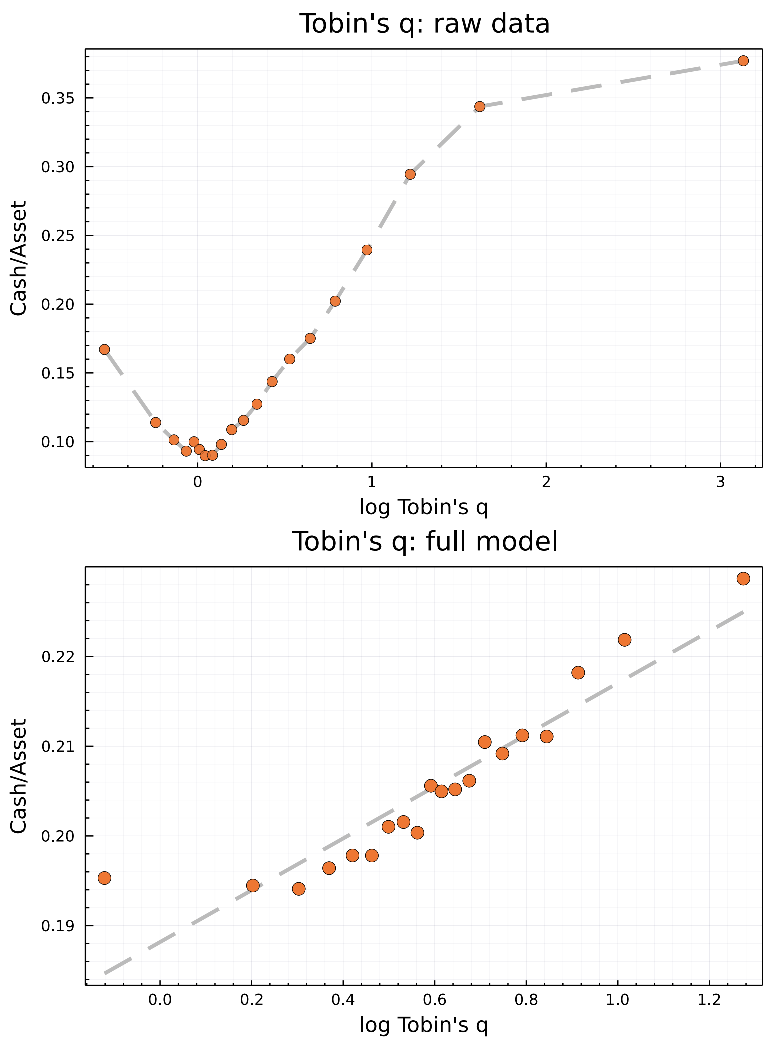
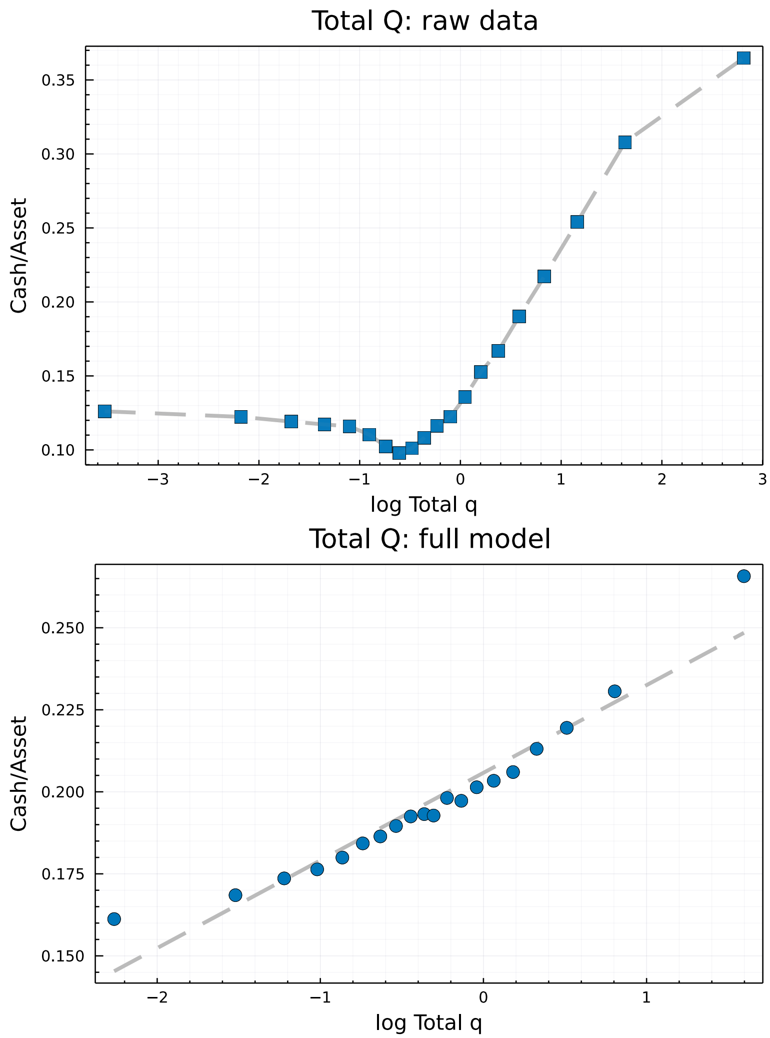


Figure A9: Tobin's q and corporate cash holdings: Binscatter plots



Notes: This figure presents the Binscatter plots between corporate cash holdings and different measures of q or markup. Main data source for this figure is the Compustat North American Annual data file. Firm-level cash-to-asset ratio is measured as the ratio of cash and short-term investments (Compustat series CHE) to firm's lagged total assets (Compustat series AT). Tobin's q is measured as the market value of firm's total assets divided by the book value of its assets. The market value of assets is computed as book value of assets (Compustat series AT) plus the market value of common stock (Compustat series $PRCC\_C$ times Compustat series $CSHO$) minus the book value of equity, where the equity book value is estimated as the sum of shareholder equity (Compustat series SEQ), deferred taxes (Compustat series $TXDB$), and investment tax credit (Compustat series $ITCB$), minus the value of preferred stocks (coalesce outcomes of Compustat series $PSTKRV$, $PSTKL$, and $PSTK$). Total q is obtained from Peters and Taylor (2017). Firm-level markup is estimated with Loecker, Eeckhout and Unger (2020)'s production cost function approach.

Figure A10: Total q and corporate cash holdings: Binscatter plots



Notes: This figure presents the Binscatter plots between corporate cash holdings and different measures of q or markup. Main data source for this figure is the Compustat North American Annual data file. Firm-level cash-to-asset ratio is measured as the ratio of cash and short-term investments (Compustat series CHE) to firm's lagged total assets (Compustat series AT). Tobin's q is measured as the market value of firm's total assets divided by the book value of its assets. The market value of assets is computed as book value of assets (Compustat series AT) plus the market value of common stock (Compustat series $PRCC\_C$ times Compustat series $CSHO$) minus the book value of equity, where the equity book value is estimated as the sum of shareholder equity (Compustat series SEQ), deferred taxes (Compustat series $TXDB$), and investment tax credit (Compustat series $ITCB$), minus the value of preferred stocks (coalesce outcomes of Compustat series $PSTKRV$, $PSTKL$, and $PSTK$). Total q is obtained from Peters and Taylor (2017). Firm-level markup is estimated with Loecker, Eeckhout and Unger (2020)'s production cost function approach.

Figure A11: Cash and misallocation

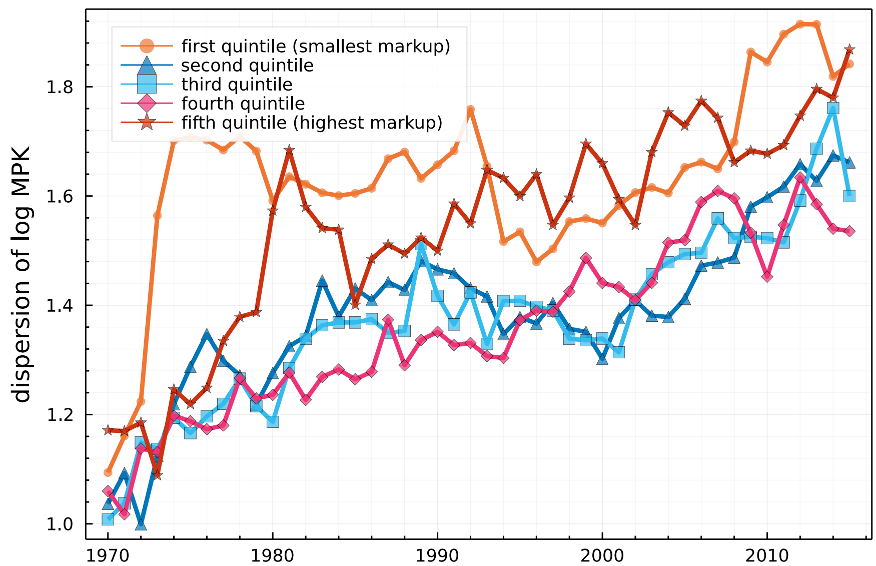
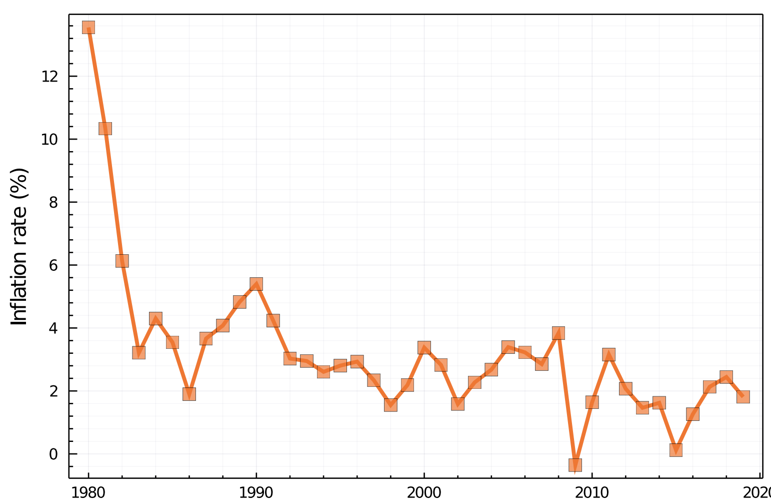


Figure A12: Historical inflation



Notes: This figure presents the time-varying inflation rate in the U.S.. Inflation rate is measured by changes in the consumer price index. The data is from World Bank and retrieved from FRED, Federal Reserve Bank of St. Louis.

Figure A13: Financing superstars

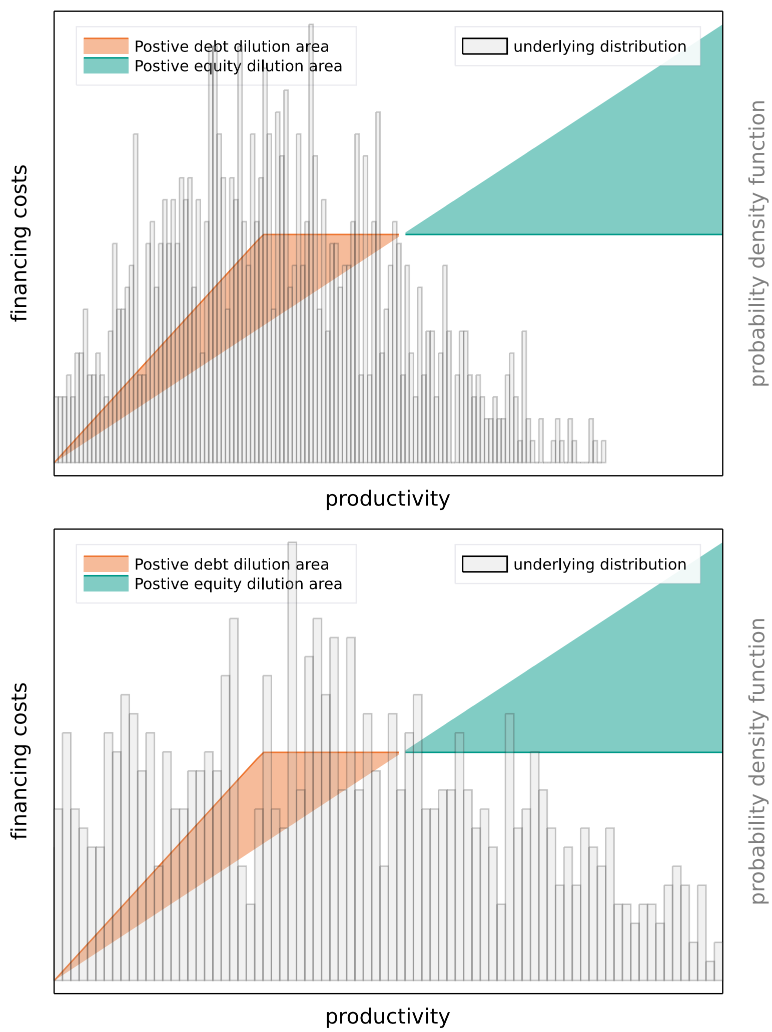


Table A1: Reduced-form evidence: Tobin's q, total q, and cash holdings

Panel A	**Cash/Asset**											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
log Tobin's q	0.032***	0.040***	0.035***	0.040***	0.030***	0.040***	0.034***	0.040***	0.039***	0.025***	0.058***	0.059***
	(43.631)	(81.418)	(75.085)	(77.681)	(56.976)	(80.571)	(45.755)	(79.956)	(71.954)	(21.982)	(49.698)	(50.905)
log Tobin's q square	0.002***									0.001***	-0.006***	-0.006***
	(14.186)									(3.397)	(-20.130)	(-20.658)
return of assets		-0.000*								0.000***	0.000***	0.000***
		(-1.739)								(4.767)	(3.009)	(3.446)
tangibility			-0.389***							-0.513***	-0.430***	-0.424***
			(-164.980)							(-119.984)	(-118.080)	(-116.121)
investment				-0.001***						0.001	0.010***	0.011***
				(-3.411)						(0.379)	(3.341)	(3.882)
size					-0.016***					-0.001	-0.009***	-0.009***
					(-46.307)					(-1.056)	(-36.048)	(-33.811)
profitability						-0.000				-0.000***	-0.000**	-0.000***
						(-1.447)				(-6.514)	(-2.367)	(-2.768)
R&D							-0.001***			-0.000***	-0.000	-0.000
							(-7.908)			(-5.766)	(-0.658)	(-1.330)
book leverage								0.000***		0.000	0.000	0.000
								(3.683)		(1.506)	(1.072)	(1.413)
payout									0.000***	0.000*	0.001***	0.001***
									(4.241)	(1.742)	(4.325)	(4.225)
	Fixed effects											
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		
Industry (naics3)											Yes	Yes
Industry × Year												Yes
N	288,949	288,189	283,855	268,911	288,949	282,698	134,581	288,949	243,667	117,935	119,293	119,293
Adjusted R^2	0.629	0.631	0.664	0.630	0.632	0.631	0.690	0.629	0.615	0.717	0.338	0.351

Panel B	**Cash/Asset**											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
log total q	0.0282***	0.0226***	0.0183***	0.0233***	0.0230***	0.0227***	0.0261***	0.0226***	0.0221***	0.0251***	0.0372***	0.0380***
	(93.252)	(79.124)	(66.282)	(80.539)	(80.829)	(79.198)	(67.004)	(79.124)	(71.265)	(58.662)	(79.109)	(81.575)
log total q square	0.0033***									0.0022***	0.0036***	0.0035***
	(52.885)									(31.480)	(36.758)	(36.955)
return of assets		0.0000								0.0004	0.0032***	0.0030***
		(0.440)								(1.596)	(10.341)	(10.067)
tangibility			-0.3349***							-0.4662***	-0.4045***	-0.3998***
			(-120.411)							(-91.428)	(-91.109)	(-89.794)
investment				-0.1048***						-0.0030	0.0773***	0.0779***
				(-25.495)						(-0.373)	(7.778)	(7.901)
size					-0.0128***					-0.0058***	-0.0097***	-0.0094***
					(-30.633)					(-9.335)	(-36.307)	(-35.356)
profitability						-0.0002***				-0.0007***	-0.0014***	-0.0013***
						(-2.942)				(-2.710)	(-3.671)	(-3.671)
R&D							-0.0142***			-0.0093***	0.0412***	0.0385***
							(-13.377)			(-6.692)	(24.713)	(23.301)
book leverage								0.0001***		0.0001*	-0.0002***	-0.0002***
								(4.899)		(1.748)	(-4.082)	(-3.474)
payout									0.0003	0.0015	-0.0007	-0.0013
									(0.161)	(0.599)	(-0.204)	(-0.388)
	Fixed effects											
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		
Industry (naics3)											Yes	Yes
Industry × Year												Yes
N	198,539	198,537	198,532	195,365	198,539	197,840	101,013	198,539	175,219	90,479	92,091	92,091
Adjusted R^2	0.652	0.647	0.673	0.649	0.649	0.647	0.706	0.647	0.629	0.733	0.359	0.374

Figure A14: SMM-MCMC parameters posterior distribution: traditional economy sub-sample (1980-1999)

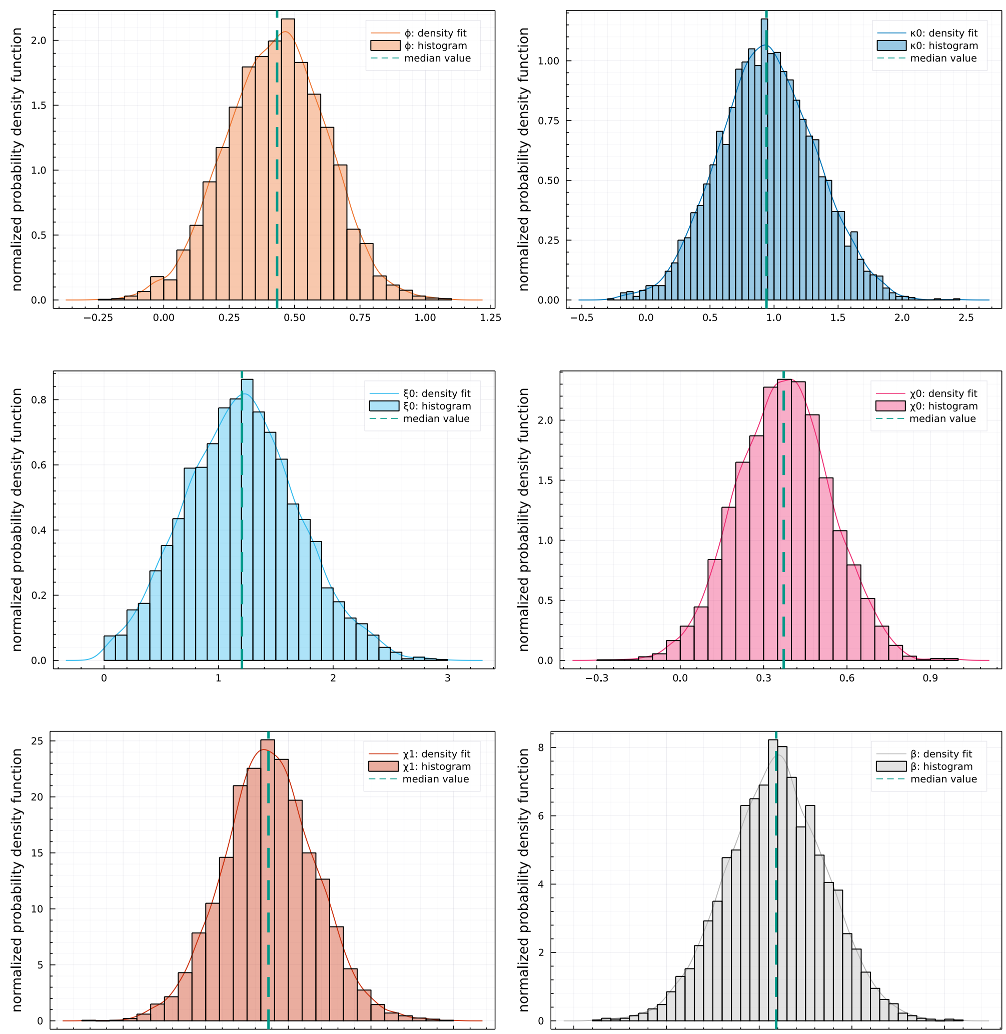
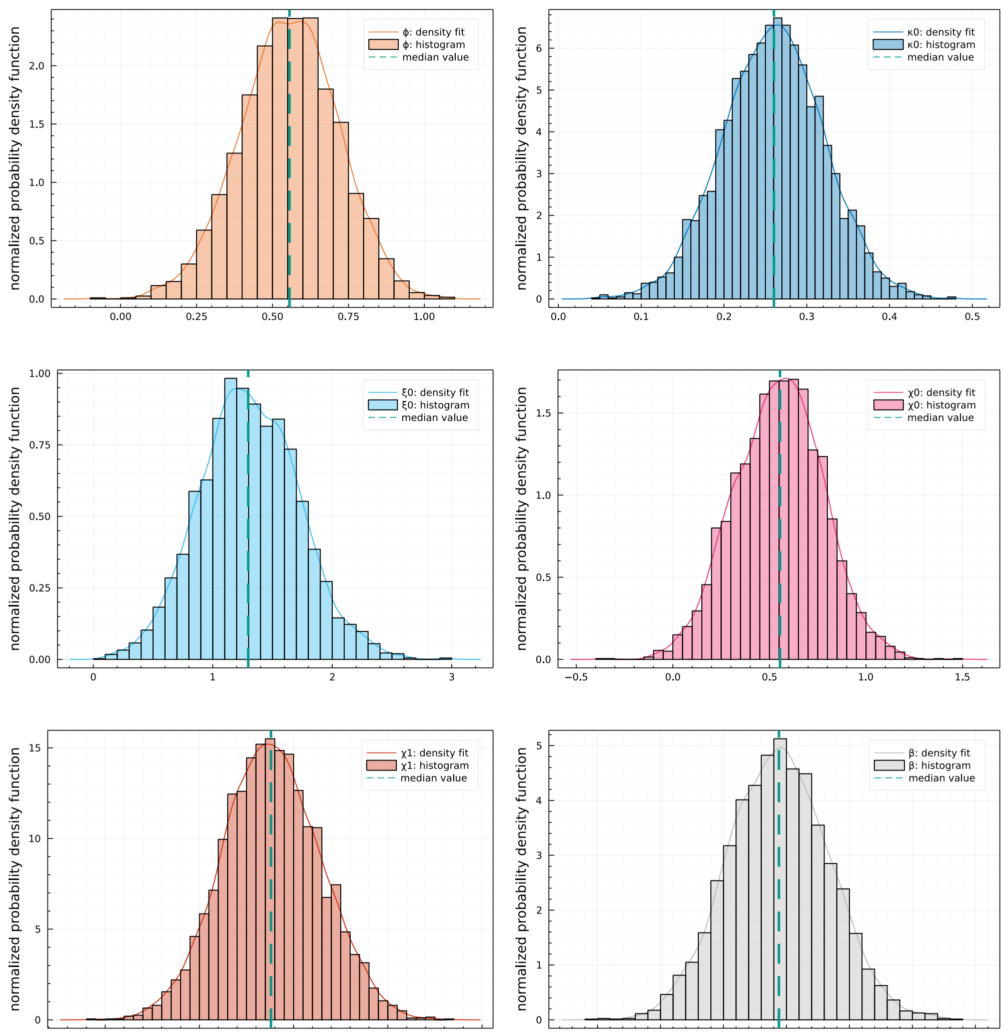


Figure A15: SMM-MCMC parameters posterior distribution: superstar economy sub-sample (2000-2015)



Appendix B

Appendix for Chapter 3

Figure B1: The rise of firms with negative net earnings: ten different industries

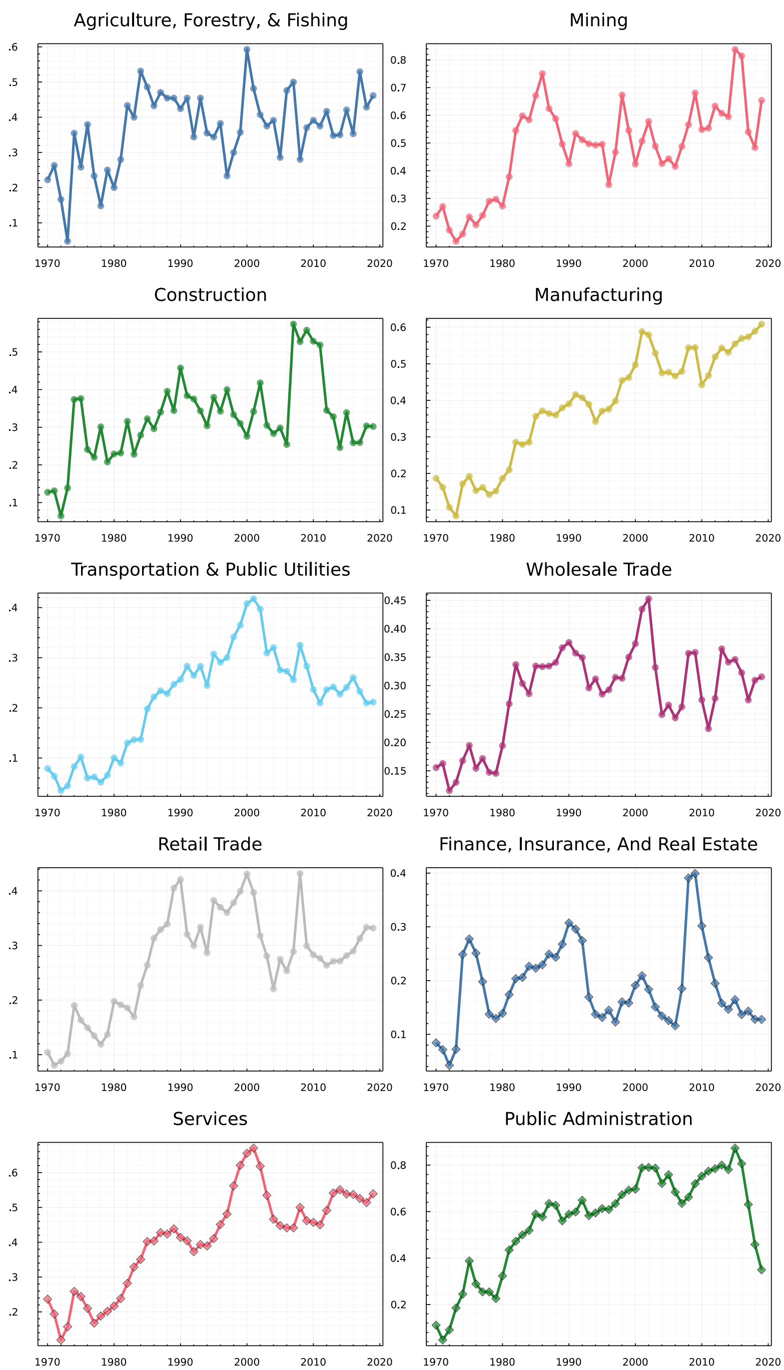


Figure B2: Average firm age

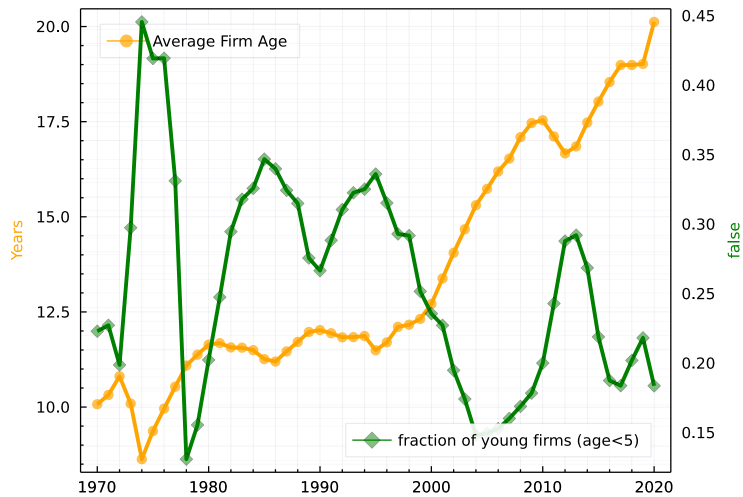


Figure B3: The rise of firms with negative net earnings: different stock exchanges

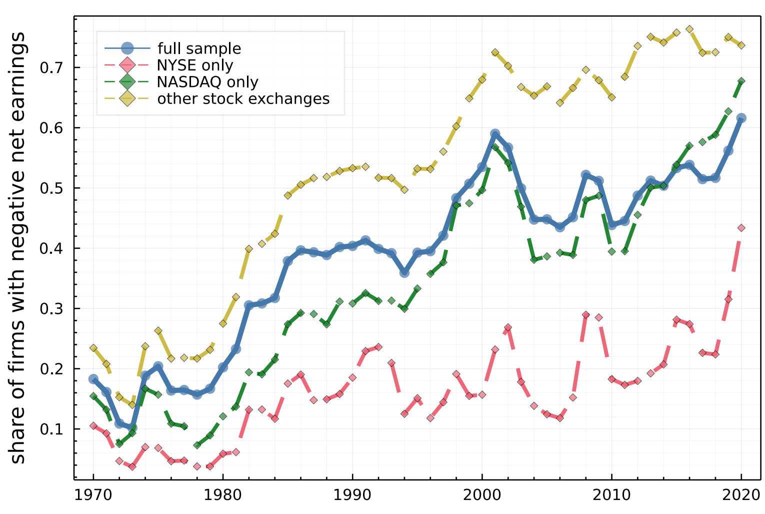


Figure B4: The rise of firms with negative net earnings: different percentiles

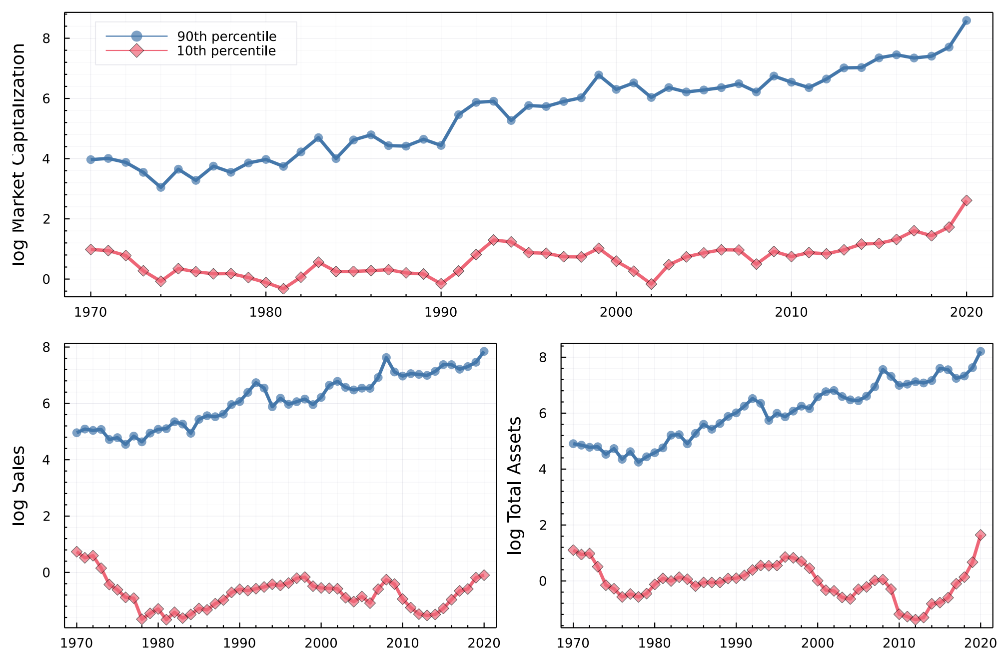
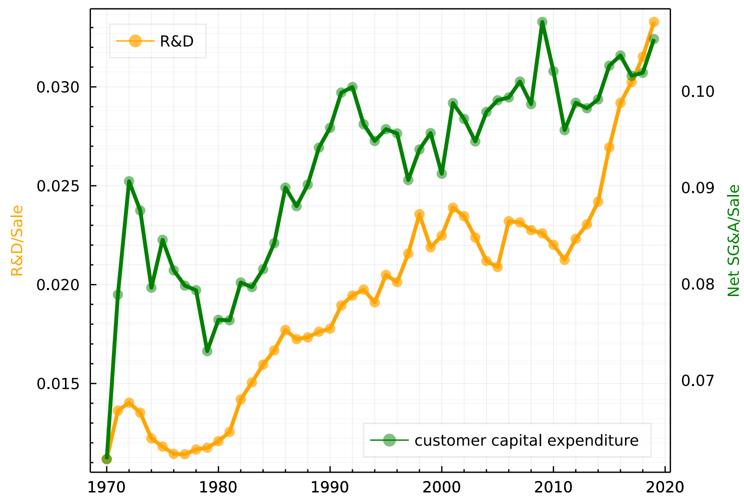


Figure B5: Changing business model: weighted by 3-digit industry sale share

(A) R&D and customer capital expenses



(B) production costs and investment

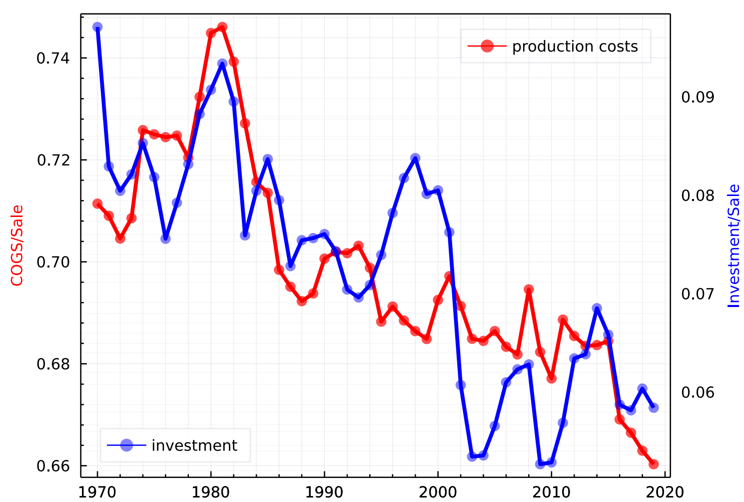
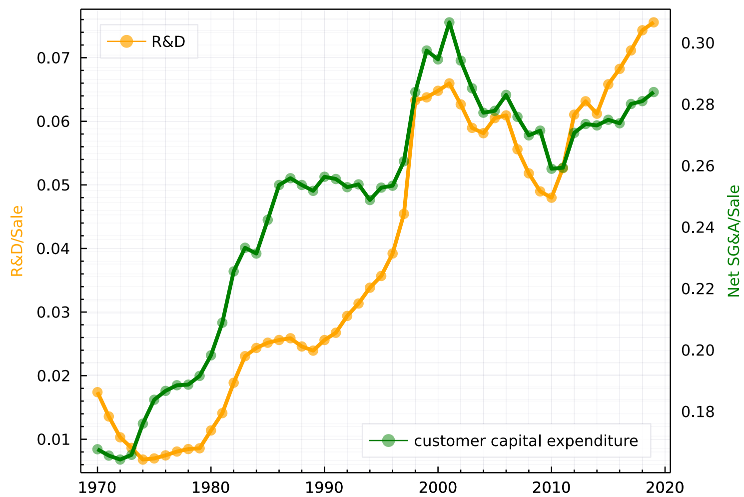


Figure B6: Changing business model: median average

(A) R&D and customer capital expenses



(B) production costs and investment

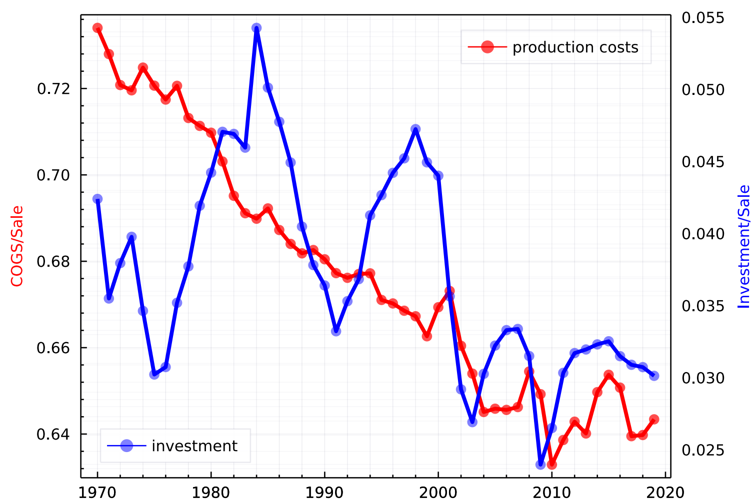


Figure B7: Examples: Amazon v.s. Walmart

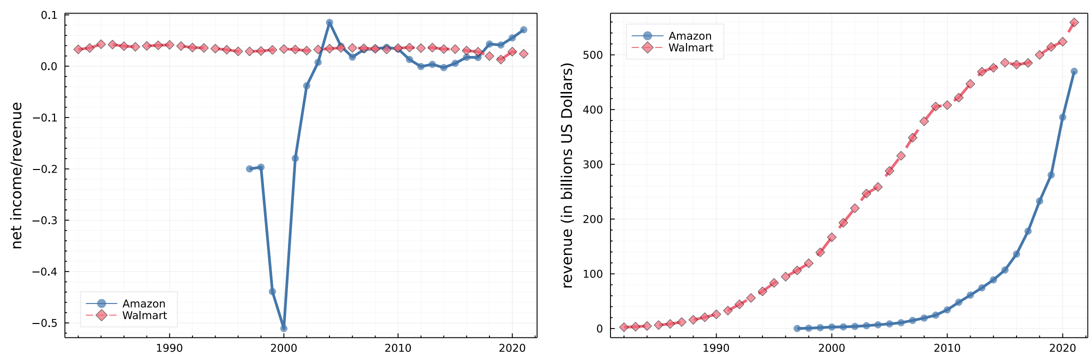
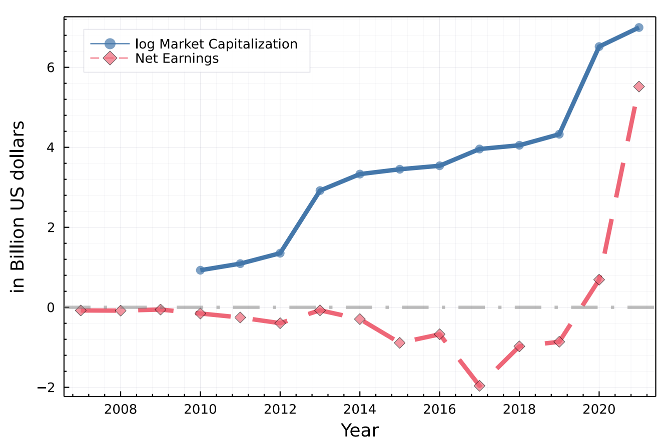


Figure B8: Example: Tesla



Appendix C

Appendix for Chapter 4

C.1 Proof

Proof of Lemma 4.1 and 4.2

To begin with, it can be easily shown that lenders will not verify if entrepreneur announces the good state as entrepreneurs have no incentives to announce the good state when the actual is a bad one.

There are several conditions on this optimal contract. First, the lenders must be break-even. As shown in Equation (4.13), the funding cost for investors is $(1+r)\,b$, and the expected return should be a weighted average of his payoff in the good state and his payoff in the bad state. Since there is no verification in the good state, investor's payoff is $z_G k - c_G$. However, in the bad state, there is a probability of q that investors will verify entrepreneurs' earnings and the cost of verification is assumed to be f. Therefore, his expected payoff in the bad state is $z_B k - q\,(c_B^v + f) - (1-q)\,c_B^{nv}$.

The second condition that an optimal contract needs to satisfy is that the entrepreneurs have no incentives to lie about the realized outcomes. Of course, entrepreneurs do not have any incentives to lie when it is a bad state. Entrepreneurs consumption in good state is always c_G. If he lies about the state and say it is a bad state. Then his expected consumption should be $(1-q)\,[(z_G - z_B)\,k + c_B]$.

The third condition that an optimal contract needs to meet is that the executions of contracts are feasible, which means all of c_G, c_B^v, and c_B^{nv} should be at least higher than or equal to zero. More importantly, p is a probability so it must lie between 0 and 1.

In this incomplete-collateralization case, the optimal verification probability is

$$q = \frac{(1+r)\,b - z_B k}{p\,(z_G - z_B)\,k - (1-p)\,f} \tag{C1}$$

when f is too large, q will be higher than 1 or even negative, which makes this earnings-based borrowing constraint infeasible. Therefore, in this paper, the micro-foundation for these two types of borrowing constraints is from the different earnings verification costs for banks and TechFin firms. As a result, they choose different ways of lending contracts, TechFin leads to the specialization-induced fragmentation in the financial services industry. In the following part of the paper, we will take these two types of borrowings as given, and investigate their macroeconomic implications.

Therefore, the expected profits of using these two types of lending can be shown as follows:

$$\pi = \begin{cases} (1+r)\,[pz_G k + (1-p)\,z_B k - (1-p)\,f] & \text{if} \quad \text{cash flow-based lending} \\ (1+r)\,lk & \text{if} \quad \text{asset-based lending} \end{cases} \tag{C2}$$

Therefore, if $f < f^* = \frac{pz_G + (1-p)z_B - l}{1-p} k$ or $l < l^* = pz_G + (1-p)\,z_B - (1-p)\,\frac{f}{k}$, then the lenders will strictly prefer cash flow-based lending over asset-based lending.

Proof of Lemma 4.3

The entrepreneur's wealth in banking sector evolves according to

$$da = (y - wl - \delta k - rb - c - \chi)\,dt = \left[(zk)^{\alpha} l^{1-\alpha} - wl - (r+\delta)\,k + ra - c - \chi\right] dt \tag{C3}$$

subject to the collateral-based borrowing constraint

$$k \leq \frac{a}{1 - \lambda_{\mathcal{B}}}$$

The first order conditions show that the optimal capital to labor ratio for firm with productivity z satisfies the following condition

$$\frac{l}{k} = \left(\frac{1-\alpha}{w}\right)^{\frac{1}{\alpha}} z \tag{C4}$$

Therefore, the firm's equilibrium profits can be written as

$$\pi = \alpha \left(\frac{1-\alpha}{w} \right)^{\frac{1-\alpha}{\alpha}} zk \tag{C5}$$

As the profit is a linear function of z, the firm's optimal choice on capital stock is a corner solution and meets the following condition

$$k_{\mathcal{B}}(a,z) = \begin{cases} \frac{a}{1-\lambda_{\mathcal{B}}} & z \geq \underline{z} \\ 0 & z < \underline{z} \end{cases} \tag{C6}$$

where $\underline{z} = \frac{r+\delta}{\xi}$ and $\xi = \alpha \left(\frac{1-\alpha}{w} \right)^{\frac{1-\alpha}{\alpha}}$.

Meanwhile, the firm's optimal debt holdings are

$$b_{\mathcal{B}}(a,z) = \begin{cases} \frac{\lambda_{\mathcal{B}} a}{1-\lambda_{\mathcal{B}}} & z \geq \underline{z} \\ -a & z < \underline{z} \end{cases} \tag{C7}$$

Equations (C6) and (C7) lead to the wealth growth for entrepreneurs in the banking sector as

$$da_{\mathcal{B}}(a,z) = \begin{cases} \left[\frac{(\xi z - r - \delta)a}{1-\lambda_{\mathcal{B}}} + ra - c - \chi \right] dt & z \geq \underline{z} \\ (ra - c - \chi)\, dt & z < \underline{z} \end{cases} \tag{C8}$$

Similarly, for TechFin entrepreneurs, we can derive the entrepreneur's optimal choice on capital stock, debt holdings, and the wealth growth rate.

Proof of Lemma 4.4

As shown in the previous lemma, the wealth follows a process of $da_j = [\Gamma_j(z)\, a_j - c_j]\, dt$ for the entrepreneurs in sector j. Therefore, the Bellman equation $\mathcal{V}_j$ should satisfy the following equation

$$\rho \mathcal{V}_j(t,a,z) = \max_{c_j} \left\{ \log c_j + \frac{1}{dt} E\left[d\mathcal{V}_j(t,a,z) \right] \right\} \tag{C9}$$

subject to the condition that $da_j = [\Gamma_j(t,z)\, a_j - c_j]\, dt$.

With the guess and verify approach, we can show that the optimal consumption choice is $c_j = \rho a_j$ for all entrepreneurs in the economy. Assume that the value function takes the form of $\mathcal{V}_j(t,a,z) = \mathcal{B}_j v_j(t,z) + \mathcal{B}_j \log a_j$. Then we have

$$E\left[d\mathcal{V}_j(t,a,z)\right] = \frac{\mathcal{B}_j}{a_j} da + \mathcal{B}_j E\left[dv_j(t,z)\right] \tag{C10}$$

Combining Equations (C9) and (C10) gives us the following equation:

$$\rho\mathcal{B}_j v_j(t,z) + \rho\mathcal{B}_j \log a_j = \max_{c_j}\left\{\log c_j + \frac{\mathcal{B}_j}{a_j}\left[\Gamma_j(t,z)\, a_j - c_j\right] + \mathcal{B}_j \frac{1}{dt} E\left[dv(t,z)\right]\right\} \tag{C11}$$

The first-order condition gives us $c_j = \frac{a_j}{\mathcal{B}_j}$. Substituting back in, we have

$$\rho\mathcal{B}_j v_j(t,z) + \rho\mathcal{B}_j log a_j = \log a_j - \log\mathcal{B}_j + \mathcal{B}_j\Gamma_j(t,z) - 1 + \mathcal{B}_j\frac{1}{dt}E\left[dv_j(t,z)\right]$$

which is

$$(\rho\mathcal{B}_j - 1)\log a_j = -\rho\mathcal{B}_j v_j(t,z) - \log\mathcal{B}_j + \mathcal{B}_j\Gamma_j(t,z) - 1 + \mathcal{B}_j\frac{1}{dt}E\left[dv_j(t,z)\right] \tag{C12}$$

Therefore, we can conclude that $\mathcal{B}_j = \frac{1}{\rho}$ for both sectors, and we have

$$c_j = \rho a_j \tag{C13}$$

$$da_j = \left[\Gamma_j(z)\, a_j - \rho\right] dt \tag{C14}$$

Finally, the value function is

$$\mathcal{V}_j(t,a,z) = \frac{1}{\rho}\left[v_j(t,z) + \log a_j\right] \tag{C15}$$

and $v_j(t,z)$ satisfies the following condition:

$$\rho v_j(t,z) = \rho\log\rho + \Gamma_j(t,z) - \rho + \mathcal{B}_j\frac{1}{dt}E\left[dv_j(t,z)\right] \tag{C16}$$

In addition, we also need to prove that $\Gamma_{\mathcal{F}}(z)$ is a convex function of z. This step is relatively easy as we only need to check the signs of first and second derivatives:

$$\frac{\partial \Gamma_{\mathcal{F}}(z)}{\partial z} = \frac{\xi\left[1-(r+\delta)\lambda_{\mathcal{F}}\right]}{\left[1-\lambda_{\mathcal{F}}\xi z\right]^{2}} > 0 \tag{C17}$$

$$\frac{\partial^{2} \Gamma_{\mathcal{F}}(z)}{\partial z^{2}} = \frac{2\lambda_{\mathcal{F}}\xi^{2}\left[1-(r+\delta)\lambda_{\mathcal{F}}\right]}{\left(1-\lambda_{\mathcal{F}}\xi z\right)^{3}} > 0 \tag{C18}$$

The above two equations are both positive because $1-\lambda_{\mathcal{F}}\zeta z > 0$ and $z > \underline{z} = \frac{r+\delta}{\xi}$.

Therefore, the wealth growth rate in the TechFin sector $\Gamma_{\mathcal{F}}(z)$ is a strictly convex function of productivity z.

C.2 A Note on Lian and Ma (2021)

Lian and Ma (2021) document the prevalence of cash flow-based lending. They argue that 20% of debt by value is based on tangible assets, whereas 80% is based predominantly on cash flows from corporate operations. Here I want to argue that their main conclusion, especially the dominating use of cash flow-based lending, is not robust. A better and less controversial way of interpreting the empirical result is the co-existence of earnings-based and collateral-based borrowing constraints.

Graph (A) in Figure C1 replicates one of their main results in the paper. To be clear, all the raw data used in this section are directly obtained from Quarterly Journal of Economics Dataverse.[1] In their paper, Lian and Ma (2021) argue that "Figure I, Panel A, shows that the median share of asset-based and cash flow–based lending among large nonfinancial firms is generally less than 20% and slightly over 80%, respectively, in recent years." The key words in their original statement are **large** and **median**. More specifically, when they prepare the data for this graph, first they classify all the firms in Compustat dataset into five different groups according to their total asset levels. Then they drop the bottom 20% firms out of the sample. Finally, they compute the *median* share of asset-based and cash flow–based lending. As we can see from the replicated result in Graph (A), the median share of asset-based lending on average is 17.8%, while that of cash flow–based lending is 77.2%.

To begin with, I want to point out that these numbers are sensitive to the choice of sub-samples and the use of median. In Graph (B) of Figure C1, I plot the same results but without dropping the smallest firms out of the sample. In Graph (C) of Figure C1,

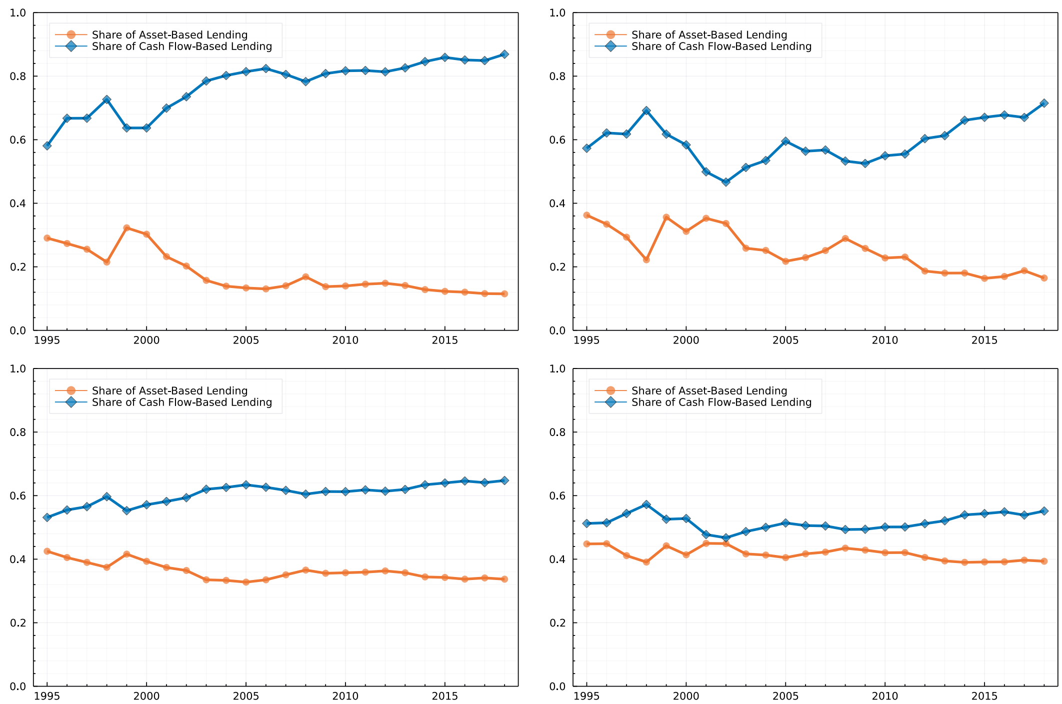
[1] Replication data and codes for Lian and Ma (2021) can be downloaded from here.

I drop all the firms in the lowest quintile but use mean instead of median. In Graph (D), I include all the firms and use mean to calculate the average value. As we can see from these graphs, whether cash flow-based lending is really prevalent depends on the specific choice of our empirical measure. For example, in Graph (D), the average use of cash flow-based lending is 51.7% while the average use of asset-based lending is 41.6%. In this way, both types of lending are important financial frictions in the real economy.

The subsample selection is not the most problematic issue in their work. In fact, Lian and Ma (2021) do mention this point. They find that for small firms, asset-based lending is more common and the median value of asset-based lending among these small firms is roughly 54%.

The real problem comes from the use of median because the actual distribution of the borrowing constraints is a **bimodal** one. It is true that both median and mean can be interpreted as the "representative" value for the data sample, and sometimes the median is used as an alternative to the mean. However, if the underlying distribution is a bimodal one, both indicators can be misleading, as there is no such a representative borrower in the data. Graph (A) and (B) in Figure C2 present the distribution of individual firm's use of asset-based and cash flow-based lending, respectively. As we can see from these two graphs, when we attempt to describe the use of borrowing constraint by individual firm, there is no such a representative firm in this economy because some firms rely heavily on cash-flow based lending while the other firms use more collateral-based lending. The detailed breakdown for each year throughout the data sample period can be shown in Figure C3 and C4. Generally speaking, the less controversial way of describing the reality is the co-existence of two types of borrowing constraints.

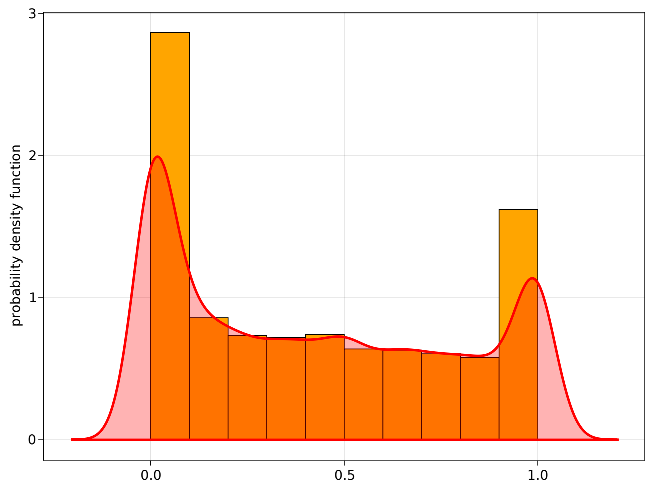
Figure C1: Anatomy of corporate borrowing constraints



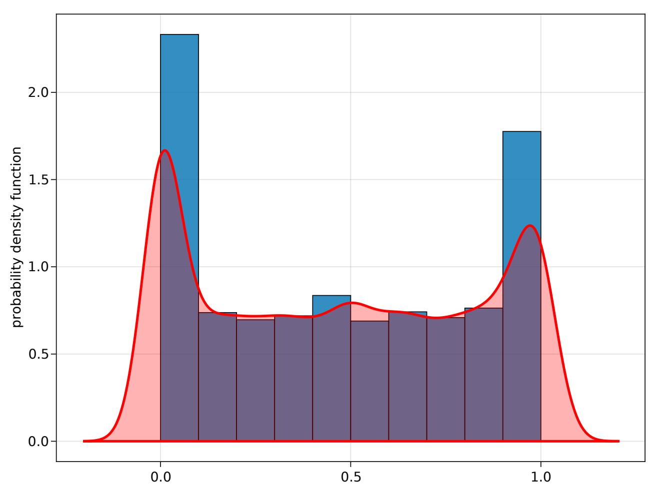
Notes: This figure presents anatomy of corporate borrowing constraints with different choices of summarizing the data. Left above represents the original plot with large firms only and median measure. Right above is the plot with all firms and median measure. Left below represents the plot with large firms and mean measure. Right below is the plot with all firms and mean meausre. Main data source for this figure is obtained directly from the replication package for Lian and Ma (2021).

Figure C2: Distributions on the types of borrowing constraints

(A) distribution of the share of asset-based lending



(B) distribution of the share of cash-flow-based lending



Notes: This figure presents the distributions of inividual firm's use of two types of lending. Main data source for this figure is obtained directly from the replication package for Lian and Ma (2021). Orange and blue rectangles represent histogram distributions with normalized probability density. Red lines are the Kernel smoothing function fits.

Figure C3: Asset-based lending distribution in each year

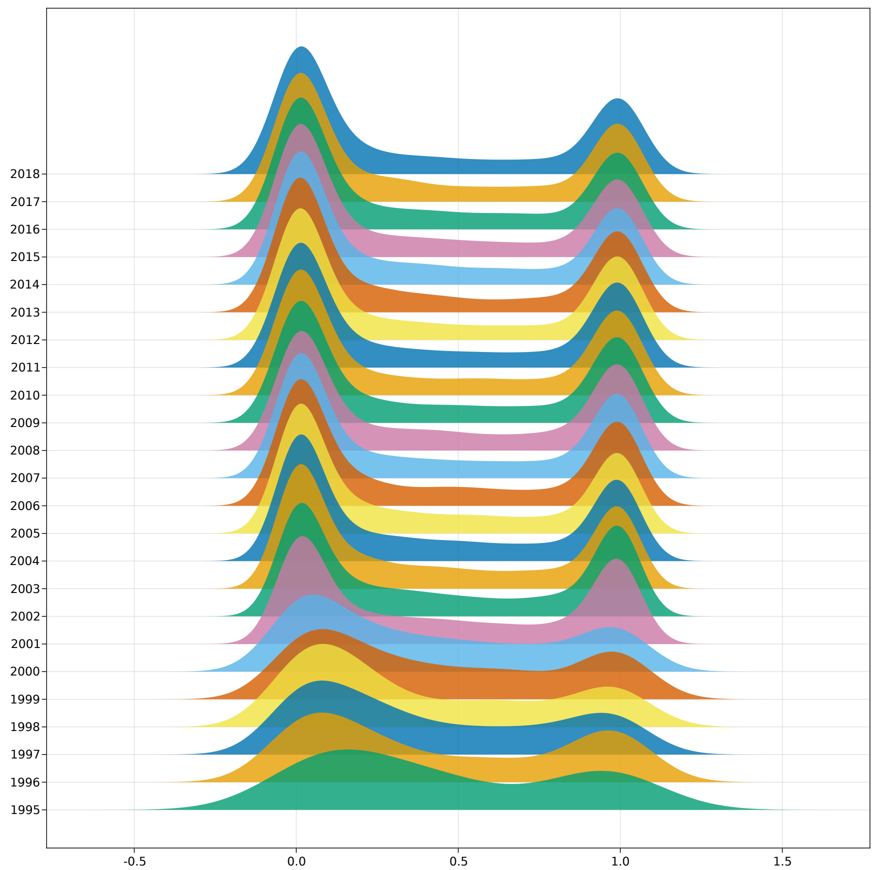
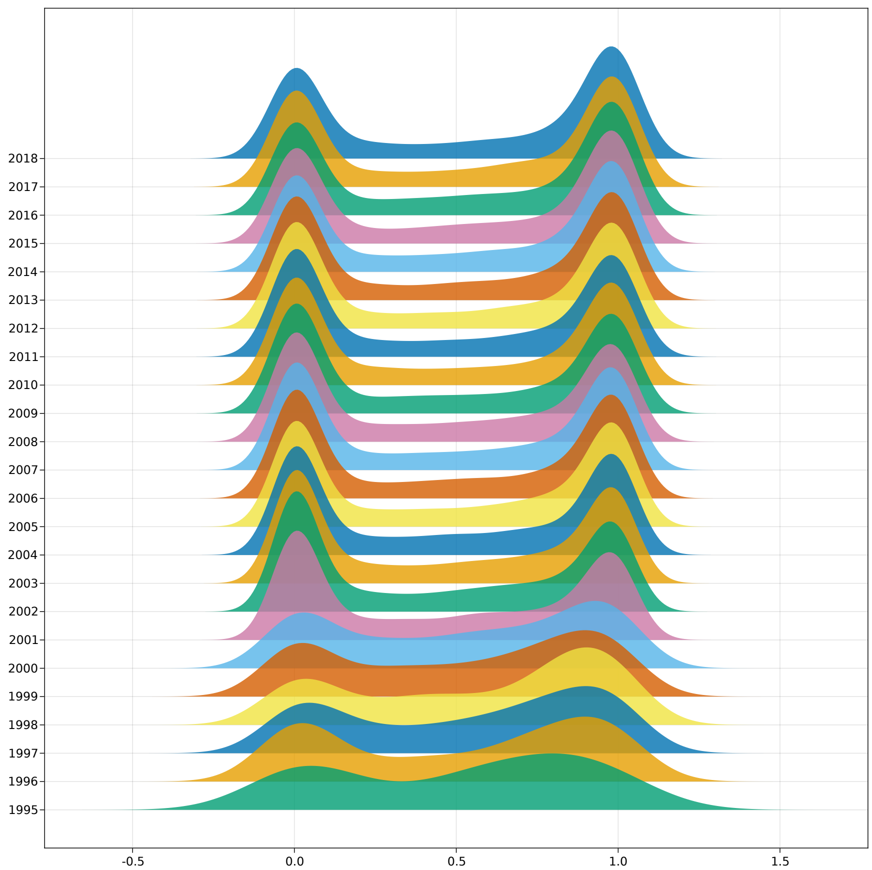
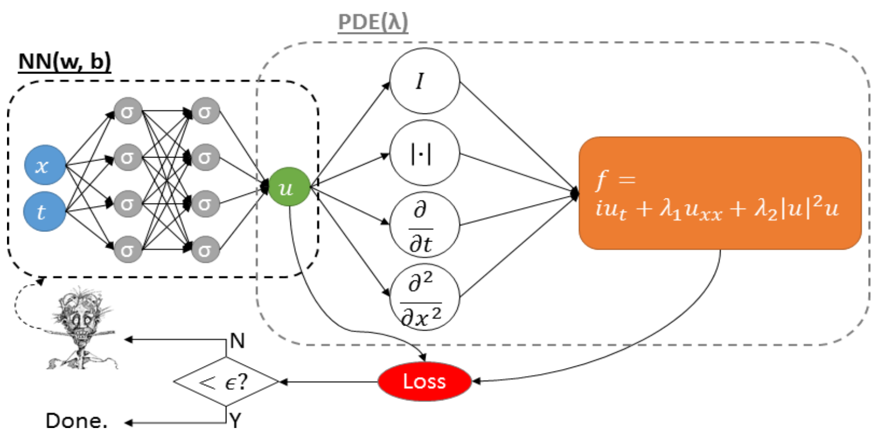


Figure C4: Cash flow-based lending distribution in each year



C.3 Physics-Informed Neural Networks Algorithm

Physics-Informed Neural Networks (PINN) algorithm is proposed by Raissi, Perdikaris and Karniadakis (2019) and represents a new deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. The basic idea of this algorithm can be summarized as in the following graph:



Generally speaking, the idea of PINN is to employ two or more neural networks that share the same parameters. In addition, the objection function is to minimize the sum of mean squared errors of original neural network and those of partial derivatives. In this way, we can make full use of the synergy between machine learning and classical computational physics to solve some high dimensional partial differential equations without encountering the curse of dimensionality. More importantly, this approach is feasible because the PINN approximation theorem guarantees that feed-forward neural nets with a sufficiently large enough number of neural nodes can simultaneously and uniquely approximate any partial differential equations and their partial derivatives.