

# Rise of Superstar Firms and Fall of the Price Mechanism\*

Dan Su<sup>†</sup>

University of Minnesota

## Job Market Paper

### Abstract

This paper investigates the impacts of superstar firms and their cash hoarding behaviors on capital misallocation through the lens of an endogenous firm-market boundary. Empirically, I find that since the 1980s, capital allocative efficiency has been deteriorating in the United States. To explain this aggregate trend, I introduce product market competition and corporate risk management into a standard continuous-time heterogeneous agent model with incomplete markets. In this framework, entrepreneurs face non-homogeneous, quality-based earnings processes, and [Coase \(1937\)](#)'s firm-(financial) market boundary exists in general equilibrium. The price mechanism is bounded by corporate internal financing as there is no market to equalize the marginal value of internal resources across firms. Shifts in supply and demand curves not only generate a “winners-take-most” phenomenon, but also make current winners inherently riskier and cause them to rely less on external financing. Accordingly, the increasing corporate self-financing moves the firm-market boundary and thus *increases* misallocation. Parameterizations of the model consistent with the data suggest that the area disciplined by the price mechanism, measured as the wealth-weighted share of firms using external financial market, has declined by roughly 11% over the past forty years.

*Keywords:* firm-market boundary; misallocation; superstars; corporate risk management

*JEL codes:* D21; D25; D33; E14; E44; G30; L20; O33

---

\*This version: December 26, 2021. First draft: September 2020. This paper is a substantially revised draft of my earlier work titled “Shrinking Boundary of the Invisible Hand”. For their constructive comments and suggestions, I thank Anmol Bhandari, Murray Frank, Robert Goldstein, Ramin Hassan, Gerard Hoberg, Yunzhi Hu, Xiang Li, Yu Li, Yucheng Liang, Xiaoji Lin, Ellen McGrattan, Bruno Pellegrino, Juliana Salomao, Xiaomei Sui, Michal Szkup, Colin Ward, Mindy Xiaolan, Yu Zheng, and other participants in Minnesota finance brownbag seminar, Minnesota workshop in macroeconomics, Western Finance Association, Society for Economic Dynamics, Econometric Society European Summer & Winter Meeting, Money, Macro and Finance annual conference, Washington University in St. Louis Economics Graduate Student Conference, and Colorado Finance Summit. Of course, all errors are my own. Please click [here](#) for the latest version and [here](#) for the online appendix.

<sup>†</sup>Affiliation: Department of Finance, Carlson School of Management, University of Minnesota. Postal address: Office 4-267, Carlson School of Management, 321 19th Ave S, Minneapolis, Minnesota, USA, 55455. Email address: [suxxx478@umn.edu](mailto:suxxx478@umn.edu); [stanleydansu@gmail.com](mailto:stanleydansu@gmail.com)

“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*

In the past several decades, superstar firms have become increasingly relevant in many different economic aspects. For example, on the production side, industry sales have over time been concentrated in a small number of firms (Autor et al., 2020). At the same time, on the financing side, some large companies, including the tech giants, hold billions of dollars of cash on their balance sheets.<sup>1</sup> In this paper, I attempt to investigate the impacts of these superstar firms and their cash hoarding behaviors on misallocation from an endogenous firm-market boundary perspective.

The motivation fact is that since the 1980s, capital allocation efficiency has been deteriorating in the United States. As will be shown in Section 1, three empirical findings support this claim. 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 less and less finance flows to productive firms. Third, the positive gap between the marginal product of capital and the real interest rate increases over time. These facts are 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.<sup>2</sup>

Throughout this paper, I argue that the rise of (superstar) firms and changes in corporate risk management policies are important drivers of these aggregate trends. The intuition is based on Coase (1937)’s seminal idea that the nature of the firm is a substitution of the price mechanism. 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

---

<sup>1</sup>According to what rating agency Moody’s said in a recent report, in 2020, the total cash holdings of U.S. non-financial companies have reached a level of 2.12 trillion dollars, which accounts for roughly 10% of total U.S. real GDP. The top five cash-abundant companies – Apple, Alphabet, Amazon, Facebook, and Microsoft – account for 27% of the total U.S. corporate cash balance in 2020.

<sup>2</sup>These findings are consistent with Hsieh and Klenow (2018)’s *reallocation myth* 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).

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. Therefore, the emergence of superstar firms will exaggerate the competition between firms and the market for allocating resources, which eventually deteriorates the aggregate capital allocation efficiency.

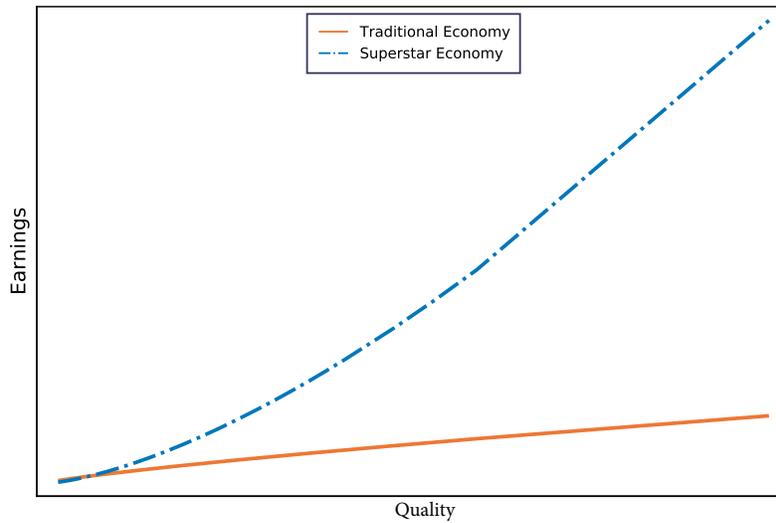
Guided by this intuition, I provide a theoretical and quantitative framework linking economic fundamental changes, endogenous firm-market boundary, and capital misallocation. More specifically, I introduce product market competition and corporate risk management into a standard heterogeneous agent model with incomplete markets. The model is written in continuous time and has three main conclusions. First, some economic fundamental changes can lead to both income and risk redistribution in the economy. One example is digitalization. The availability of the internet allows firms to drastically increase their scale of operations, which makes it possible for high-quality firms to become superstars. As shown in Figure 1, this type of economic fundamental change is redistributive in income and wealth. Compared to the low-concentration traditional economy, in a Superstar Economy, a relatively small number of talented people can obtain enormous earnings and dominate the activities in which they engage. This static income redistribution effect has been widely recognized in the existing literature (e.g., [Sattinger, 1993](#); [Tervio, 2008](#); [Scheuer and Werning, 2017](#)). However, when entrepreneurs face uncertainty in their ability, from a dynamic perspective, superstars are inherently risky. A small variation in ability can translate into considerable earnings fluctuations, especially on the right-tail side; in other words, these economic changes are also redistributive in risks. They generate a “winners-take-most” phenomenon but also make current winners bear more volatile income fluctuations in the future.

Second, this risk-redistribution nature of economic fundamental changes will substantially and heterogeneously affect corporate risk management policy. Firms finance their activities with current profit flows, external debt and equity, and internal cash. When external finance is costly, firms demand cash so that they can lower the probability of funding their operations with costly external finance. Such an incentive is stronger for firms with more volatile earnings processes. Although this conclusion is widely documented in the corporate finance literature (e.g., [Riddick and Whited, 2009](#))<sup>3</sup>,

---

<sup>3</sup>According to the “trade-off” theory (e.g., [Miller and Orr, 1966](#)), the optimal cash holdings should be determined by the benefits and costs of holding liquid assets. The main costs of holding cash include opportunity costs and costs from agency frictions. The benefits of holding cash include lowering the transaction costs associated with raising external funds or liquidating assets. Therefore, firms demand more cash when their expected future earnings are more volatile.

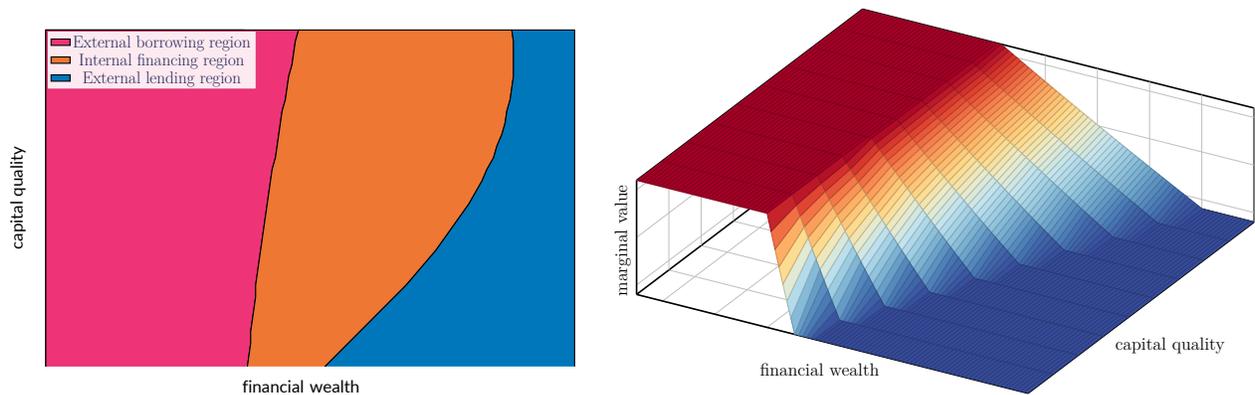
**Figure 1: Risky Superstar Economy**



its macroeconomic implications have not yet been investigated. As shown in the left graph in Figure 2, if we allow the entrepreneurs to optimally choose their reliance on external financing, then there will be an endogenous firm-(financial)market boundary as suggested in [Coase \(1937\)](#). The existence of these cash control boundaries will split the economy into three different regions. Some agents in this economy are borrowing from the market as they have low net financial wealth, while some agents with relatively high net financial wealth are lending to the market. Between these two regions, there is an internal financing region where companies internally save their financial wealth. Of course, in the model, this internal saving region is pinned down by the optimal choices made by all the individual entrepreneurs. When technical development such as digitalization transfers the society from a traditional economy to a Superstar Economy, the size of this internal financing region will expand as corporate earnings become more volatile. This incentive is stronger for superstars as their expected future earnings are becoming much riskier.

Third, the expansion of the self-financing region will increase capital misallocation in this economy. According to the right graph in Figure 2, this internal financing region is entirely different from the external financing region, as the marginal cost of financing will not be equalized when companies save resources internally. The intuition is as follows: If investment is financed externally, then resource allocation is still based on the *price system*; a frictionless capital market will equalize all the difference in investment returns. However, if investment is financed internally, then resource allocation now primarily depends on *entrepreneurs*: the marginal product of capital does not need to be equal across dif-

**Figure 2: Boundary of the Invisible Hand**



ferent firms as there is no such market to equalize the marginal value of internal resources. Therefore, the existence and movement of this firm-market boundary have an interesting implication: different from the common wisdom that self-financing can undo misallocation (e.g., [Midrigan and Xu, 2014](#)), in this paper, an expansion of the internal financing region lowers the aggregate capital allocation efficiency.

To follow, I implement some reduced-form empirical investigations on whether some testable predictions derived from my theory actually hold in the data. The 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 not only have on average the largest degree of capital misallocation, but they have also experienced the fastest increase since the 1970s. 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 the main thrust of the paper.

Finally, I investigate the model's quantitative implications. I estimate the structural parameters of the model through the simulated method of moments (SMM) approach. When taken to data, the model is able to match both the aggregate trends and cross-sectional patterns. In addition, my quantitative analysis suggests that the most important driver behind these aggregate micro-finance trends is the change in production technology. This type of technical change generates an increased fixed cost component and decreased marginal supply cost for companies, which fundamentally changes the earnings-quality relationship in the economy. Finally, my model suggests that the area disciplined by the price mechanism has declined by approximately 10.88% during the past four decades.

**Contributions** The contributions of this paper are fourfold. First, I 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 papers focus on the markup level channel. In contrast, this paper highlights the markup risk channel of economic fundamental changes. Second, this paper provides a different finance view on the origins of misallocation. 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, I formally establish [Coase \(1937\)](#)'s firm-(financial)market boundary in general equilibrium. More specifically, I 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. Fourth, I extend the  $R - g$  framework on inequality to allow for the existence of two types of capital returns (two  $R$ s). 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 5](#).

**Layouts** The paper is organized as follows. In [Section 1](#), I summarize the facts that motivated the entire paper. [Section 2](#) describes the model setup and provides some preliminary discussions on the underlying mechanism. [Section 3](#) presents the reduced-form evidence for the key prediction derived from theory. [Section 4](#) shows the quantitative results, including the cross-sectional implications and the aggregate effects of the scale-related technical change. In [Section 5](#), I review the existing literature and also discuss the validity of two important assumptions used in this paper. [Section 6](#) offers a conclusion.

## 1 Facts

I start by presenting several aggregate macro-finance trends over the past forty years. First, the U.S. has experienced significant declines in its capital allocation efficiency since the 1980s. Second, the society has been transitioning into a new and risky Superstar Economy, where both income and risk concentrations have increased substantially. Third, compared with their counterparts of 40 years ago,

firms today have become heavily reliant on internal financing and invest less. This paper aims to provide a unified quantitative framework to account for all these macro-finance trends, as studying each of them in isolation may generate misleading interpretations.

The data used in this section comes primarily from the Compustat North America Fundamentals dataset. Explanations on variable constructions in detail are presented in Appendix B. In this section, I will focus on documenting these facts. Exhaustive interpretations and discussions on the related literature will be provided in the following quantitative analysis section.

## 1.1 Deteriorating Efficiency of Capital Allocation

### 1.1.1 Increasing “misallocation”

First, I investigate the dynamics of capital misallocation among U.S. public firms. The definition of misallocation here is the dispersion in the static marginal product of capital. Following [David, Schmid and Zeke \(2019\)](#), I 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, I 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 A8 and A9 in the appendix, using alternative measures (e.g., including intangible capital) yields similar patterns.<sup>4</sup>

The solid orange line in panel A of Figure 3 presents the time-series plot of the baseline measure on misallocation. For this full sample, each year, I 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.

---

<sup>4</sup>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 I didn’t 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.

One possible concern for my baseline result is that I 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, I only include the top 2000 firms according to their market capitalization. Then for this subset of firms, I calculate the degree of capital misallocation using the same approach as before. Again, Figure 3 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 3, I 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, I create a balanced panel where I only include the firms that are continuously present within this decade. For each year, I 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 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.

### 1.1.2 Eclipse of the public market

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. Here I investigate a simple and related 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 4, where I calculate the annual cross-sectional correlation between the firm's net finance dependence and its productivity. In the baseline result in panel A, I 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 4](#), 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 I measure firm-level productivity. In panel B, I 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 Akerberg-Caves-Frazer ([Akerberg, 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.<sup>5</sup>

The pattern documented in [Figure 4](#) 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](#)). My 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 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.

---

<sup>5</sup>In the baseline analysis, I 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.

### 1.1.3 MPK $\gg$ $r$

The third piece of evidence 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 5 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).<sup>6</sup> The development approach adopts [Feenstra, Inklaar and Timmer \(2015\)](#)'s method and the Penn World Table dataset.<sup>7</sup> 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.

The reason I 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 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 ( $\delta$ ) and risk premium (RP). Therefore, the total cost of capital in this case should be  $r + \delta + RP$ . I plot the time series of this adjusted return-cost gap in Figure A15 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.

---

<sup>6</sup>[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.

<sup>7</sup>[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.

## 1.2 Transitioning into a Risky Superstar Economy

Over the past several decades, there has been a sharp rise in income inequality among many different professions and industries. Some studies find that many advanced economies have become increasingly concentrated, and a few firms have become more and more dominated. For instance, using the U.S. Census panel dataset, [Autor et al. \(2020\)](#) find a secular upward trend in industry concentration over time, and the concentration in sales is stronger than the concentration in employment. Meanwhile, [De Loecker, Eeckhout and Unger \(2020\)](#) document the rise of market power by using the U.S. Compustat dataset, and find that average markup in the 1980s was only 21% above marginal cost. However, this number has increased to 61% today.

Here I want to show that income and risk concentrations are *both* increasing in the U.S.. This medium-run trend in risk concentration is new to the literature and essential to this paper's story. In [Figure 6](#), I present the time series plot of two Herfindahl-Hirschman indexes: one for the level of sales and the other for the volatility of sales. The level-HHI index captures the dynamics of income concentration, while the volatility-HHI index measures how risk concentration changes over time. Both of them are constructed with the Compustat Quarterly dataset. The way I construct them is the following. First, for each firm in each year, I compute the mean and standard deviation of its quarterly sales. Second, I calculate the ratios of the firm's sales and volatility to total industry sales and volatility for every firm in a given industry-year. Following the existing literature, an industry is defined according to NAICS three-digit industry classification. Third, I square these obtained ratios. Finally, I sum up the squared ratios across all firms in a given industry-year. In this way, the sale's mean-HHI and volatility-HHI at the industry level can be obtained. To aggregate the index across industries, I use a sales-weighted approach, where the relative importance of industry sales in the whole economy determines the weights. The methodology here follows the work of [Grullon, Larkin and Michaely \(2019\)](#) and many other related works.

As shown in the top panel in [Figure 6](#), the increasing concentration patterns are evident in the data. Both income and risk concentrations remain relatively stable in the 1980s, but these two indicators have risen, especially post-2000. In terms of the magnitudes, at the beginning of the 1980s, sales' level-HHI was about 0.136. By the end of 2019, this number had increased to 0.201. The same pattern holds for the HHI in sales volatility. In 1980, the degree of risk concentration was only at the level of 0.130. However, in 2019, the value of volatility-HHI had increased to 0.219. In a nutshell, since the

1980s, income concentration and risk concentration have surged almost 48% and 68%, respectively.

At the same time, in addition, in panel B, I plot the kernel density of the sales distribution in 1980 and in 2019. A similar figure for the distribution of sales volatility is provided in panel C. Comparison of these two distribution plots reveal that the rise in concentration, especially for the sales level distribution, is exclusively driven by the increase in the upper percentiles. This result indicates that more firms have incredibly large sales in today's economy and that they dominate the economic activities. This shift in the distribution of sales is similar to the pattern documented in [De Loecker, Eeckhout and Unger \(2020\)](#), which focuses on the changes in the distribution of markup. In Figure A14 in the appendix, I follow [De Loecker, Eeckhout and Unger \(2020\)](#) and replicate their results on markup distribution in both the traditional and the new economy. Consistent with the results on sales distribution, there is also a substantial reallocation of economic activity towards firms with extremely high markups.

### 1.3 The Rise of Internal Financing

During the same time that the income and risk concentrations are seeing increases, companies have substantially changed their risk management policies and accumulated a massive amount of cash. Although this paper provides a different explanation, the focus here is more on the macroeconomic consequences of corporate internal financing. In this section, I first document these empirical patterns.

Following the seminal work done by [Bates, Kahle and Stulz \(2009\)](#), in the top panel of Figure 7, I present the time-series plot of both the median and aggregate cash-to-asset ratios among the U.S. public firms. Intuitively speaking, corporate cash represents the profits retained by firms instead of being paid out as dividends. As shown in the data, in the 1980s, the average cash ratio is about 6%. In contrast, the ratio has become more than doubled. The average number now is about 12% (for the aggregate cash-to-asset ratio) or 15% (for the medium cash-to-asset ratio). Therefore, compared to their counterparts of 40 years ago, companies today have accumulated a significant amount of cash on their balance sheet.

In addition, the increase in corporate cash-to-asset ratio has much exceeded that in corporate capital-to-asset ratio. In a similar way, in the bottom panel of Figure 7, I present the time-series plot of the difference between (median or aggregate) cash-to-asset ratio and capital-to-asset ratio. These

two indicators represent the changes in corporate preference for cash over capital. As this graph makes clear, in the 1980s, on average, the difference between cash-to-asset ratio and the capital-to-asset ratio is approximately -3.3% (for the aggregate measure) or 0.4% (for the median measure). However, this gap has been substantially increasing over time. In the 2010s, the average difference between cash and capital ratio becomes 7.4% (for the aggregate measure) or 9.7% (for the median measure). Therefore, corporate saving has not only increased significantly over the past four decades but also far exceeded the needs of long-term capital that is vital to the corporate business.

## 1.4 International Evidence

These macro-finance trends are not unique to the U.S.. In Appendix C, I conduct similar exercises for some other developed economies, including Canada, France, Germany, Japan, South Korea, and the United Kingdom. The main conclusion from Appendix C is that over the past several decades, developed economies have experienced similar changes in *some* critical macro-finance indicators. At the same time, they also differ from each other in certain trends. Exploring the similarities and differences across different countries is beyond the scope of this paper. Therefore, for the rest of this paper, the discussion remains focused on the U.S..

# 2 Theory

## 2.1 Model Setup

### 2.1.1 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:

$$\mathcal{J}_t = \mathbb{E}_t \left[ \int_t^\infty e^{-\rho(s-t)} u(c_s) ds \right], \forall t \geq 0 \quad (1)$$

where  $\rho$  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, I introduce a normalized aggregator  $f$

for consumption  $c$  and continuation value  $\mathcal{J}$  in each period:

$$f(c, \mathcal{J}) = \frac{\rho}{1 - \theta} \frac{c^{1-\theta} - [(1 - \gamma) \mathcal{J}]^{\frac{1-\theta}{1-\gamma}}}{[(1 - \gamma) \mathcal{J}]^{\frac{1-\theta}{1-\gamma} - 1}} \quad (2)$$

where  $\gamma$  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  $\zeta$ , cash holdings  $\omega$ , 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, I drop the individual and time subscript for simplicity unless otherwise needed.

### 2.1.2 Product market competition

In order to characterize the origins of superstar firms, I model the product market competition as a modified Dixit-Stiglitz demand system (Dixit and Stiglitz, 1977), as this simple framework provides a tractable way to introduce monopolistic competition and increasing returns. More specifically, there is one homogeneous final goods produced by a number of perfectly competitive firms. 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:

$$\mathcal{Y} = \int_0^1 \mathcal{M}(\zeta_i)^\phi y_i(\zeta_i) di = \int_0^1 \zeta_i^\phi y_i(\zeta_i) di \quad (3)$$

where  $\mathcal{M}$  is a strictly increasing function of capital quality  $\zeta$  and  $\phi \geq 0$  represents the consumer's preference on product quality. In the baseline model, I use the simplest functional form  $\mathcal{M}(\zeta) = \zeta$ , but my story can be extended to allow for a more general  $\mathcal{M}$  function. The key difference between this paper and the original Dixit-Stiglitz framework is that I 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, I 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  $\phi$ . 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,  $\phi$  denotes how much these consumers care about the quality. With a relatively higher value  $\phi$ , 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 \(1981\)](#).<sup>8</sup>

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. 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 I 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  $\phi$  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 I 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, I 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

---

<sup>8</sup>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 Roldan-Blanco \(2021\)](#) use this term as a proxy for the total *perceived* quality of different goods.

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  $\pi$  and it can be summarized as below:

$$\pi \equiv \max_y p(\zeta)y - \Theta(y) = p(\zeta)y - f_0 - \xi_0 y^{\frac{1}{\eta}} \quad (4)$$

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 I assume that capital quality is subject to some random shocks with the following process:

$$d\zeta_t = \left( \bar{\mu} + \iota_t^\zeta - \delta\zeta_t \right) dt + \sigma \sqrt{\zeta_t} dZ_t \quad (5)$$

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 (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,  $\iota_t^\zeta$  is the entrepreneur's total (dis)investment on capital quality at time  $t$ , and  $\delta$  is the capital quality depreciation rate.  $Z_t$  is the standard exogenous Brownian shock, and it is independent and identically distributed (i.i.d.) across different firms.  $\sigma$  represents the sensitivity of  $\zeta$  to  $Z_t$ . As my purpose is to investigate how shifts in demand and supply curves amplify the impacts of incomplete market frictions on the value of cash, here I assume that entrepreneurs cannot fully hedge their business risk  $Z_t$ .<sup>9</sup>

Following [Hayashi \(1982\)](#) and references thereafter, I 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  $\kappa_0$  measures the degree of investment inflexibility and

---

<sup>9</sup>An implicit assumption here is that there is no perfect capital market for entrepreneurs to rent or lease. I 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.

$\kappa_0 > 0$ .

### 2.1.3 Corporate risk management

The Modigliani-Miller theorem holds in a standard neoclassical macro model, and corporate capital structure does not matter. However, in reality, entrepreneurs need to make complex financial decisions such as choosing leverage, the timing of equity issuance, and debt maturity. More importantly, these corporate behaviors do have important macroeconomic implications. In this paper, for tractability, I consider a simple financing environment, where entrepreneurs can fund their investment projects with current earnings, cash holdings, and externally by issuing instantaneous risk-free bonds.<sup>10</sup> 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, I 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, I 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:

$$db_t = l_t^b dt \quad (6)$$

$b_t$  denotes the amount of short-term debt issued by each entrepreneur at time  $t$  and  $l_t^b$  is the net changes in debt issuance, i.e.,  $l_t^b = b_{t+dt} - b_t$ . As mentioned before, for simplicity, I 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, I introduce the following two modifications. First, as we can see from the timeline in Section 2.1.4, I 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, I introduce the following borrowing constraint to make sure that entrepreneurs can only borrow a

---

<sup>10</sup>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.

fraction of their earnings:

$$b_t \leq \beta \frac{\pi_t - \chi_0}{1 + r + \chi_1}, \eta \in [0, 1] \quad (7)$$

where  $\pi$  are the entrepreneur's profits as I have defined before. With these two adjustments, all the borrowing and lending in this economy become **credit risk-free**. In some aspects, Equation (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  $\beta = 1$ , I say 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, I assume that all the entrepreneurs will face the same  $\beta$ .<sup>11</sup>

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  $\chi_0$  and  $\chi_1$  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,  $\chi_0$  and  $\chi_1$  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, I 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](#)).

---

<sup>11</sup>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.

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

$$d\omega_t = (\iota_t^\omega - \lambda\omega_t) dt, \text{ and } \omega_t \geq 0 \quad \forall t \quad (8)$$

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

In Equation (8),  $\lambda$  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  $\lambda$  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,  $\lambda$  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 I use the reduced-form approach to introduce the cash carry costs, and the carry costs can be summarized as a positive parameter  $\lambda$ .<sup>12</sup> 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 my story.

Another important feature about Equation (8) is this *time-invariant*  $\lambda$ . 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 I

---

<sup>12</sup>In the baseline model, I 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.

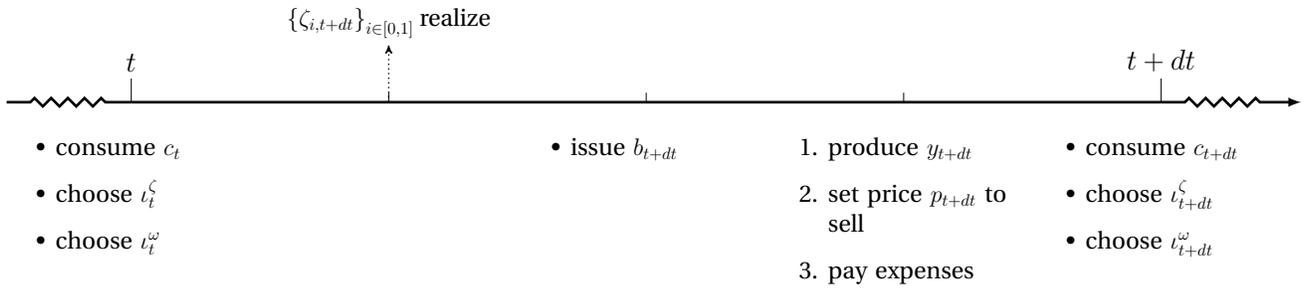
will show later, this difference is crucial when I explore the macroeconomic consequences of internal financing. I put more discussions on this assumption in Section 5.2.2.

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

$$c + l^k + l^\omega - l^b = \pi - rb - \mathbf{1}_{b \neq 0} (\chi_0 + \chi_1 |b|) - \frac{\kappa_0}{2} \frac{(l^k)^2}{k} \quad (9)$$

## 2.1.4 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.

## 2.1.5 Equilibrium definition

The equilibrium definition can be summarized in Definition 1.

**Definition 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  $\left\{ \left( y_{i,t}, l_{i,t}^\zeta, l_{i,t}^\omega, l_{i,t}^b, c_{i,t} \right)_{i \in [0,1]} \right\}_{t=0}^{\infty}$  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  $\left\{ \left( y_{i,t}, l_{i,t}^\zeta, l_{i,t}^\omega, l_{i,t}^b, c_{i,t} \right)_{i \in [0,1]} \right\}_{t=0}^{\infty}$

maximize each entrepreneur's life-time utility (1) subject to constraints (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:*

$$p(\zeta) = \zeta^\phi \quad (10)$$

- *bond market:*

$$\int b_{i,t} di = 0 \quad (11)$$

- *final goods market:*

$$C_t + \mathcal{I}_t + \mathcal{G}_t + \mathcal{X}_t = \mathcal{Y}_t \quad (12)$$

where  $\mathcal{Y}_t$ ,  $C_t$ ,  $\mathcal{I}_t$ ,  $\mathcal{G}_t$ , 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

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

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  $\{p^*, l^*, l^{k,*}, \iota^{\omega,*}, \iota^{b,*}, c^*\}$  and satisfies the following Kolmogorov forward equation<sup>13</sup>:

$$\frac{\partial \Lambda_t(\zeta, \omega, b)}{\partial t} = -\frac{\partial}{\partial \zeta} [\mu^{\zeta,*}(\zeta) \Lambda_t(\zeta, \omega, b)] - \frac{\partial}{\partial \omega} [\mu^{\omega,*}(\omega) \Lambda_t(\zeta, \omega, b)] - \frac{\partial}{\partial b} [\mu^{b,*}(b) \Lambda_t(\zeta, \omega, b)] + \frac{\partial^2}{\partial \zeta^2} [\sigma^{\zeta,*}(\zeta)^2 \Lambda_t(\zeta, \omega, b)] \quad (13)$$

<sup>13</sup>The derivation of Kolmogorov forward equation follows Stokey (2008).

where

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

By definition, stationary equilibrium means that  $\frac{\partial \Lambda_t(\zeta, \omega, b)}{\partial t} = 0$ , i.e., Equation (13) becomes

$$0 = -\frac{\partial}{\partial \zeta} [\mu^{\zeta,*}(\zeta) \Lambda_t(\zeta, \omega, b)] - \frac{\partial}{\partial \omega} [\mu^{\omega,*}(\omega) \Lambda_t(\zeta, \omega, b)] - \frac{\partial}{\partial b} [\mu^{b,*}(b) \Lambda_t(\zeta, \omega, b)] + \frac{1}{2} \frac{\partial^2}{\partial \zeta^2} [\sigma^{\zeta,*}(\zeta)^2 \Lambda_t(\zeta, \omega, b)] \quad (14)$$

## 2.2 Risky Superstar Economy

After describing the basic setup, now I turn to analyzing the model. The main purpose here is to show 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** I begin with discussing how a firm's markup  $\mu$  and earnings  $\pi$  are linked to its underlying capital quality  $\zeta$ . In this paper, the definition of a firm's markup  $\mu$  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  $\pi$  is simply the difference between profits and total costs, i.e.,  $\pi = py - f_0 - \xi_0 y^{\frac{1}{\eta}}$ . I summarize the key results related to the origins of markup in Lemma 1. All the proofs are provided in Appendix A.

**Lemma 1.** *With the optimization problem shown in Equation (4), the markup and earnings for the firm with capital quality  $\zeta$  are shown as follows*

$$\mu = \frac{py}{f_0 + \xi_0 y^{\frac{1}{\eta}}} = \frac{1}{\eta + f_0 \left[ \zeta^{\frac{\phi}{1-\eta}} \left( \frac{\eta}{\xi_0} \right)^{\frac{\eta}{1-\eta}} \right]^{-1}} \quad (15)$$

$$\pi = (1 - \eta) \left( \frac{\eta}{\xi_0} \right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi}{1-\eta}} - f_0 \quad (16)$$

Three important conclusions can be drawn from Lemma 1. First, both markup and earnings are increasing in capital quality  $\zeta$ . This outcome is quite intuitive given the model setup used in this paper. To begin with, I 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, I 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 1 shows, the earnings-markup relationship is eventually pinned down by the preference parameter  $\phi$  and some production technology parameters such as  $\eta$ . 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  $\zeta$ . 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  $\zeta$ . 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, my story here relies crucially on the existence of fixed costs  $f_0$ , but I 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, my 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. However, I will not limit the story in this paper to intangible capital because the recent robotic automation creates efficiencies to the operation scale but requires manufacturing firms to pay more on fixed production costs.

**Quality-based non-homogeneous earnings process** With the previous model result, it can easily show that with the endogenous monopolistic competition, the actual earnings process becomes different from the underlying capital quality process. I summarize the key results in the following lemma.

**Lemma 2.** *Given the homogeneous underlying process shown as in Equation (5), the actual earnings process is a quality-based non-homogeneous one:*

$$d\pi_t = \underbrace{\left[ \pi'(\zeta_t) \left( \bar{\mu} + \iota_t^\zeta - \delta\zeta_t \right) + \frac{\sigma^2 \zeta_t}{2} \pi''(\zeta_t) \right]}_{\text{drift}} dt + \underbrace{\pi'(\zeta_t) \sigma \sqrt{\zeta_t}}_{\text{volatility}} dZ_t \quad (17)$$

where

$$\pi'(\zeta) = \phi \left( \frac{\eta}{\xi_0} \right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi}{1-\eta}-1} \quad (18)$$

$$\pi''(\zeta) = \phi \left( \frac{\phi}{1-\eta} - 1 \right) \left( \frac{\eta}{\xi_0} \right)^{\frac{\eta}{1-\eta}} \zeta^{\frac{\phi}{1-\eta}-2} \quad (19)$$

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

More importantly, my 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  $\zeta$  follows any stochastic differential equation  $d\zeta = f(\zeta) dt + g(\zeta) d\mathcal{W}$ , and earnings  $\pi$  are a function of  $\zeta$ , i.e.,  $\pi = \pi(\zeta)$ . By Ito's Lemma, I can derive the entrepreneur's earning as the following stochastic process:

$$d\pi = \pi'(\zeta) d\zeta + \frac{1}{2}\pi''(\zeta) g(\zeta)^2 dt = \left[ \pi'(\zeta) f(\zeta) + \frac{1}{2}\pi''(\zeta) g(\zeta)^2 \right] dt + \pi'(\zeta) g(\zeta) d\mathcal{W} \quad (20)$$

Based on the equation above, as long as  $\pi$  is not a linear function of  $\zeta$ , i.e.,  $\pi'$  is not independent of  $\zeta$ , 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.

**Impacts of shifts in demand and supply curves** Now I turn to investigating 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. My key conclusions here can be summarized in the following proposition.

**Proposition 1.** *Given that there is a fundamental change that increases both the taste for quality  $\phi$  and the operation scale  $\eta$  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:*

- i. earnings  $\pi(\zeta)$  are convex in capital quality  $\zeta$ ; entrepreneur's output  $y(\zeta)$  is also a convex function of capital quality  $\zeta$  if  $\phi > \frac{1-\eta}{\eta}$*
- ii. in steady state, drift and volatility components of the entrepreneur's earnings process  $d\pi$  are both increasing in capital quality  $\zeta$*

The formal proof is provided in Appendix A and here I focus on the main intuition. As can be seen from Proposition 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.

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  $\pi$  with respect to  $\zeta$ , two implications arise directly from the earnings process (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 risky. Therefore, changes in economic fundamentals will 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.

### 2.3 Endogenous Firm-Market Boundary

In this part, I will show how 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.

**Proposition 2.** *Given the model setup, the optimal policy of an entrepreneur with capital quality  $\zeta$  involves a quality-based downward control boundary  $\bar{\Omega}^\zeta$  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.,  $\omega = \bar{\Omega}^\zeta$  and  $b < 0$ . The Hamilton-Jacobi-Bellman (HJB) equation for this external lending region is*

$$0 = \max_{\iota^\zeta, c} \left\{ \begin{array}{l} f(c, \mathcal{J}) + (\bar{\mu} + \iota^\zeta - \delta\zeta) \mathcal{J}_\zeta + \frac{\zeta\sigma^2}{2} \mathcal{J}_{\zeta\zeta} - \zeta\sigma^2\pi'(\zeta) \mathcal{J}_{\zeta b} + \frac{\zeta\sigma^2}{2} (\pi'(\zeta))^2 \mathcal{J}_{bb} \\ - \left[ \pi'(\zeta) (\bar{\mu} + \iota^\zeta - \delta\zeta) + \frac{1}{2}\pi''(\zeta) \zeta\sigma^2 - c - \iota^\zeta - rb - \chi_0 + \chi_1 b - \lambda\bar{\Omega}^\zeta - \frac{\kappa_0(\iota^\zeta)^2}{2\zeta} \right] \mathcal{J}_b \end{array} \right\} \quad (21)$$

where  $c$  and  $\iota^\zeta$  satisfy the following first-order conditions

$$f_c(c, \mathcal{J}) = -\mathcal{J}_b \quad (22)$$

$$-\frac{\mathcal{J}_\zeta}{\mathcal{J}_b} + \pi'(\zeta) = 1 + \kappa_0 \frac{\iota^\zeta}{\zeta} \quad (23)$$

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

$$0 = \max_{\iota^\zeta, c} \left\{ \begin{array}{l} f(c, \mathcal{J}) + (\bar{\mu} + \iota^\zeta - \delta\zeta) \mathcal{J}_\zeta + \frac{\zeta\sigma^2}{2} \mathcal{J}_{\zeta\zeta} - \zeta\sigma^2 \pi'(\zeta) \mathcal{J}_{\zeta b} + \frac{\zeta\sigma^2}{2} (\pi'(\zeta))^2 \mathcal{J}_{bb} \\ - \left[ \pi'(\zeta) (\bar{\mu} + \iota^\zeta - \delta\zeta) + \frac{1}{2} \pi''(\zeta) \zeta\sigma^2 - c - \iota^\zeta - rb - \chi_0 - \chi_1 b - \lambda \underline{\Omega}^\zeta - \frac{\kappa_0 (\iota^\zeta)^2}{2\zeta} \right] \mathcal{J}_b \end{array} \right\} \quad (24)$$

where  $c$  and  $\iota^\zeta$  satisfies the following first-order conditions:

$$f_c(c, \mathcal{J}) = \max \left\{ -\mathcal{J}_b, -\mathcal{J}_b|_{b=\beta \frac{\pi-\chi_0}{1+r+\chi_1}} \right\} \quad (25)$$

$$1 + \kappa_0 \frac{\iota^\zeta}{\zeta} = \frac{\mathcal{J}_\zeta}{\max \left\{ -\mathcal{J}_b, -\mathcal{J}_b|_{b=\beta \frac{\pi-\chi_0}{1+r+\chi_1}} \right\}} + \pi'(\zeta) \quad (26)$$

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

$$0 = \max_{\iota^\zeta, c} \left\{ \begin{array}{l} f(c, \mathcal{J}) + (\bar{\mu} + \iota^\zeta - \delta\zeta) \mathcal{J}_\zeta + \frac{\zeta\sigma^2}{2} \mathcal{J}_{\zeta\zeta} + \zeta\sigma^2 \pi'(\zeta) \mathcal{J}_{\zeta\omega} + \frac{\zeta\sigma^2}{2} (\pi'(\zeta))^2 \mathcal{J}_{\omega\omega} \\ + \left[ \pi'(\zeta) (\bar{\mu} + \iota^\zeta - \delta\zeta) + \frac{1}{2} \pi''(\zeta) \zeta\sigma^2 - c - \iota^\zeta - \lambda\omega - \frac{\kappa_0 (\iota^\zeta)^2}{2\zeta} \right] \mathcal{J}_\omega \end{array} \right\} \quad (27)$$

where  $c$  and  $\iota^\zeta$  satisfy the following first-order conditions

$$f_c(c, \mathcal{J}) = \mathcal{J}_\omega \quad (28)$$

$$\mathcal{J}_\zeta = \left[ 1 + \kappa_0 \frac{\iota^\zeta}{\zeta} - \pi'(\zeta) \right] \mathcal{J}_\omega \quad (29)$$

In addition, the accompanying boundary conditions consist of the Neumann boundary conditions in the  $\zeta$ -dimension

$$\mathcal{J}_\zeta(\zeta_{min}, \omega, 0) = 0, \forall \omega \quad (30)$$

$$\mathcal{J}_\zeta(\zeta_{max}, \omega, 0) = 0, \forall \omega \quad (31)$$

the Neumann boundary conditions in the  $\omega$ -dimension

$$\mathcal{J}_\omega(\zeta, \underline{\Omega}^\zeta, 0) = 1 + r + \chi_1, \forall \zeta \quad (32)$$

$$\mathcal{J}_\omega(\zeta, \overline{\Omega}^\zeta, 0) = 1 + r - \chi_1, \forall \zeta \quad (33)$$

$$\mathcal{J}_{\omega\omega}(\zeta, \overline{\Omega}^\zeta, 0) = 0, \forall \zeta \quad (34)$$

and the Neumann boundary conditions in the  $b$ -dimension

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

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

**Boundary of the invisible hand** Results in Proposition 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 very standard cash inventory model, the optimal policy is made from 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, I introduce the following definition on the boundary of the invisible hand.

**Definition 2.** The *boundary of the invisible hand* is made of a set of upward and downward control boundaries  $\{\overline{\Omega}^i, \underline{\Omega}^i\}_{i \in [0,1]}$ . For each entrepreneur  $i \in [0, 1]$ ,  $\overline{\Omega}^i$  and  $\underline{\Omega}^i$  are the solutions to the Neumann boundary conditions (32)-(34) of a nonlinear partial differential equation (27). At time  $t$ , the area con-

trolled by the price mechanism, i.e., the fraction of entrepreneurs outside the internal financing region, can be calculated as follows:

$$\Psi_t \equiv \int \left(1 - \mathbb{I}_{\Omega^i < \omega^i < \underline{\Omega}^i}\right) di = \iiint \left(1 - \mathbb{I}_{\Omega^\zeta < \omega < \underline{\Omega}^\zeta}\right) \Lambda_t(\zeta, \omega, b) d\zeta d\omega db \quad (35)$$

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

The idea behind Definition 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, I introduce the transaction cost in a reduced-form but add it to a heterogeneous agent general equilibrium framework so that I 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, I formally introduce the boundary between firms and the financial market.

With Definition 2, I can use Equation (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,  $\Psi$  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  $\Psi$  through their impacts on the earnings process. More importantly, we should expect  $\Psi$  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, I will perform a quantitative exercise to estimate the magnitudes of these impacts.

**HJB equation** In this part, I will 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 extent. Here I 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 (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 will increase 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 (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  $\iota^\zeta$  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 I turn to explaining the first-order conditions derived from the entrepreneur's optimal decision. Again, I will use the internal financing region as an example. The optimal consumption choice is shown in Equation (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 (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  $1 + \kappa_0 \frac{\zeta}{\zeta}$  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  $1 + \kappa_0 \frac{\zeta}{\zeta}$  and marginal cash value  $\mathcal{J}_\omega$ . Meanwhile, different from the existing literature, in Equation (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  $\left[1 + \kappa_0 \frac{\zeta}{\zeta} - \pi'(\zeta)\right] \mathcal{J}_\omega$ .

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  $1 + \kappa_0 \frac{\zeta}{\zeta}$ . In contrast, Equation (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.

### 3 Reduced-form Evidence

Here I 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.

### 3.1 Risky Superstars

The first testable implication from my model is that superstar firms are riskier than the other firms. Here I provide some empirical evidence to support this claim. In Figure 8, I plot the time series of markup volatility for five groups of firms with different levels of markup. Consistent with the model, I define superstar firms to be those with the highest markup. I obtain the results in Figure 8 using the following steps. First, for each individual company in each period, I compute its markup volatility by using a five-year rolling window. Second, in each year, I classify all the firms with available markup measures into five different quantiles according to their markup level. Finally, I calculate the annual average markup volatility for each quantile of firms. As shown in Figure 8, 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.

Two caveats are worth noting for interpreting the results here. First, my 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, I 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, I 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, I argue that superstar firms 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.

## 3.2 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 9, I again classify all the firms into five different quintiles as what I did in Figure 8. This time, however, I investigate the dispersion of  $mpk$  within each group of firms. As shown in Figure 9, 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 9, 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 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 A18 in the appendix, when I classify firms into different groups according to their cash holdings, misallocation patterns are similar among them. Another supporting evidence is that when plotting the level of  $mpk$  for different quintiles of firms, Figure A19

shows that average  $mpk$  of superstar firms is the lowest while that of firms with smallest markup is the highest. These two results further confirm the previous hypothesis that the origin of self-financing matters for its impact on misallocation.

### 3.3 Markup and Cash Holding

The third testable implication is that there should be a positive relationship between firm-level markup and cash holdings. In the previous model, I 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 I provide some related reduced-form evidence for this prediction. Following the existing literature, I 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, I 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, I will only discuss the empirical results related to markup-cash relationship.  $q$ -cash results are provided in the appendix.

#### 3.3.1 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 10 presents the main result. I provide two graphs: one is plotted with the raw data and the other with a preferred model specification used in the later regression analysis. My 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 10, 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 10 suggests that this financial frictions story can help make sense of the cash-markup relationship for the left-tail firms, while my story can better explain cash holding behaviors for firms in the right-tail distribution.

The Binscatter plots for cash-Tobin- $q$  and cash-Total- $q$  relationships are provided in Figure A16 and Figure A17, respectively. Based on these two graphs, on average, firms with relatively higher  $q$  hold more cash on their balance sheets.

### 3.3.2 Regression analysis

Table 1 shows the regression results for investigating the relationship between markup and cash holdings. I adopt two-way fixed-effects regressions to estimate the impacts of markup on corporate cash holding behaviors. The general model specification in Table 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  $\mu$  represents the empirical measure for firm-level markup. I am primarily interested in the sign and statistical significance of the estimated coefficient  $\beta$ . 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, I also add the square markup term  $\mu^2$  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, I include some other firm-level indicators such as return of assets, tangibility, investment, size, profitability, R&D, book leverage, and dividend payout. In addition, I 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 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, I have included all the firm-level controls. The difference between the last three columns comes from

the choice of fixed effects. In Column (10), I control for firm and year fixed effects. In Column (11), I include 3-digit NAICS industry and year-fixed effects. In the last column, I 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 1, I 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 my model. In addition, I 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, the standard error is small, and the estimation is accurate. As is shown in the bottom panel in Figure 10, 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, I provide the corresponding regression results for the  $q$ -cash relationship. Similarly, Table A1 shows that across different model specifications, I 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, the three pieces of reduced-form evidence provides additional support for the underlying mechanism in the model section. Now I turn to quantitative analysis.

## 4 Quantitative Analysis

This section outlines how the model is parameterized and then investigates the model's quantitative implications. To start, I discuss the external calibration and structural estimation strategy used in this paper. Next, I show the extent to which my model is able to match both the aggregate trends and cross-sectional patterns in the data; furthermore, I discuss the model's aggregate implications and their connections with the existing literature. Finally, I present some quantitative results from several counterfactual exercises.

## 4.1 Parametrization

The model is calibrated at an annual frequency. To reduce the computational burden, I externally calibrate a subset of parameters, then consider the rest estimated within the model. For estimating those structural parameters, I 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, I 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.

### 4.1.1 External calibration

Those externally calibrated parameters are shown in the top panel in [Table 2](#). The rate of time preference  $\rho$  is set to be 0.046. The risk aversion  $\gamma$  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 I will use the data to calibrate them. The capital depreciation rate  $\delta$  can be computed with the fixed asset tables (FAT) obtained from the U.S. Bureau of Economic Analysis (BEA). Consistent with the previous model, I do not limit my analysis here to intangible capital only. Instead, I will calculate the average depreciation rate for all assets. Of course, this paper focuses on firms, so I 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, I compute the depreciation rate for intangible and tangible assets, respectively. Then I calculate the annual weighted average depreciation rate by using the relative importance of each asset as the weight. Lastly, I compute the average annual depreciation rate for these two subsamples. The 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.

I 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, I rescale the average output to be one here. In order to estimate operating scale  $\eta$  and fixed production cost  $f_0$ , I 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:

$$f_0 = \left(1 - \frac{1}{\mu}\right) py - \pi = \left(1 - \frac{1}{\text{markup}}\right) \text{SALE} - \text{IB} \quad (36)$$

$$\eta = \frac{\xi_0 y^{\frac{1}{\eta}}}{py} = \frac{\text{COGS} - f_0}{\text{SALE}} \quad (37)$$

Following the standard literature, I measure firm-level revenue by using total sales (Compustat data variable *SALE*). In addition, I obtain the total production cost data by using Compustat variable *COGS*, which contains all the direct costs involved with production. The operating profits  $\pi$  are measured with income before extraordinary items (Compustat data variable *IB*). As the firm-level price information is not available in Compustat, I cannot directly estimate variable production cost  $\xi_0$ . Therefore, I 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 my 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  $\sigma$  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.

#### 4.1.2 Internal estimation: SMM-MCMC approach

The rest of the parameters are jointly calibrated with the SMM approach by targeting some moments in the data. I 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  $\phi$ , marginal production cost  $\xi_0$ , investment adjustment cost  $\kappa_0$ , fixed external financing cost  $\chi_0$ , variable external financing cost  $\chi_1$ , and tightness of borrowing constraint  $\beta$ . The data moments I choose are mean and dispersion of markup, investment-to-output ratio, and cash-to-output ratio. In addition, I include the relative markup ratio of the 90th percentile to the median value in these two subsamples. It turns out that quality taste parameter  $\phi$  is sensitive to this data moment, so including it can help identify it more accurately.

In the top panel in Table 3, I 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 presents the median values of these structurally estimated parameters. I choose median instead of mean so that the result suffers less from some extreme outcomes. In addition, I can take full advantage of this MCMC approach to build the corresponding confidence intervals and posterior distributions for each of these estimates. In Figure 11, I 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 3. For instance, in the traditional economy, the median estimated value of  $\phi$  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,  $\phi$  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  $\xi_0$  is from 0.466–1.433 and 0.183–0.338 for two subsamples,

respectively.

## 4.2 Cross-Sectional Validation

In this section, I provide some cross-sectional evidence to validate the model and the choice of parameters. I 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, I do not directly target these moments. Therefore, the goodness of fit on these moments can be informative for evaluating the model's validity.

I first look at how the model fits the cross-sectional cash-to-output ratios. The results are presented in panel A in Figure 12, where I 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, I 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, I compute the firm-level cash-to-output ratios, then take averages for each group in each year. Finally, I 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 I 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 12 is obtained by computing the difference in these numbers between the traditional and the new economy.

Figure 12 shows that my model is able to quantitatively match the cross-sectional patterns of cash-to-output ratios in both subsamples. My model implies that all the firms in different groups tend to save more in the Superstar Economy, especially for the firms with the highest markups. This cross-sectional pattern is consistent with what I 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 I 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 my 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.

I 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 12, which is generated in the same way as I 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, my 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.

### 4.3 Policy Functions

Now I turn to discussing the entrepreneur's optimal investment and cash holding decisions. Figure 13 presents the numerical solutions with estimated parameters from the Superstar Economy subsample. In panel A, I 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 my 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 I 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 13 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.

## 4.4 Aggregate implications with M.I.T. Shocks

As shown in Section 1, over the past several decades, the U.S. economy has been characterized by several “puzzling” macro-finance trends. Here I will use the general equilibrium model developed in Section 2 and the previous parameterization strategy to quantify its underlying mechanism and investigate whether the model is able to jointly explain these aggregate trends. Here I inspect the mechanism 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.

### 4.4.1 Self-financing and “misallocation”

Table 4 summarizes the results of comparing these macro-finance trends in the data and those obtained from the model. I start with misallocation. In the data, misallocation is measured the same way I did for Figure 3, 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 4, my 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.

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. My 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.

My explanation here is similar to the risk and inaction story in [Bloom \(2009\)](#) and [Akerberg, Caves and Frazer \(2015\)](#). A standard result from this type of model is the existence of an inaction region. However, [Bloom \(2009\)](#) and [Akerberg, 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 I 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 my 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.

#### **4.4.2 Eclipse of the public markets**

The second row in [Table 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

I did for Figure 4 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  $\zeta$ . According to Table 4, my 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.<sup>14</sup> 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 through public capital markets. My 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 effectiveness 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](#)

---

<sup>14</sup>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.

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 I argue that it could come from some economic fundamental changes that permanently change the risk of corporate earnings.

#### 4.4.3 Two rates of return

The third row in Table 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 5. In the model, MPK is obtained by calculating the average value of  $\mathcal{J}_\zeta$ , while  $r$  is simply the equilibrium interest rate. Table 4 shows that in the data, the measured MPK &  $r$  gap has increased by 5.00%, from 5.27% to 10.27%. Meanwhile, my 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 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 my paper is that there are actually two types of  $R$ s: 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  $R$ s to be equal. However, if there are some transaction costs of using the external market, then the gap between these two  $R$ s 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, my 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.

#### **4.4.4 Moving firm-market boundary**

Another interesting result from my 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. I present the result in the last part of [Table 4](#), which is obtained by calculating the difference in the steady-state values of  $\Psi$  between these two subsamples.  $\Psi$  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 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  $\Psi$  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.

## 4.5 Counterfactual Exercises

In order to show the relative contributions of different parameters, I conduct some counterfactual exercises. The corresponding quantitative results are provided in [Table 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  $\xi_0$ . When I fix  $\xi_0$  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  $\phi$  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 significantly to the rise of superstar firms and their cash holding behaviors. Therefore, I 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 5](#), changes in the degree of financial frictions are not an important factor behind these macro-finance trends. When I fix the value of  $\beta$  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.

## 5 Related Literature and Additional Discussions

### 5.1 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.<sup>15</sup> 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

---

<sup>15</sup>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.

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, my 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 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, I 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 [Imrohorglu \(1989\)](#) 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](#)). 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 I 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.

## **5.2 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, I 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, I 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, I 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, I will provide some additional discussions on their validity.

### 5.2.1 Incomplete market

As mentioned before, to obtain firm-market boundaries, one crucial assumption is that the market is not (dynamically) complete. Here I 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 I 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 A21 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.

### 5.2.2 Safe assets

At the same time, in order to obtain [Coase \(1937\)](#)'s second conjecture that a firm is a substitution of the market, I need an additional assumption: firms accumulate their internal savings through some safe assets. In order to validate this assumption, first, I want to explain what is the fundamental difference between cash and any other type of financial asset. In the model, I 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 I 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, I 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 A20](#) 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.<sup>16</sup> In addition, many recent studies have 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, my 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 \(2018\)](#). 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.

## 6 Conclusion

This paper investigates how the rise of superstar firms affects the effectiveness of the price mechanism. Empirically I find that the capital allocation efficiency in the U.S. has been deteriorating over the last several decades. First, the dispersion of the marginal product of capital has increased sharply among U.S. public firms. Second, the correlation between firm-level total factor productivity and external financing dependence has changed from positive to negative, indicating that less and less finance flows to productive firms. Third, the positive gap between the marginal product of capital and the real interest rate increases over time. I argue that the rise of (superstar) firms and their risk management policies are the main drivers behind these aggregate trends. To formally establish this idea, I 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

---

<sup>16</sup>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.

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, which prompts superstar firms to rely more on internal financing. The expansion of internal financing region weakens the role of price mechanism in allocating resources, thus leading to the increase in misallocation. When taken to the data, the model can quantitatively match some important macro-finance trends, and it shows that the area disciplined by the market system has declined by about 10.88% during the past forty years.

In terms of the policy implications, this paper points out two types of market failures in the Superstar Economy. First, the boundary of the price mechanism is shrinking. The massive corporate cash savings prevent firms from being disciplined by the market system. 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 the 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.

## 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.
- 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.
- Bates, Thomas W., Kathleen M. Kahle, and Rene M. Stulz.** 2009. “Why Do U.S. Firms Hold So Much More Cash than They Used To?” *Journal of Finance*, 64(5): 1985–2021.
- 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.
- 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.
- 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.
- Brander, James A., and Tracy R. Lewis.** 1986. “Oligopoly and Financial Structure: The Limited Liability Effect.” *American Economic Review*, 76(5): 956–970.
- Brunnermeier, Markus K., and Yuliy Sannikov.** 2014. “A Macroeconomic Model with a Financial Sector.” *American Economic Review*, 104(2): 379–421.
- 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.
- 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.

- Caselli, Francesco, and James Feyrer.** 2007. "The Marginal Product of Capital." *Quarterly Journal of Economics*, 122(2): 535–568.
- Cavenaile, Laurent, and Pau Roldan-Blanco.** 2021. "Advertising, Innovation, and Economic Growth." *American Economic Journal: Macroeconomics*, 13(3): 251–303.
- Coase, Ronald.** 1937. "The Nature of the Firm." *Economica*, 4(16): 386–405.
- 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 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."
- 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 Ridder, Maarten.** 2019. "Market Power and Innovation in the Intangible Economy."
- Di Tella, Sebastian.** 2017. "Uncertainty Shocks and Balance Sheet Recessions." *Journal of Political Economy*, 125(6): 2038–2081.
- Dixit, Avinash K., and Joseph E. Stiglitz.** 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review*, 67(3): 297–308.
- 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.
- 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.
- Duffie, Darrell, and Larry G. Epstein.** 1992. "Stochastic Differential Utility." *Econometrica*, 60(2): 353–394.
- Eckbo, B. Espen, and Markus Lithell.** 2021. "Merger-Driven Listing Dynamics."
- 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.
- 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.”
- Gabaix, Xavier, and Augustin Landier.** 2008. “Why has CEO Pay Increased So Much?” *Quarterly Journal of Economics*, 123(1): 49–100.
- 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.
- Graham, John R.** 2000. “How Big Are the Tax Benefits of Debt?” *Journal of Finance*, 55(5): 1901–1941.
- 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.
- 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.
- 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.
- 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.
- Hoberg, Gerard, and Gordon M. Phillips.** 2021. "Scope, Scale and Competition: The 21st Century Firm."
- 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."
- 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." *Harvard Business Review*.
- Jiang, Zhengyang, Hanno Lustig, Stijn Van Nieuwerburgh, and Mindy Z. Xiaolan.** 2021. "Manufacturing Risk-free Government Debt." NBER Working Paper w27786.
- 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. "Monetary Policy According to HANK." *American Economic Review*, 108(3): 697–743.

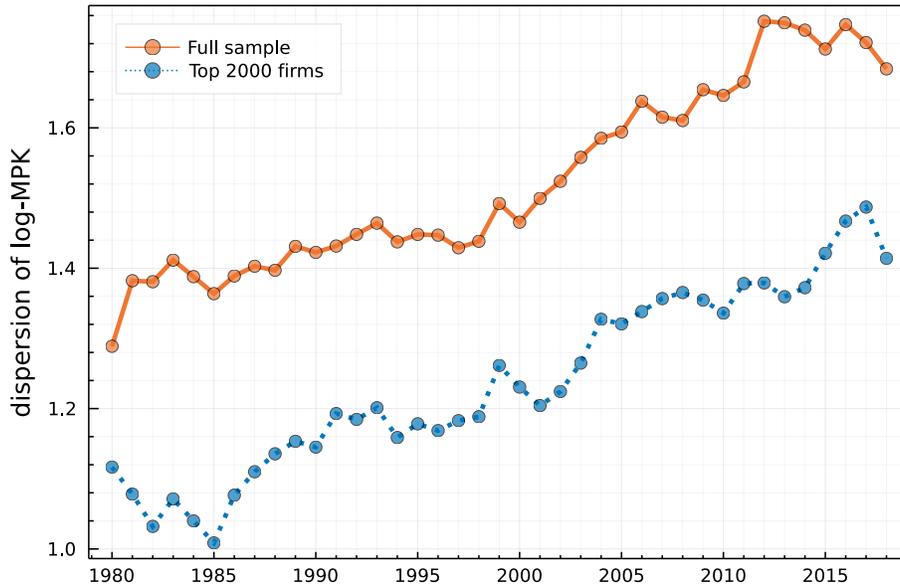
- Karabarbounis, Loukas, and Brent Neiman.** 2018. "Accounting for Factorless Income." NBER Working Paper No. 24404.
- Kehrig, Matthias, and Nicolas Vincent.** 2020. "The Micro-Level Anatomy of the Labor Share Decline." NBER Working Paper No. 25275.
- Kilic, Mete, Louis Yang, and Miao Ben Zhang.** 2019. "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.
- 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."
- 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.
- 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.
- 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.** 2014. "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?" *American Economic Review*, 104(10): 3186–3221.
- 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.
- 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.
- Peters, Ryan, and Lucian Taylor.** 2017. "Intangible capital and the investment- $q$  relation." *Journal of Financial Economics*, 123(2): 251–272.
- Piketty, Thomas.** 2013. *Capital in the Twenty-First Century*. Harvard University Press.
- 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.
- 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.
- Stokey, Nancy.** 2008. *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton University Press.
- 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.

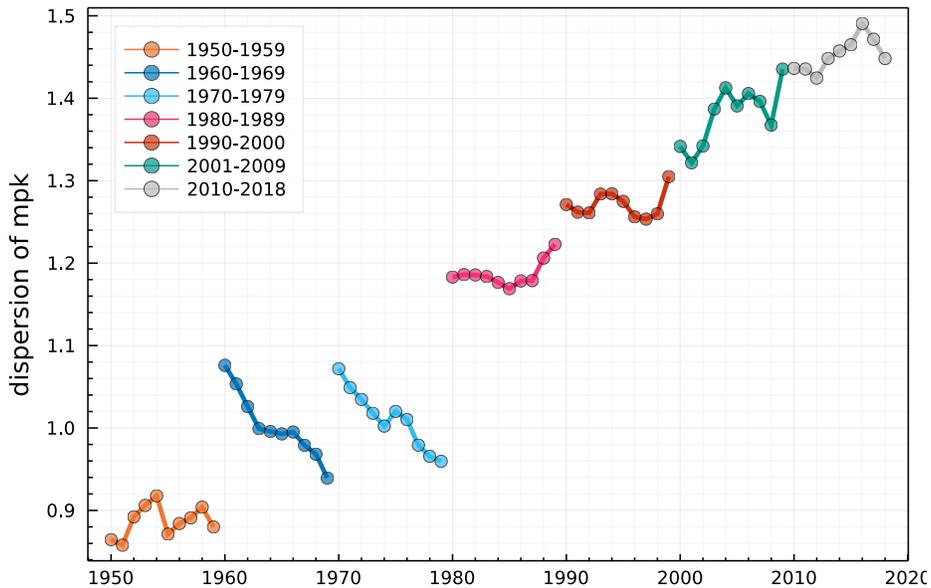
- Tervio, Marko.** 2008. "The Difference That CEOs Make: An Assignment Model Approach." *American Economic Review*, 98(3): 642–668.
- Wang, Chong, Neng Wang, and Jinqiang Yang.** 2012. "A Unified Model of Entrepreneurship Dynamics." *Journal of Financial Economics*, 106: 1–23.
- Williamson, Oliver.** 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.

**Figure 3: Increasing “misallocation”**

**A. unbalanced panels**



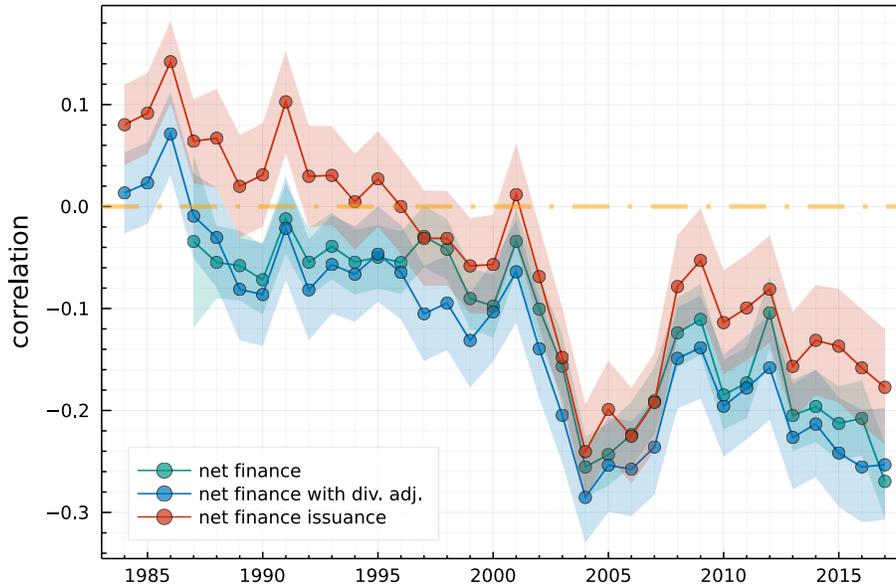
**B. seven small balanced panels**



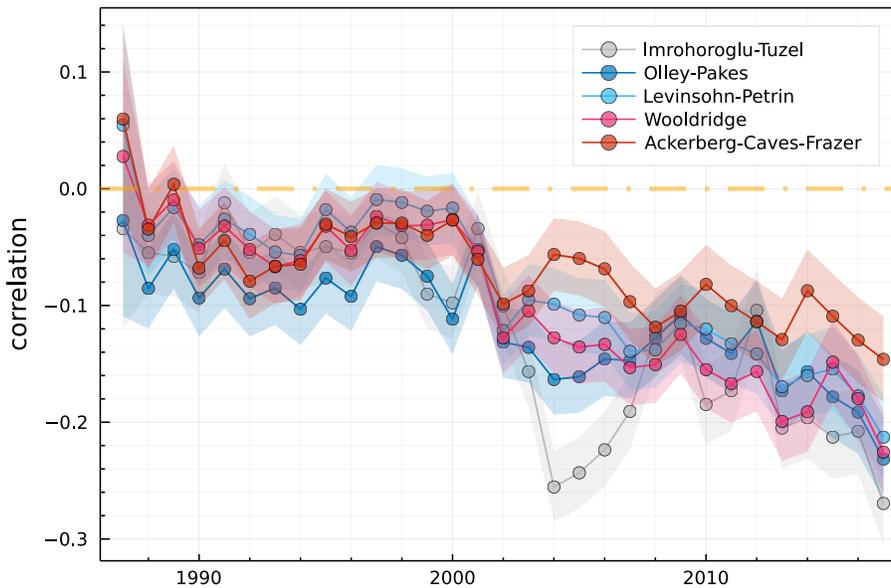
*Notes:* This figure presents different measures on the time-series of capital misallocation degree among U.S. public firms. Misallocation is defined as the dispersion of firm-level log marginal product of capital  $mpk$ . The orange solid line in the top graph represents the full-sample result, where I 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, I present measured misallocation degrees for seven small and balanced panels for each of the seven decades between 1950 - 2020. For each balanced panel, I 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*).

**Figure 4: Eclipse of the public markets**

**A. different measures of net finance**

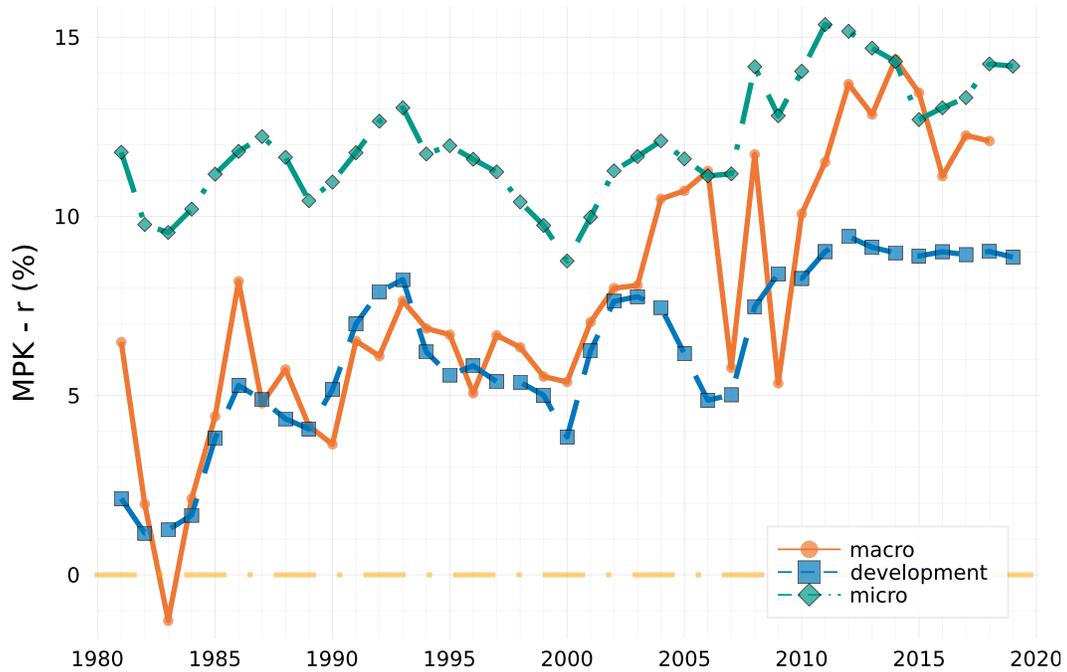


**B. different measures of firm-level TFP**



*Notes:* This figure presents the cross-sectional correlation between firm-level TFP and firm-level net finance dependence. I use five different approaches to measure firm-level TFP, including Imrohorglu-Tuzel (Imrohorglu and Tuzel, 2014), Olley-Pakes (Olley and Pakes, 1996), Levinsohn-Petrin (Levinsohn and Petrin, 2003), Wooldridge (Wooldridge, 2009), and Akerberg-Caves-Frazer (Akerberg, Caves and Frazer, 2015) methods. In addition, I 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.

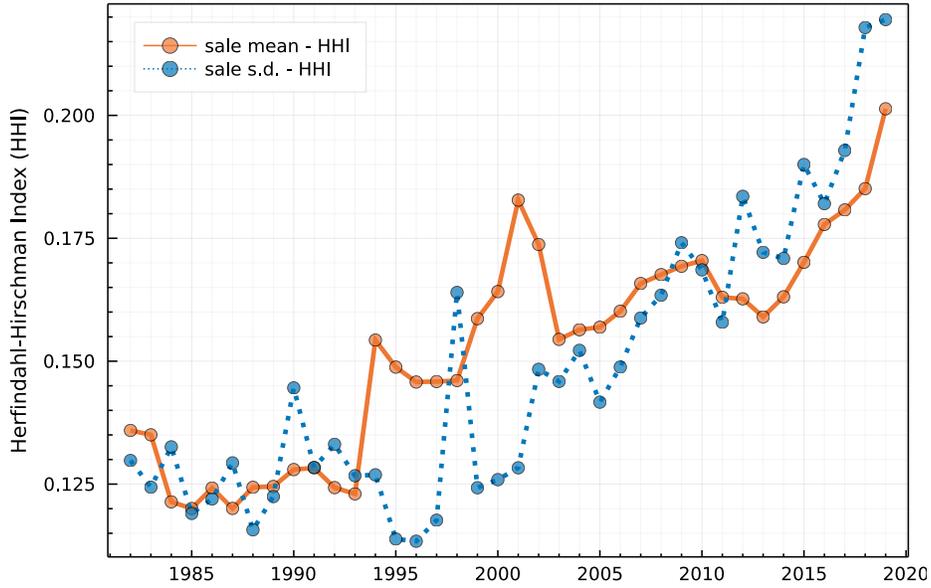
**Figure 5: MPK minus  $r$**



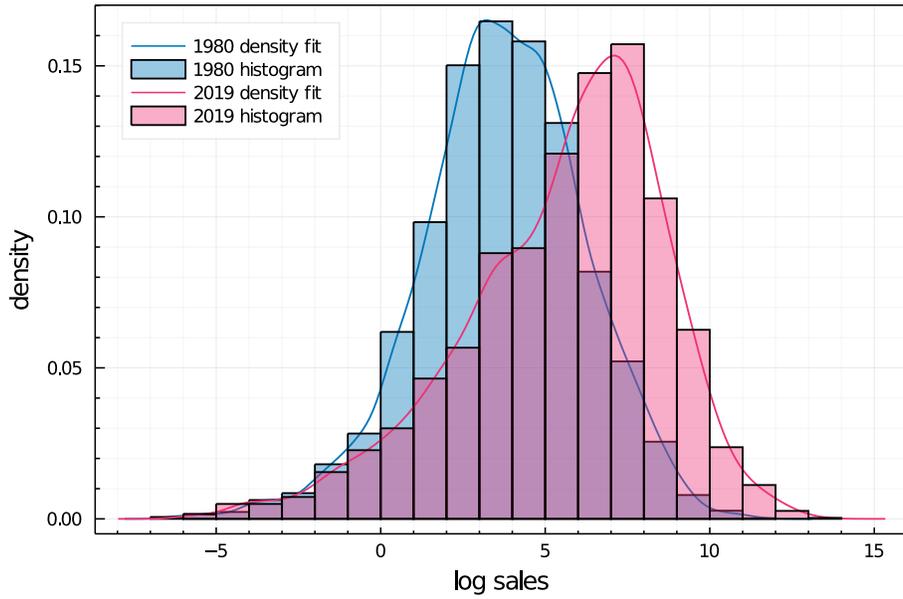
*Notes:* This figure presents the time-series difference between aggregate marginal product of capital (MPK) and real interest rate in the U.S.. MPK is measured with three different methods. The macro approach measures the aggregate MPK by following [Gomme, Ravikumar and Rupert \(2011\)](#) and using 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.

**Figure 6: The risky Superstar Economy**

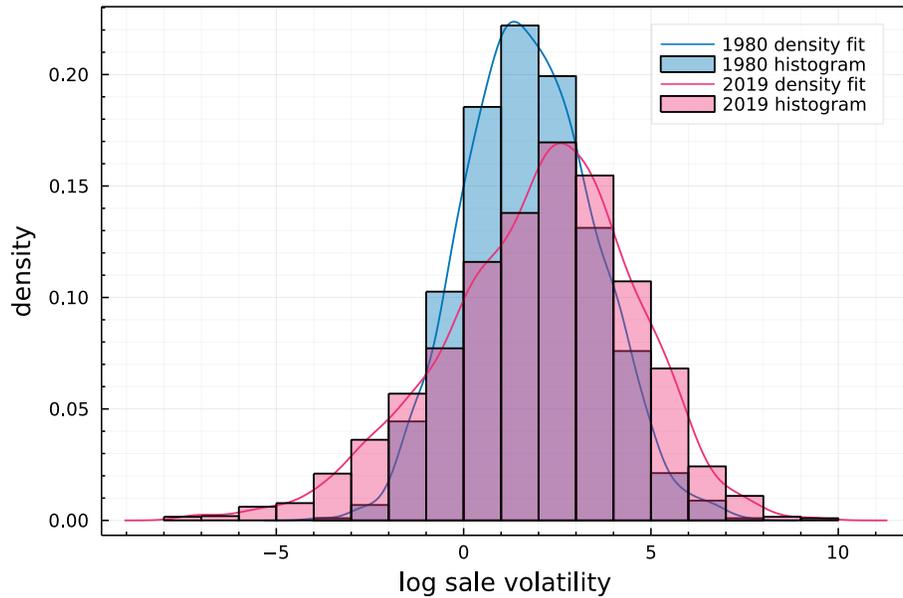
**A. changes in HHI of sale and sale volatility**



**B. shifts in sales distributions**



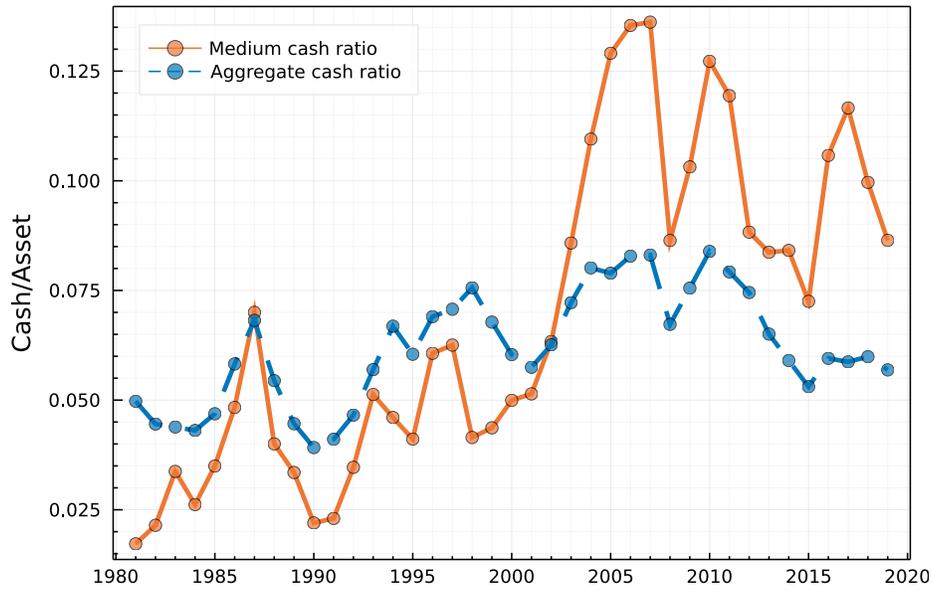
### C. shifts in sales volatility distributions



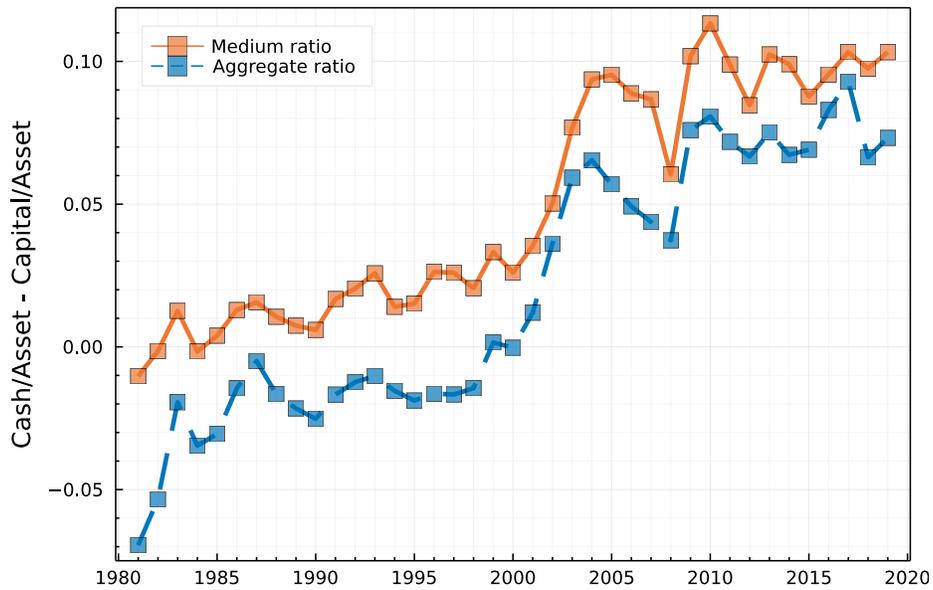
*Notes:* In Graph (A), I follow [Grullon, Larkin and Michaely \(2019\)](#)'s methodology and present the time series plot of two Herfindahl-Hirschman indexes (HHI): the orange line is for the level of sales while the blue one is for the volatility of sales. These two indexes are constructed in the following steps. First, for each firm in each year, I compute the average and standard deviation of its quarterly sales. Second, I calculate the ratios of the firm's sales and volatility to total industry sales and volatility for every firm in a given industry-year. Following the related literature, an industry is defined according to NAICS three-digit industry classification. Third, I square these obtained ratios. Forth, I sum up the squared ratios across all firms in a given industry-year. In Graph (B), I plot the kernel density of the sales distribution in 1980 and in 2019. Graph (C) presents the kernel density of the sales volatility distribution in 1980 and in 2019. Main data source for this figure is the Compustat North American Quarterly data file.

**Figure 7: Corporate cash holding**

**A. changes in cash/asset**

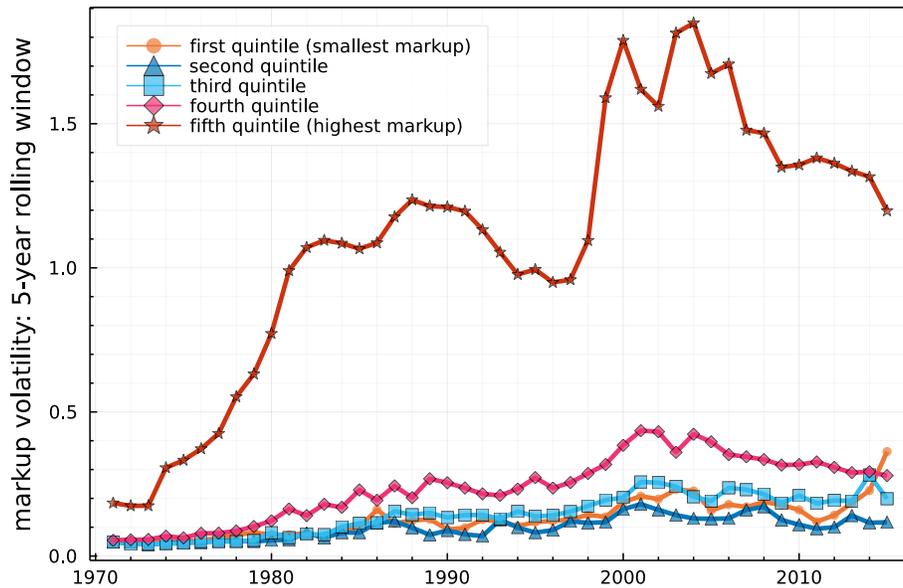


**B. changes in the difference between cash/asset and capital/asset**



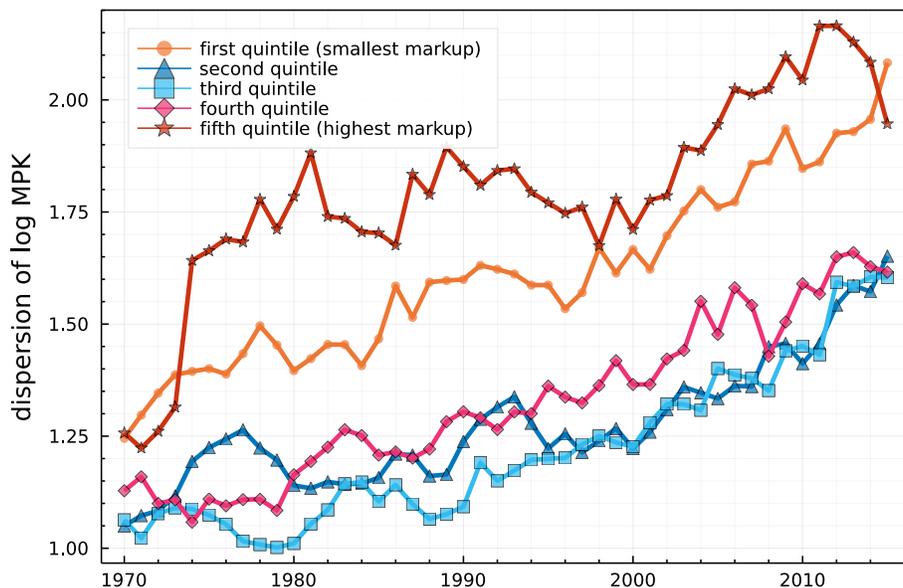
*Notes:* The top graph of this figure presents the time-varying cash-to-asset ratios among the U.S. public firms since the 1980s. In the bottom graph, I present the time-series plot of the difference between cash-to-asset ratio and capital-to-asset ratio. Main data source for this figure is the Compustat 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 capital-to-asset ratio is measured as the ratio of net property, plant and equipment (Compustat series *PPENT*) to firm's lagged total assets (Compustat series *AT*).

**Figure 8: Risky superstars**



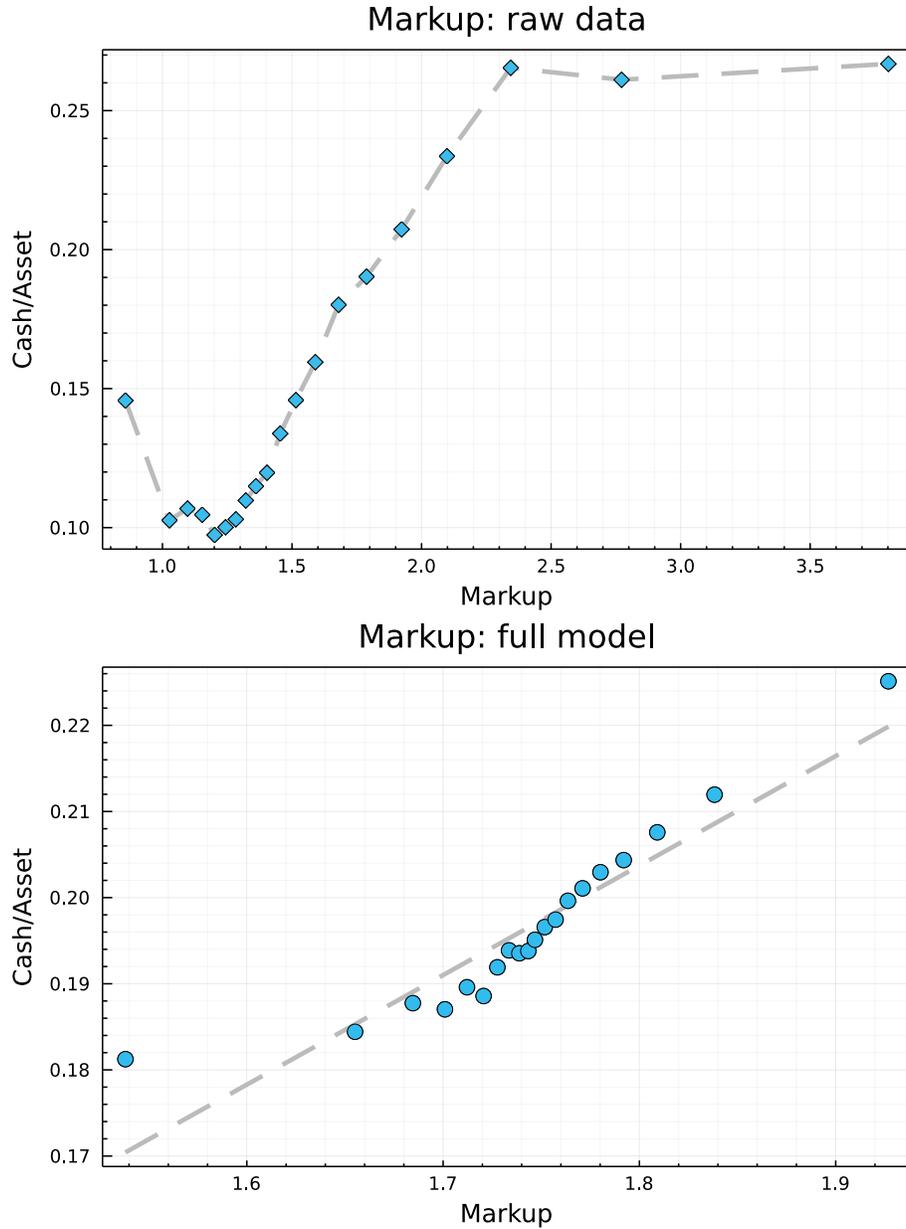
*Notes:* This figure presents the time-series plots of the markup volatility for five groups of firms with different levels of markups. Main data source for this figure is the Compustat North American Annual data file. Firm-level markup is estimated with [Loecker, Eeckhout and Unger \(2020\)](#)'s production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

**Figure 9: 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*).

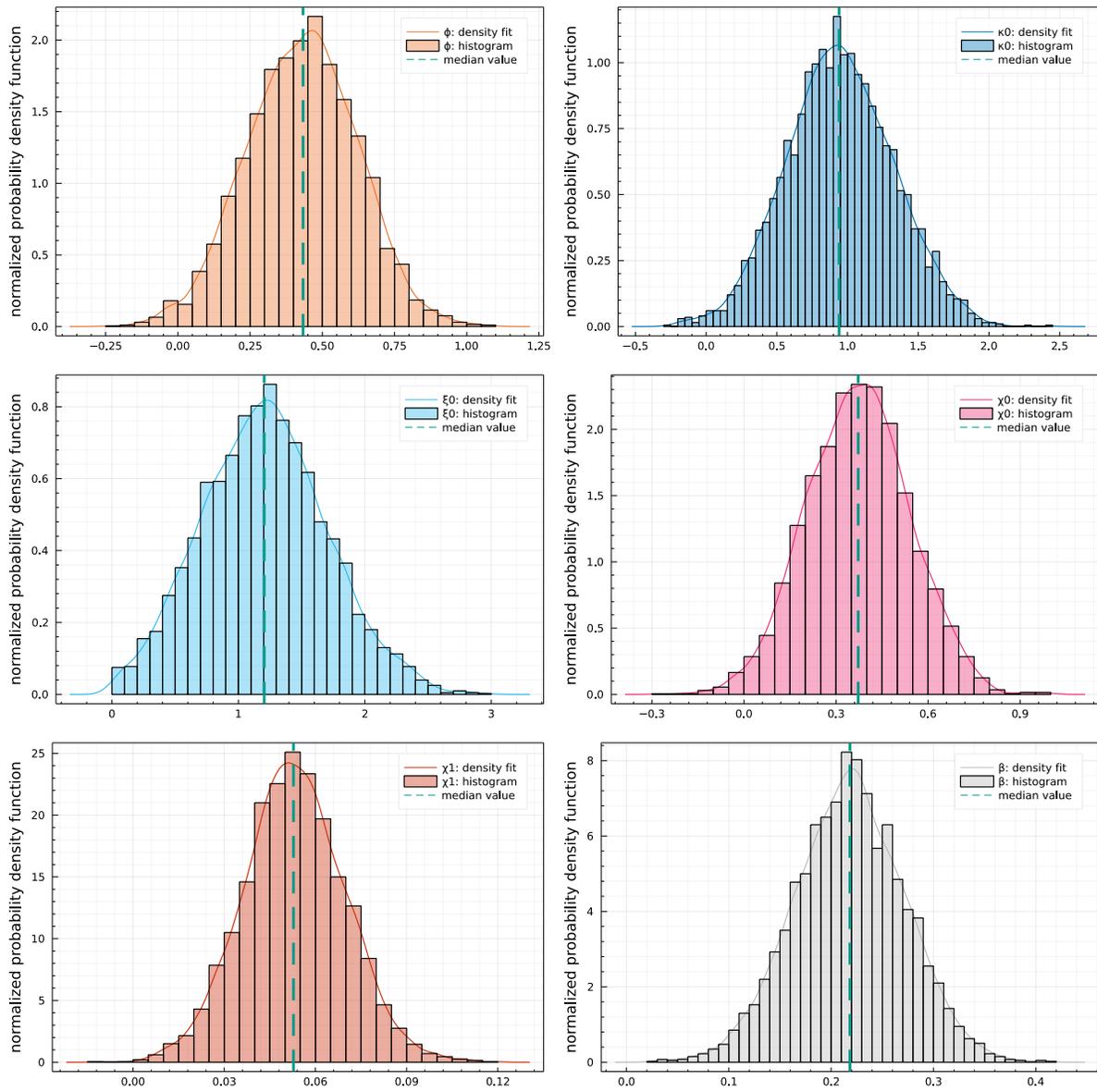
**Figure 10: Tobin's  $q$ , total  $q$ , markup, and corporate cash holdings: Binscatter plots**



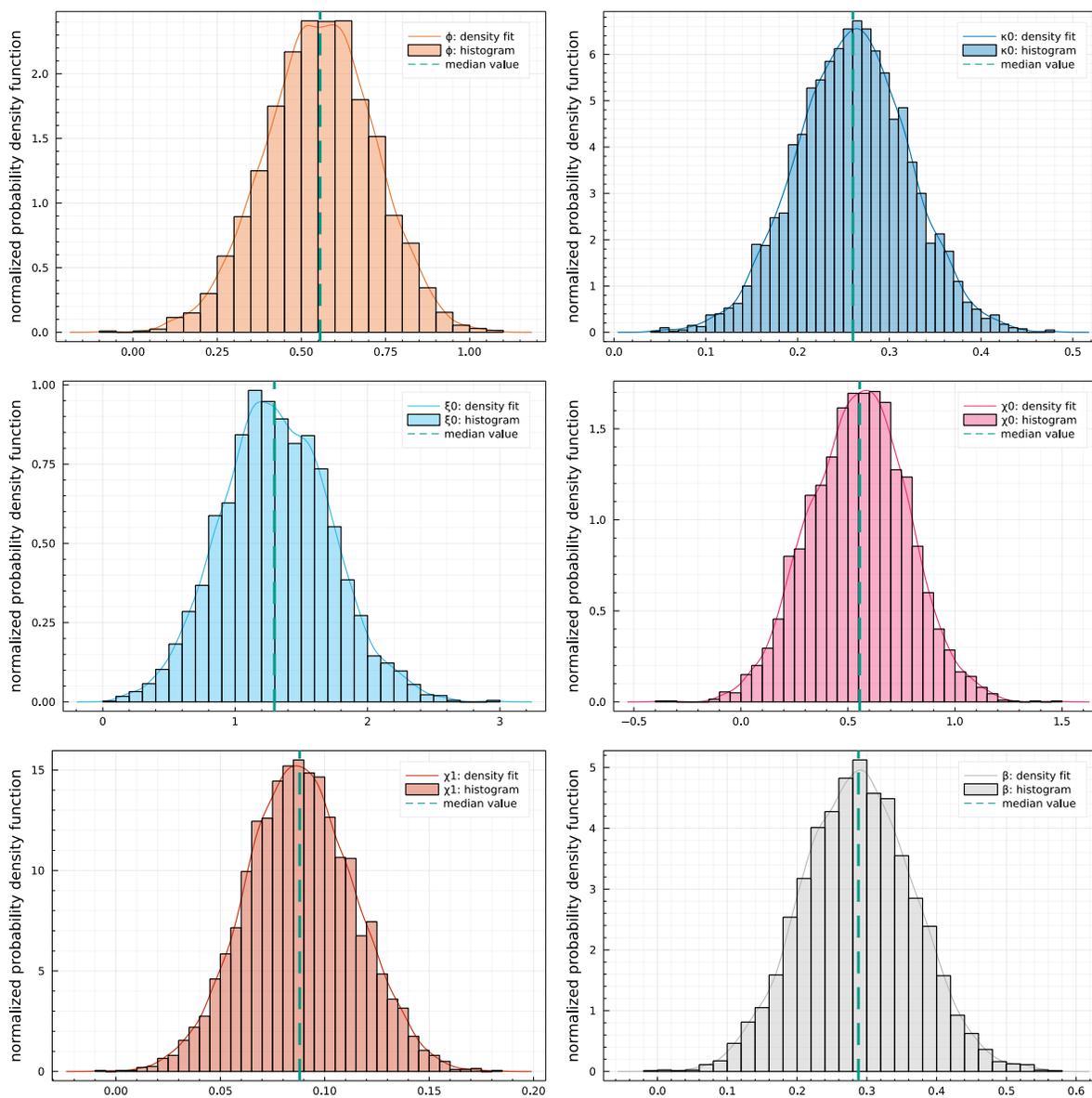
*Notes:* This figure presents the Binscatter plots between corporate cash holdings and markup. Main data source for this figure is the Compustat North American Annual data file. Firm-level cash-to-asset ratio is measured as the ratio of cash and short-term investments (Compustat series *CHE*) to firm's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

Figure 11: SMM-MCMC parameters posterior distribution

A. Traditional Economy: 1980-1999

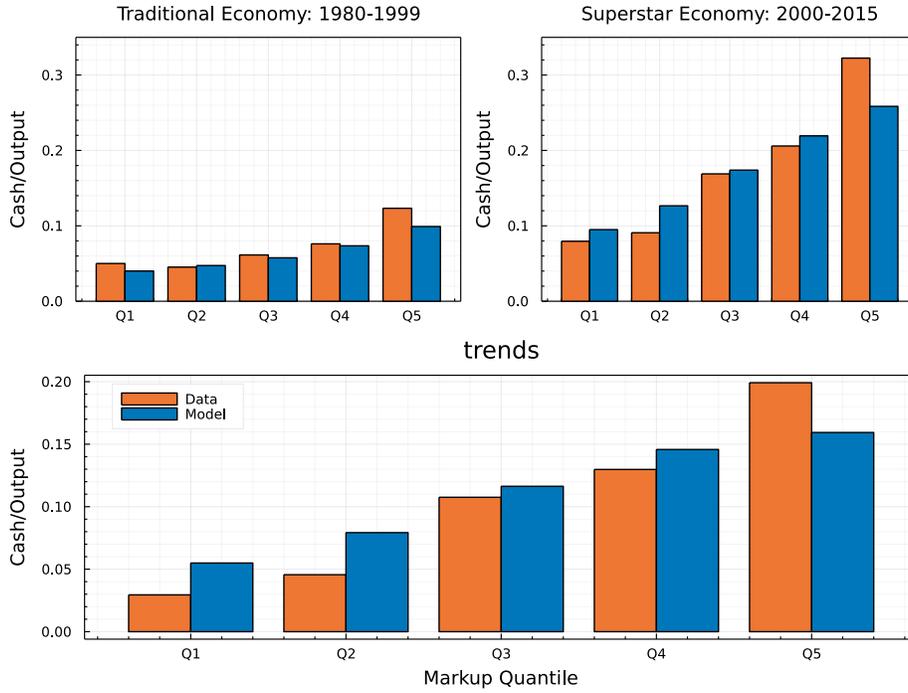


## B. Superstar Economy: 2000-2015

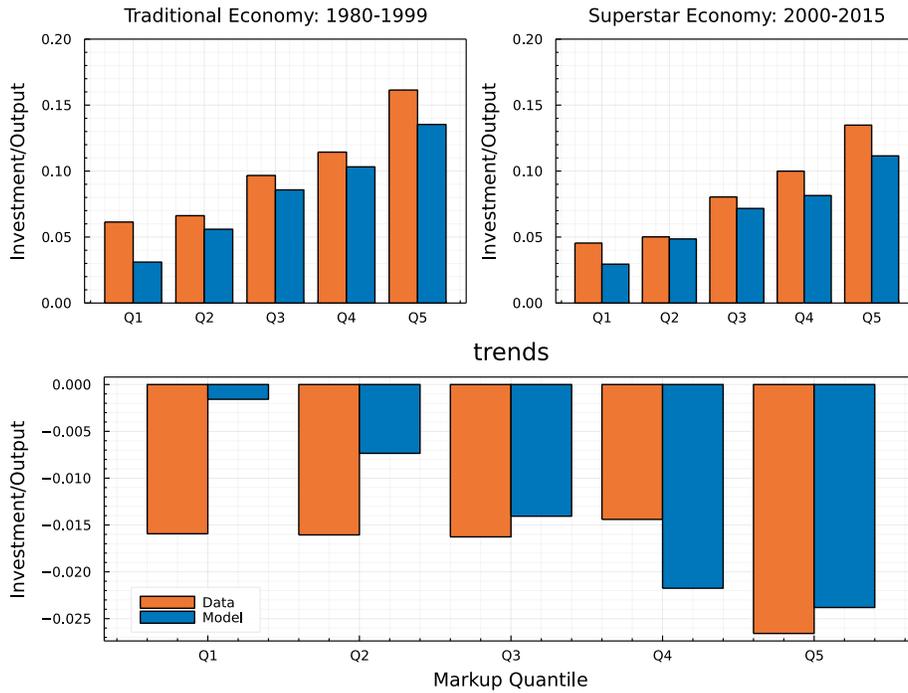


**Figure 12: Cross-sectional validation: data v.s. model**

**A. cash/output ratio**

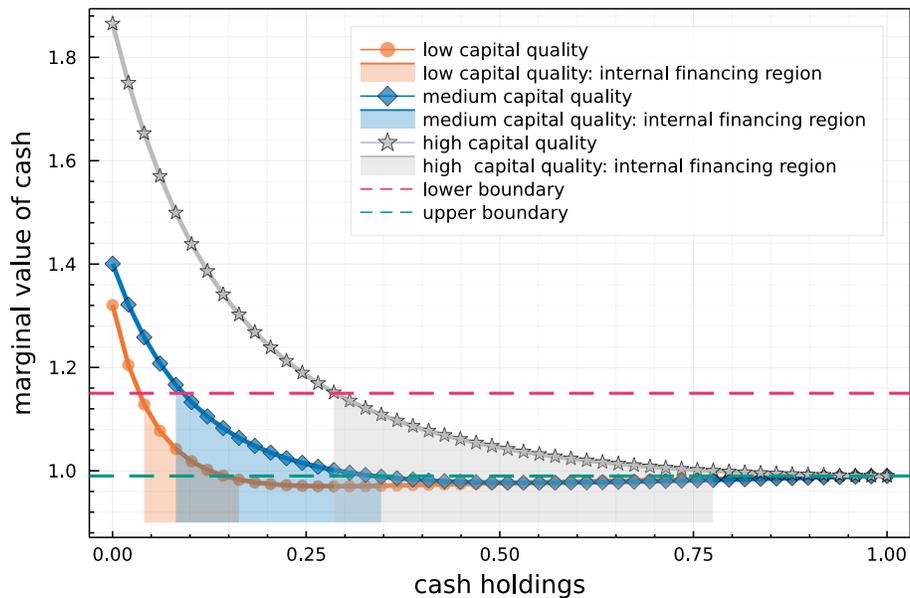


**B. investment/output ratio**

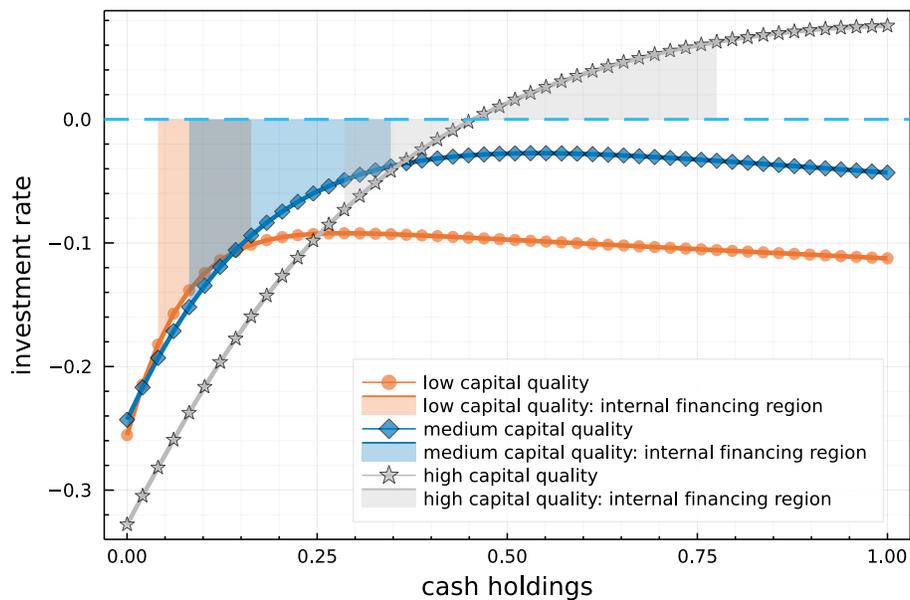


**Figure 13: Policy functions**

**A. cash holding**



**B. investment**



**Table 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*** (3.012)	0.0080*** (9.910)	0.0269*** (34.457)	0.0086*** (10.583)	0.0054*** (6.626)	0.0080*** (9.961)	0.0113*** (9.377)	0.0080*** (9.922)	0.0083*** (9.585)	0.1177*** (24.798)	0.1783*** (39.656)	0.1763*** (38.899)
markup square	-0.0002 (-0.422)									-0.0160*** (-18.619)	-0.0251*** (-29.197)	-0.0251*** (-29.032)
return of assets		-0.0000 (-0.071)								0.0002 (0.863)	0.0031*** (10.100)	0.0030*** (9.852)
tangibility			-0.3761*** (-132.893)							-0.5582*** (-108.150)	-0.5386*** (-118.906)	-0.5385*** (-117.821)
investment				-0.0767*** (-18.388)						0.0741*** (9.165)	0.2033*** (20.450)	0.2064*** (20.854)
size					-0.0106*** (-24.885)					-0.0003 (-0.537)	-0.0050*** (-17.237)	-0.0049*** (-16.717)
profitability						-0.0003*** (-3.289)				-0.0010*** (-3.683)	-0.0022*** (-5.790)	-0.0022*** (-5.813)
R&D							-0.0165*** (-15.186)			-0.0122*** (-8.715)	0.0369*** (21.741)	0.0344*** (20.426)
book leverage								0.0001*** (4.901)		0.0001* (1.800)	-0.0002*** (-3.772)	-0.0002*** (-3.376)
payout									0.0020 (1.180)	0.0040 (1.623)	0.0046 (1.394)	0.0047 (1.406)
	<b>Fixed effects</b>											
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry (naics3)												Yes
Industry × Year												Yes
<i>N</i>	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 <i>R</i> <sup>2</sup>	0.635	0.635	0.667	0.637	0.636	0.634	0.692	0.635	0.618	0.725	0.336	0.347

**Table 2: Model calibration and estimation**

<b>Part A. External Calibration</b>				
PARAMETER	DESCRIPTION	TRADITIONAL ECONOMY 1980-1999	SUPERSTAR ECONOMY 2000-2015	SOURCE/REFERENCE
$\rho$	rate of time preference		0.046	
$\gamma$	risk aversion		4.0	Wang, Wang and Yang (2012)
$\theta$	EIS reciprocal		2.0	
$\lambda$	cash carry cost		0.01	Bolton, Chen and Wang (2011)
$\delta$	capital depreciation rate	0.053	0.056	BEA-FAT
$\eta$	operating scale	0.48	0.64	
$f_0$	fixed production cost	0.11	0.32	Compustat
$\bar{\mu}$	capital quality: long-run mean		1.48	
$\sigma$	capital quality: volatility		0.76	
<b>Part B. Internal Estimation</b>				
PARAMETER	DESCRIPTION	TRADITIONAL ECONOMY 1980-1999	SUPERSTAR ECONOMY 2000-2015	DIFFERENCE
$\phi$	taste for quality	0.43	0.56	+ 0.13
$\xi_0$	variable production cost	0.94	0.26	-0.68
$\kappa_0$	investment adjustment cost	1.20	1.30	+0.10
$\chi_0$	fixed external financing cost	0.37	0.55	+0.18
$\chi_1$	variable external financing cost	0.053	0.088	+0.035
$\beta$	tightness of borrowing constraint	0.22	0.29	+0.07

**Table 3: Target moments and parameters uncertainty**

<b>Part A. Goodness of Fit</b>					
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	

<b>Part B. Parameters Uncertainty</b>								
PARAMETER	TRADITIONAL ECONOMY: 1980-1999				SUPERSTAR ECONOMY: 2000-2015			
	P10	P90	S.D.	MEDIAN	P10	P90	S.D.	MEDIAN
$\phi$ (taste for quality)	0.186	0.672	0.187	0.432	0.346	0.759	0.161	0.556
$\kappa_0$ (investment adjustment cost)	0.466	1.433	0.376	0.941	0.183	0.338	0.0597	0.260
$\xi_0$ (variable production cost)	0.565	1.844	0.492	1.204	0.789	1.829	0.411	1.296
$\chi_0$ (fixed external financing cost)	0.161	0.591	0.167	0.372	0.249	0.833	0.227	0.555
$\chi_1$ (variable external financing cost)	0.0320	0.0744	0.0167	0.0528	0.0559	0.122	0.0258	0.0880
$\beta$ (tightness of borrowing constraint)	0.149	0.286	0.0535	0.218	0.188	0.390	0.0796	0.288

**Table 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%

**Table 5: Counterfactual exercises**

MACRO-FINANCE TRENDS	DATA	MODEL						
		Fix $\phi$	Fix $\eta$	Fix $f_0$	Fix $\xi_0$	Fix $\eta, f_0, \& \xi_0$	Fix $\phi, \eta, f_0, \& \xi_0$	Fix $\beta$
degree of “misallocation” (% of the full model)	+0.22 -	+0.18 (58.06%)	+0.25 (80.65%)	+0.26 (83.87%)	+0.14 (45.16%)	+0.10 (32.26%)	+0.08 (25.81%)	+0.28 (90.32%)
correlation between TFP and net finance (% of the full model)	-0.164 -	-0.0713 (54.43%)	-0.116 (88.55%)	-0.113 (86.26%)	-0.0868 (66.26%)	-0.0404 (30.84%)	+0.023 (-17.56%)	-0.129 (98.47%)
MPK - $r$ (% of the full model)	+5.00% -	+3.33% (64.66%)	+3.91% (75.92%)	+3.85% (74.76%)	+2.80% (54.37%)	+1.60% (31.07%)	+1.04% (20.19%)	+4.72% (91.65%)
area disciplined by the price mechanism (% of the full model)	N/A -	-7.28% (66.91%)	-9.34% (85.85%)	-9.26% (85.11%)	-5.50% (50.55%)	-3.25% (29.87%)	-3.17% (29.14%)	-10.68% (98.17%)