Firm Life-Cycle Learning and Misallocation

Ying Feng

UC San Diego

February, 2019

Latest version here.

Abstract. Dispersion in marginal revenue products of capital (MRPK) across firms may lower aggregate productivity through misallocation. Using firm-level panel data from China, Columbia, and Chile, I document that MRPK dispersion decreases substantially with firm age, particularly before age five. Building on this fact, I provide a new interpretation of MRPK dispersion as resulting from firm life-cycle learning. I formalize this idea in a dynamic model in which firms learn about their fundamental productivity as they age and choose capital inputs in a frictional market based on their priors. Within each cohort of firms, imprecise priors lead firms to differ in their ex-post MRPK even in the absence of firm-level distortions. As firms learn over time and adjust their capital stocks, possibly through exiting the market, dispersion in MRPK decreases. Quantitative analysis of the model shows that firm life-cycle learning accounts for 7% of MRPK dispersion in the economy, which can lead to a 10% loss in aggregate productivity.

Email: yif014@ucsd.edu. I am grateful to my advisors David Lagakos and James Rauch for their guidance and support. For helpful comments, I thank Jim Hamilton, Pete Klenow, Richard Rogerson, Tommaso Porzio, and seminar/conference audiences at Cal State Fullerton, CityUHK, HKU, ITAM, New Structural Economics Conference, NUS, SAIF, Tsinghua, and UCSD. I also benefit from numerous discussions with Emilien Gouin-Bonenfant and Xiao Ma. All potential errors are my own.

1 Introduction

Differences in average income levels across countries are vast. Development accounting points to differences in total factor productivity (TFP) as an important proximate cause of cross-country income differences (Caselli, 2005). Yet the determinants of TFP are still not well understood. A prominent theory of TFP emphasized in the recent literature is misallocation. Two influential papers, Hsieh and Klenow (2009) and Restuccia and Rogerson (2008), interpret dispersion in marginal revenue products of capital (MRPK) across firms as the result of firm-level 'distortions' that cause misallocation. They argue that misallocation leads to large TFP losses in developing countries. Bachas, Fattal-Jaef, and Jensen (2018) provide evidence that size-dependent tax, a form of distortions, are more prevalent in low-income countries.

However, the literature is still very much undecided about how to interpret dispersion in MRPK across firms (Restuccia and Rogerson, 2017). A large body of work has provided alternative interpretations. For example, Asker, Collard-Wexler, and De Loecker (2015) emphasize the role of capital adjustment costs under volatility of productivity, and David, Hopenhayn, and Venkateswaran (2016) emphasize the role of uncertainty in contemporaneous productivity. Both channels lead to MRPK dispersion but do not imply misallocation from distortions. David and Venkateswaran (2017) further develop a quantitative framework to decompose sources of MRPK dispersion and conclude that, while these channels are present, a large share of dispersion still results from firm-level distortions. An open question in the literature remains: What are the sources of MRPK dispersion?

This paper provides a new source of MRPK dispersion, building on a new pattern I document in the data. Following firm cohorts using firm-level panel data from China for the period 1998 - 2007, I document that MRPK dispersion across firms decreases substantially with firm age, particularly before age 5. The magnitude of this life-cycle decrease is similar to the difference in MRPK dispersion between China and the US as reported in Hsieh and Klenow (2009). Furthermore, for young firms, MRPK dispersion decreases at a decreasing rate with firm age.

Yet the challenge is that identifying the age effects separately without any additional assumptions is impossible, because the age, year, and cohort indicators are collinear. In particular, during my sample period, China experienced massive privatization reforms so that revenue share of the state-owned firms in the industrial sector declined by 20 percent (Hsieh and Song, 2015). One can expect large year effects as China underwent such reforms and opened up to international trade. Hence, the decrease in MRPK dispersion over a firm cohort's life cycle could be the result of year effects, rather than age effects. Similarly, one can expect that each successive cohort may be founded with less MRPK dispersion across firms, as they entered the market more for economic reasons rather than political reasons. Thus, including controls for year effects and cohort effects is crucial in any reasonable attempt to identify the age effects on MRPK dispersion.

My preferred identification approach imposes the testable assumption of a linear trend in the age effects at older ages. For example, consider a special case of linear effects as no trend in the age effects on MRPK dispersion after firm age 10. Then year effects can be identified by following the same firm cohorts aging from age 10 because all the changes over time are only due to year effects in the absence of cohort and age effects. I can subsequently identify the age effects and cohort effects after knowing the year effects. Specifically, in the preferred approach, I imposed three plausible trends of age effects at older ages for identification. I also provide two alternative identification approaches in the paper. All three estimation results show negative age effects on dispersion in MRPK across firms. In particular, the estimated profile of the standard deviation of log MRPK within a firm cohort always decreases by more than 0.2 before age five, which accounts for 13% of the initial dispersion at firm entry.

Building on the facts I document, I provide a new interpretation of MRPK dispersion as resulting from firm life-cycle learning. It reflects informational frictions over the firm cohort's life cycle when firms learn about their own fundamental productivity, as in Jovanovic (1982). Within each cohort of firms, differences in the precision of priors lead firms to differ in their ex-post MRPK even in the absence of firm-level distortions. I formalize this idea in a dynamic model in which firms learn over time and choose capital inputs based on their priors in a frictional market with firm-specific distortions. Qualitatively, as priors of the firm cohort improve through learning over time, firms with too much or too little capital stock adjust, and the less productive firms within a cohort exit. Hence, the model predicts that MRPK dispersion within the firm cohort decreases as firms age.

The main quantitative experiment is to compute the model's predictions about MRPK dispersion within a firm cohort as the firms age. To do so, I take the joint distribution of productivity and capital stocks among firm entrants as given in the data. I calibrate the model to match three key moments in the data, namely, the exit rate of firm entrants, the correlation between productivity and capital investment, and the autocorrelation of capital investments. As a result, for the first ten years of the firm cohort's life cycle, the calibrated model accounts for around two thirds of the decrease in MRPK dispersion in the data.

To understand the quantitative role of learning, I decompose changes in MRPK dispersion over the firm cohort's life cycle by sequentially adding mechanisms in the model. If the firms adjust capital stocks without updating their priors and without exiting the market, MRPK dispersion barely decreases with firm age. If firms Bayesian update their priors while adjusting capital stocks, but still do not have the exit option, the dispersion in MRPK decreases around half as much as the benchmark model prediction. Further adding endogenous firm exit under the life-cycle learning accounts for the other half of the benchmark model prediction.

What, then, are the consequences of firm life-cycle learning for aggregate TFP, rather than for a firm cohort? Taking into account the firms' age distribution in the stationary equilibrium, I compare the benchmark model predictions to a hypothetical baseline where young firm cohorts had already completed their learning process as older firms. This comparison suggests that informational frictions from firm life-cycle learning lead to a 10 percent loss in aggregate TFP. I conduct the same analysis in the model after removing firm-level distortions, which suggests that distortions and firm life-cycle learning together result in a 19 percent loss in aggregate TFP. Therefore, omitting the contribution of firm life-cycle learning to MRPK dispersion causes more than half of TFP losses to be incorrectly attributed to distortions. I regard these estimates as lower bounds of TFP losses because the quantitative analysis assumes MRPK dispersion across firms remains constant after age 10.

Before concluding, I present plant-level panel data from Colombia and Chile for an earlier period (around the 1980s). I ask whether MRPK dispersion (measured by standard deviations of log MRPK) decreases with firm-cohort age. I find that, in both countries, MRPK dispersion decreases by around 0.4 through the first five years of the firm cohort's life cycle, which accounts for 29% and 24% of the initial dispersion across age-zero firms in Colombia and Chile, respectively. I conclude that data from other developing countries broadly show decreasing life-cycle MRPK dispersion, similar to the data from China.

Related Literature. Most existing work focuses on the aggregate level of MRPK dispersion across firms and does not consider its dynamics over the firm cohort's life cycle. For example, Midrigan and Xu (2014) and Li (2015) study financial frictions, Kehrig and Vincent (2017) combine financial frictions and adjustment costs to investigate MRPK dispersion across plants within the same firms, Ho and Ruzic (2018) explore the variation in markups and returns to scale, and Song and Wu (2015) consider markup dispersion, adjustment costs, and measurement errors. In addition, all the models above are silent on endogenous firm entry and exit. Yang (2016) studies the effects of distortions on firm entry but assumes exogenous exit. Fattal-Jaef (2018) emphasizes that, in theory, endogenous exit may offset the effects of distortions on long-run TFP, but does not consider informational frictions. This paper is the first to look at life-cycle MRPK dispersion and the first to interpret MRPK dispersion as resulting from firm learning.[1]

[1] See Restuccia and Rogerson (2017) for an in-depth literature review on the causes and costs of misallocation. Other studies focus on misallocation over the business cycle: Argente, Lee, and Moreira (2018) consider the reallocation of products, and Senga (2015) emphasizes the rising uncertainty at the start of the Great

This paper also relates to the literature in macroeconomics that makes cross-country comparisons of average firm sizes over firms' life cycles. Hsieh and Klenow (2014) show that plants stay much smaller in Mexico and India than in the US over the plants' life cycles. Bento and Restuccia (2017) use data from more countries to argue that severe distortions in developing countries discourage investments, leading to smaller average firm sizes. Akcigit, Alp, and Peters (2016) and Chen, Habib, and Zhu (2018) emphasize the importance of delegation frictions and lack of selection in explaining smaller average plant sizes over the plants' life cycles in developing countries. My results pertain to MRPK dispersion across firms rather than average firm size. The fact that the dispersion decreases with firm age implies considerable improvement in how efficiently resources are allocated across firms over their life cycles.

The idea of firm life-cycle learning is built on the classic model of Jovanovic (1982). By adding capital to his original model, I bring in frictional capital markets, including adjustment costs and fire-sale discounts upon exit. These frictions are important to match the key pattern of life-cycle MRPK dispersion within a firm cohort. In addition, other studies, for example, Asturias, Hur, Kehoe, and Ruhl (2017), emphasize that exits of low-productivity firms contribute to aggregate productivity growth. By focusing on MRPK dispersion across firms, this paper can draw further implications of the consequences of firm life-cycle learning and exit for aggregate productivity through reallocating resources across firms.

Finally, this paper adds to the vast literature on the theories of TFP, aiming at advancing our understanding of income differences across countries and across time. For example, Guner, Ventura, and Xu (2008) consider the macroeconomic implication on reductions in output of size-dependent policies. Buera, Kaboski, and Shin (2011) quantify the role of financial frictions in economic development. David and Venkateswaran (2017) emphasize distortions accounts for a larger share of misallocation among Chinese manufacturing firms and adjustment costs are more salient for large US firms. Calligaris, Del Gatto, Hassan, Ottaviano, and Schivardi (2018) argue that resource misallocation has played a sizable role in slowing down Italian productivity growth. This paper points to the potentially important role of firm life-cycle informational frictions and learning.

The rest of this paper is structured as follows. Section 2 describes the data from China and presents the features of life-cycle MRPK dispersion across firms without any additional assumptions. Section 3 reports the estimated profile of MRPK dispersion with firm age while controlling for cohort effects and year effects. Section 4 presents the model with learning and its qualitative predictions. Section 5 discusses the quantitative analysis. Section 6 provides evidence from Colombia and Chile. Section 7 concludes.

Recession. Bai, Jin, and Lu (2018) consider misallocation in an open economy with trade liberalization.

2 MRPK Dispersion across Firms over Their Life Cycles

In this section, I describe the data and present cross-sectional evidence on the pattern of MRPK dispersion over the firm cohort's life cycle. When tracking each firm cohort over time, I find that the dispersion in MRPK across firms always decreases with firm age. The younger cohorts tend to have smaller MRPK dispersion than older cohorts, and the aggregate MRPK dispersion in a year also decreases during my sample period (1998 to 2007). I also find that, for firm cohorts before age five, MRPK dispersion decreases at a decreasing rate.

2.1 Data and Measurement

I use the Annual Industry Surveys for 1998 - 2007 conducted by the National Bureau of Statistics of China. The survey covers all the state-owned firms in the manufacturing sector, as well as non-state-owned manufacturers with sales revenue above 5 million RMB (around 0.7 million USD). I follow the procedure used by Brandt, Van Biesebroeck, and Zhang (2012) to construct the panel data. I start by matching the firms over time by registration ID. When firms changed their registration ID due to restructure or acquisition, I use company name, phone number, and address to identify the same firms. Note that ownership change will not cause false exits in the data, because those firms will still be identified over time through address and name. Throughout the paper, I focus on the firm cohorts founded after 1978, when the "opening-up reform" started. I drop firms founded in a planned economy before the economic reform because they may operate under very different systems. In addition, those firms cannot be observed at ages younger than 20, and are thus less relevant for studying the life cycles of firms. The remaining panel data have an average of around 180,000 firms per year, growing from 106,000 firms in 1998 to over 298,000 firms in 2007. In addition, I use the 4-digit Chinese Industry Code (CIC), birth year, wage, employee benefits, value-added, and capital stock. [2]

Let i denote an individual firm. The firm age j is calculated as the survey year minus the reported birth year. Therefore, the age-one firms are operating for a full year. Let y_{it} denote the revenue output, k_{it} the capital input, and n_{it} the labor input. Then y_{it} is measured as value added, k_{it} is measured as the book value of fixed capital net of depreciation of the year, and n_{it} is measured as the total of wage payments and employee benefits. The employee benefits include unemployment insurance, old care insurance, medical insurance, housing compensation, travailing compensation, and union expenses, but availability of the specific variable varies across years. Hence, I inflate the labor share to match those reported

[2] The share of firms younger than age 10 is around 72% in every year of my sample.

in the annual national accounts as Hsieh and Klenow (2009) did. This procedure assumes the imputed values of missing benefits are a constant fraction of labor income.To summarize dispersion of the key variables over the firms' life cycles, Appendix Figure A.1 plots the standard deviation of log value-added (y_{it}), log capital input (k_{it}), log labor input (n_{it}), and log employment by firm age. I find the dispersion of value added and labor input across firms increases with firm age until age 15, while dispersion of capital input and employment across firms increases very marginally with firm age.

Let the production function be Cobb-Douglas $y_{it} = e^{z_{it}} k_{it}^{\alpha_1} n_{it}^{\alpha_2}$. I assume decreasing returns to scale, that is, $\alpha_1 + \alpha_2 < 1$. I also allow the capital and labor input share to vary across industries but not over time as in Hsieh and Klenow (2009). Following their work, I use the NBER-CES Manufacturing Industry Database to calculate α_1 as the average values of capital share at 4-digit SIC level during the period 1987-2011, and then match them to CIC at the 2-digit level. In the empirical analysis, I set $\alpha_1 + \alpha_2$ to be the standard 0.85.[3] By definition, the MRPK of firm i at time t is $\frac{\partial y_{it}}{\partial k_{it}} = \alpha_1 \frac{y_{it}}{k_{it}}$. I measure total factor revenue productivity (TFPR), marginal revenue product of labor (MRPL), and MRPK in log terms throughout the paper:

$$tfpr_{it} = log(y_{it}) - \alpha_1 log(k_{it}) - \alpha_2 log(n_{it}) \tag{1}$$

$$mrpk_{it} = log(\alpha_1) + log(y_{it}) - log(k_{it}). \tag{2}$$

$$mrpl_{it} = log(\alpha_2) + log(y_{it}) - log(n_{it}). \tag{3}$$

I drop the observations with missing values and trim the 1% tails of measured MRPK or TFPR in each industry-year-age group. The remaining data have an average of around 169,000 firms per year, consisting of more than 480,000 unique firms recorded during the sample period. Around 48% of the unique firms survived for at least four years.

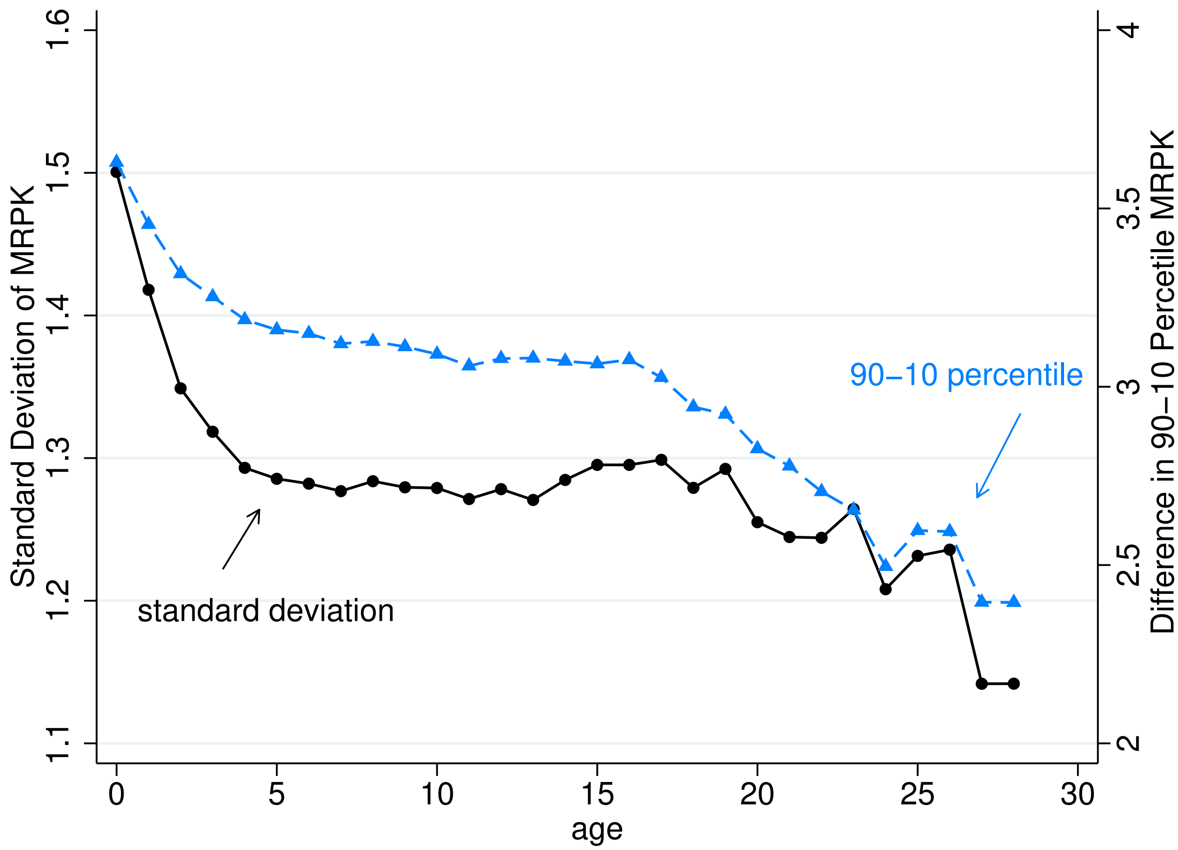
2.2 Dispersion of Marginal Products by Firm Age

Consider an industry-year-age bin, denoted as stj, consisting of firms observed in calendar year t at firm age j in the 4-digit industry s. To measure MRPK dispersion within a stj bin, I use the standard deviation of $mrpk_{it}$, denoted as $\sigma_{mrpk,stj}$, and the 90th minus the 10th percentile $mrpk_{it}$, denoted as $D^{90-10}_{mrpk,stj}$. Note that, by construction, MRPK dispersion across firms within an industry-year-age bin is always measured within the same firm cohort c of firms, which are founded in year $t - j$. When summarizing MRPK dispersion at a give

[3] I conduct the same analysis assuming constant returns to scale as in Hsieh and Klenow (2009), letting $\alpha_2 = 1 - \alpha_1$, and get similar results.

age or in a given year, I will always weight $\sigma_{mrpk,stj}$ by N_{stj}, the number of firms in an industry-year-age bin.

Figure 1: Dispersion of MRPK by Firm Age



Note: This figure plots the weighted average standard deviation of MRPK ($\bar{\sigma_j}$) and the weighted average value of the 90th minus the 10th percentile ($\bar{D}_j^{90-10}$) by firm-cohort age.

Define the weighted average dispersion in MRPK at a given firm-cohort age j in both measures

$$\bar{\sigma_j} \equiv \sum_s \sum_t \sigma_{mrpk,stj} \cdot \omega_{st}$$

$$\bar{D}_j^{90-10} \equiv \sum_s \sum_t D_{mrpk,stj}^{90-10} \cdot \omega_{st},$$

where the weight $\omega_{st} = \frac{N_{stj}}{\sum_s \sum_t N_{stj}}$. Figure 1 then plots $\bar{\sigma_j}$ (black solid line with round markers) and $\bar{D}_j^{90-10}$ (blue dashed line with triangle markers) by firm-cohort age. The weighted average standard deviation of MRPK, $\bar{\sigma_j}$, deceases substantially from 1.5 by almost 0.4 points by age 28. Note that this change is huge, and has the same magnitude as the difference between

China and the US reported in Hsieh and Klenow (2009).[4] The plotted average log differences, $\bar{D}_j^{90-10}$, show the average ratio of the 90th to the 10th percentile MRPK decreases from more than 33 ($e^{3.5}$) to only 12 ($e^{2.5}$) as the firm cohort ages from zero to around 25. Similar to the patterns of MRPK standard deviations, the 90-10 percentile difference in MRPK decreases substantially with firm age. I will focus on the standard deviation of MRPK ($\sigma_{mrpk,stj}$) as the dispersion measure for the rest of the paper because it has been used more broadly in the literature. More importantly, I will show later that $\sigma_{mrpk,stj}$ translates directly to TFP losses.[5]

Meanwhile, how does the dispersion of MRPL across firms change as firms age? Appendix Figure A.2 plots the weighted average standard deviation of MRPL and the weighted average value of the 90th minus the 10th percentile MRPL by firm age. Both measures of dispersion in MRPL decrease very marginally before age 5, and they increase slightly afterward. Therefore, this paper focuses on MRPK dispersion and abstracts from the discussion of MRPL dispersion.

A natural explanation is that, MRPK dispersion decreases with firm age because the less productive firms within a cohort learn about their type and exit gradually over their life cycles. This implies that TFPR dispersion also decreases at a decreasing rate with firm age. As learning and the selection in firm exits becomes less pronounced, the dispersion in both MRPK and TFPR does not decrease further. Denote the weighted average standard deviation of TFPR at age j as $\bar{\sigma}_{tfpr,j}$. Appendix Figure A.3 shows that $\bar{\sigma}_{tfpr,j}$ decreases at a decreasing rate from around 0.99 at entry to around 0.9 at age five, and fluctuates between 0.85 and 0.95 afterward. Alternatively, if one thinks of the production process as the less productive firms catching up with the most productive firms due to stable innovation investments or spillover effects, one should expect TFPR dispersion to continuously decrease at older ages, which is not observed in the data.[6] Furthermore, Appendix Figure A.4 plots MRPK dispersion for the balance panel, which consists of firms that I can observe every year during the sample period of 1998-2007. It shows that, when firm exits are shut down, MRPK dispersion decreases with firm age with a smaller magnitude.

However, the summary statistics by firm-cohort age presented above is the result of a combination of age, cohort, and year effects. In a fast-changing economy like China, one may expect large variations across firm cohorts born in different years. For example, as China

[4] Table 2 in their paper reports the difference of 0.14 in the standard deviation of $tfpq$ between China and the US in 2005. It implies a MRPK dispersion difference of 0.4 points in their model under the standard capital share of 0.3.

[5] As an alternative measure of dispersion in MRPK, the average ratio of the 75th to the 25th percentile MRPK decreases monotonically from 6 to 4 as the firm cohort ages from zero to around 25.

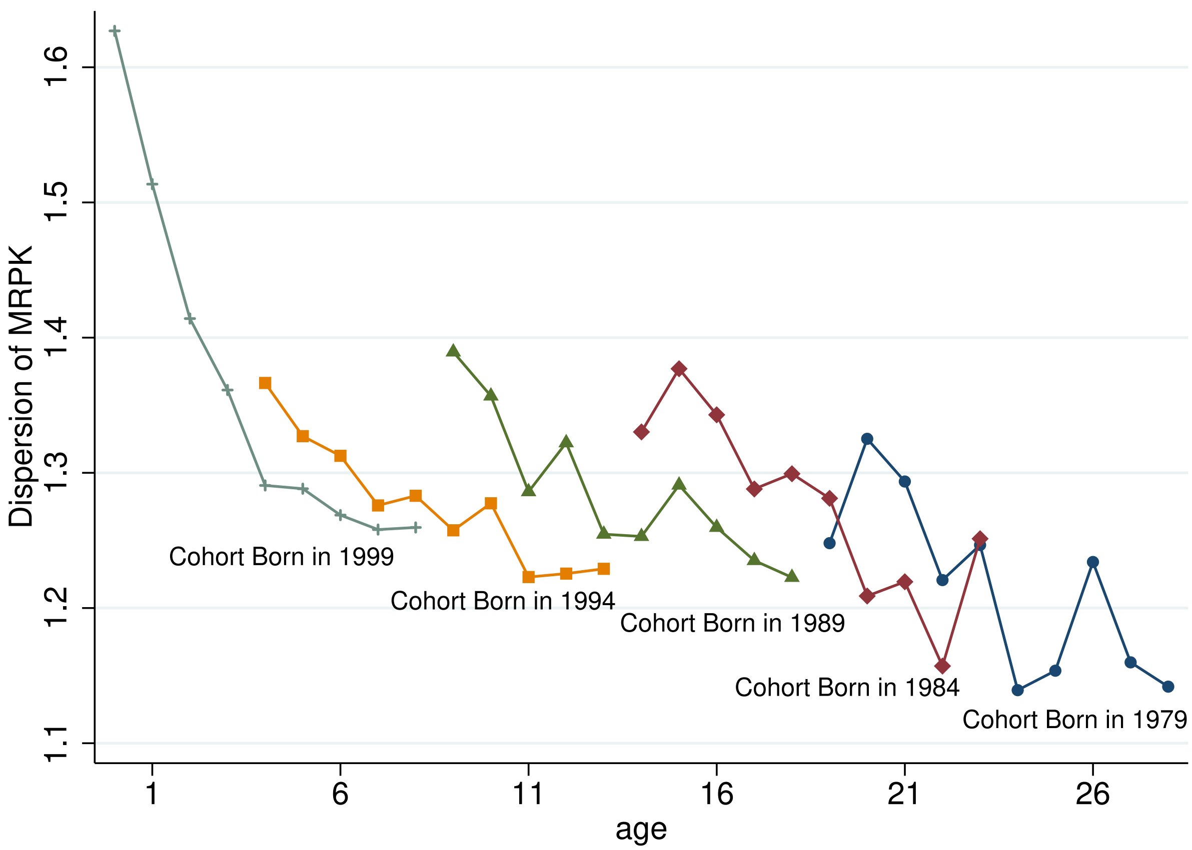
[6] Note that $\bar{\sigma}_{tfpr,j}$ is at a lower scale than $\bar{\sigma}_{mrpk,j}$, due to a large dispersion in k_{it}, which is not offset by the empirical correlation between y_{it} and k_{it}, thus being reflected in $\bar{\sigma}_{mrpk,j}$.

moves from an economy of state-owned enterprises to one with mostly private enterprises, each successive cohort of firms may be founded with a smaller dispersion in MRPK. To investigate the life-cycle pattern within each firm cohort $c = t - j$, define the weighted average dispersion across firms at age j as

$$\bar{\sigma}_{jc_{t-j}} \equiv \sum_{s} \sigma_{mrpk,stj} \cdot \omega_{sc},$$

where the weight $\omega_{sc} = \frac{N_{stj}}{\sum_s N_{stj}}$.

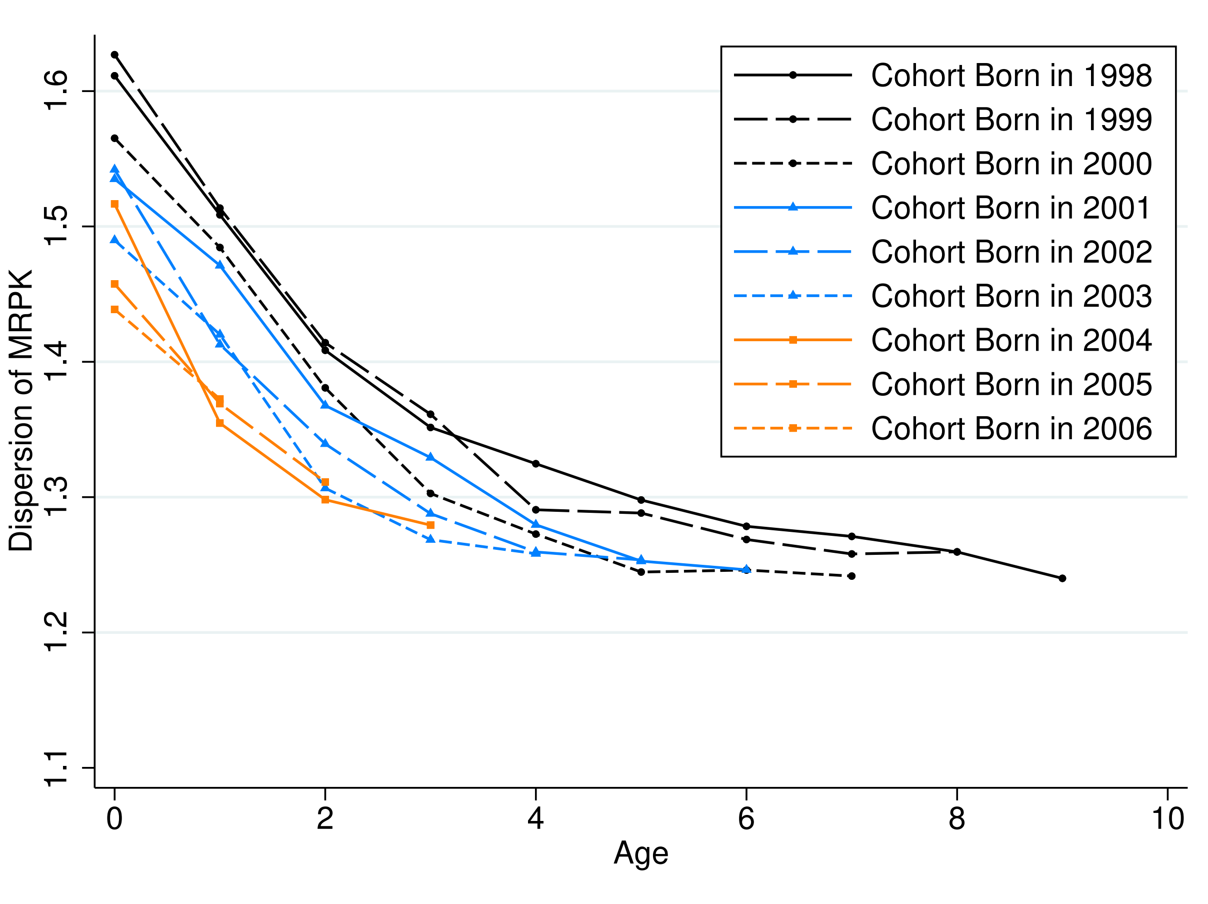
Figure 2: Dispersion of MRPK by Firm Age and Cohort



Note: This figure plots the weighted average dispersion, $\bar{\sigma}_{j,c_{t-j}}$, by firm age within each cohort born in year 1979, 1984, 1989, 1994, and 1999, respectively.

Figure 2 then plots $\bar{\sigma}_{j,c_{t-j}}$ against firm-cohort age by following each cohort born in 1979, 1984, 1989, 1994, and 1999, respectively. For all the cohorts, we see a general decrease in MRPK dispersion with firm-cohort age. The older firm cohorts tend to have a larger dispersion in MRPK than the younger cohorts at the same ages. In particular, for the firm cohort founded in 1999, MRPK dispersion declines by almost 0.4 from age zero to age eight.

Figure 3: Dispersion of MRPK by Firm Age, Young Cohorts



Note: This figure plots the weighted average dispersion, $\bar{\sigma}_{j,c_{t-j}}$, by firm age within each of the nine cohorts born during 1998 to 2006.

Furthermore, Figure 3 plots MRPK dispersion for the nine firm cohorts founded between 1998 and 2006, which can be tracked from age 0 at the entry year. Similar to the 1999 firm cohort in Figure 2, the decline in MRPK dispersion across firms is substantial through the first five years of the firms' life cycles for all nine cohorts. In addition, the older cohorts among the nine again tend to have a larger MRPK dispersion, as we see the black lines are above the blue and the blue are above the orange.[7]

Similarly, it is difficult to imagine the Chinese economy with little year effects as the privatization reforms deepen over time. Define the weighted average dispersion in a given year t as $\bar{\sigma}_t \equiv \sum_s \sum_j \sigma_{mrpk,stj} \cdot \omega_{sj}$, where the weight $\omega_{sj} = \frac{N_{stj}}{\sum_s \sum_j N_{stj}}$. Appendix Figure A.5 plots $\bar{\sigma}_t$ during my sample period. The average aggregate dispersion in MRPK decreases from 1.4

[7]Appendix Figure A.8 plots the exit rates by firm-cohort age of the same firm cohorts in Figure 2 after removing zero-sum year effects using the identification approach in Section 3.1. Patterns of the exit rates resemble those of the MRPK dispersion.

in 1998 to less than 1.3 in 2007.

In summary, by following each firm cohort over time, I find that MRPK dispersion decreases substantially with firm-cohort age. However, the decrease is not necessarily the result of age effects, because it reflects both age effects and year effects. Instead, the substantial decrease could be the result of potentially sizable year effects, because China underwent its reforms and opened up over time. Therefore, in any reasonable attempt to identify age effects, controlling for cohort effects and year effects is crucial.

2.3 Industry Variations

In different industries, MRPK dispersion decreases at different rates with respect to firm-cohort age. I use this variation to shed light on the mechanisms of decline in MRPK dispersion over the firm cohort's life cycle. Define the weighted average dispersion in MRPK across firms at age j in industry s as $\bar{\sigma}_{sj} \equiv \sum_t \sigma_{mrpk,stj} \cdot \omega_t$, where the weight $\omega_t = \frac{N_{stj}}{\sum_t N_{stj}}$. I investigate how the correlation between $\bar{\sigma}_{sj}$ and firm age j varies across industries, and discuss how it relates to the industry characteristics.

Within each industry, I use κ_s in the linear model below to summarize MRPK dispersion over the firm cohort's life cycle:

$$\sigma_{mrpk,stj} = \kappa_{0s} + \kappa_s age_{stj} + \epsilon_{stj}. \tag{4}$$

Estimate $\hat{\kappa_s}$ describes how $\bar{\sigma}_{sj}$ changes with firm-cohort age under the linear specification. The average value of $\hat{\kappa_s}$ across industries is -.014, with a standard deviation of 0.026. Therefore, in the majority of industries, MRPK dispersion decreases with firm age, and on average, $\bar{\sigma}_{sj}$ decreases more than 1% per age.

To utilize the standard industry-level characteristics commonly used in the trade literature, I mapped the 4-digit 2003 CIC to the 6-digit US Input-Output classification, and used the industry indexes from Antras (2015). Table 1 reports the results of regressing estimated $\hat{\kappa_s}$ on various industry characteristics with bootstrap standard errors. It shows that when the capital-labor ratio increases by 1 log point, the decrease of $\sigma_{mrpk,stj}$ per age is 0.005 points larger, which is more than one third of the average value 0.014 across all industries. The significant positive coefficient on log capital per worker suggests that industries with higher capital shares are better at decreasing MRPK dispersion over the firm cohort's life cycle. One possible explanation could be that the costs related to capital, such as storage and maintenance costs or adjustment costs, push firms to adjust capital more responsively, reallocating resources to the more productive incumbent firms. Meanwhile, industries that

Table 1: Industry Characteristics and Life-Cycle $\sigma_{mrpk,stj}$

$\hat{\kappa_s}$	(1)	(2)	(3)	(4)
Log(Capital Per Worker)	-.005***	-.005*	-.006**	-.006**
	(.002)	(.002)	(.003)	(.003)
Log(R&D/Sales)		-.002**	-.002**	-.002*
		(.001)	(.0008)	(.0009)
Contractibility			-.006	-.006**
			(.005)	(.003)
Financial Dependence			-.004	.003
			(.003)	(.016)
Input Substitutability			-.00004	-.00004
			(.0002)	(.0003)
Log(Capital per worker)*Financial Dependence				-.002
				(.004)
Obs.	423	408	408	408
R^2	.015	.034	.044	.044

Note: ***, **, and * indicate statistical significance at the 1-percent, 5-percent, and 10-percent levels, respectively. Independent variable Log(Capital per worker) is the industry average calculated during sample period 1998 - 2007 in China. The source of the industry indexes is Antras (2015): industry-level Log(R&D/Sales) is from Nunn-Trefler (US, 2000-2005); upstream Contractibility from Nunn (2007) based on liberal classification; Financial Dependence is measured as the External Capital Dependence from Rajan-Zingales (1997) calculated using 1980s Compustat data. Input Substitutability is measured as the Import Demand Elasticity (based on SITC33).

larger innovation expenditure shares are better at decreasing MRPK dispersion over firm-cohort age. Additionally, a slightly significant positive correlation exists between resource reallocation and contractibility: Industries in which it is easier to contract the sales of their capital inputs at less discounted values upon exit also experience a faster decrease in MRPK dispersion with firm-cohort age. I will build these ideas formally in the dynamic firm model in section 4 to assess their explanatory power.

3 Life-Cycle MRPK Dispersion: Controlling for Cohort and Time

The previous section reports the average MRPK dispersion by firm-cohort age simply. Although understanding the data with minimal structure and assumptions is useful, this exercise does not address certain issues. The identification of first-order age effects is a well-known challenge due to the collinearity between age, year, and cohort indicators. In this section, I address the identification issues.

Though Figures 2 and 3 track the same firm cohorts over time and find a consistent trend of decreasing dispersion with firm-cohort age, they still leave open the possibility that the trend is driven by year effects rather than age effects. For instance, one may expect large negative year effects as China deepens its privatization reforms and shuts down the inefficient state-owned enterprises. The year effects, which are cohort-neutral, could lead to decreasing life-cycle MRPK dispersion within every cohort. In this scenario, year effects lead to a spurious relationship between MRPK dispersion and firm-cohort age for all the cohorts.

The goal of this section is to estimate flexible versions of the MRPK dispersion profile with firm age. The specifications take the following form:

$$\sigma_{mrpk,stj} = \alpha_0 + \sum_{j \in J} \phi_j D_j + \chi_c + \psi_t + \theta_s + \epsilon_{stj}. \tag{5}$$

D_j is a dummy equal to 1 if firms in the industry-year-age bin stj are observed at age j. ψ_t captures year fixed effects, χ_c captures cohort fixed effects, θ_s captures industry fixed effects, and ϵ_{stj} is a mean-zero error term.

3.1 Three Approaches to Identifying Age Effects

The main challenge to estimat the age effects on MRPK dispersion is that age indicators are correlated with cohort indicators and year indicators. Therefore, identifying the age effects separately without any additional assumption is impossible. This section uses three approaches to identify age effects while controlling for cohort and year effects.

To resolve the difficulty of collinearity, I follow Deaton (1997) and imposes one additional linear restriction on the set of cohort and time effects to estimate equation (5). Consider the decrease in aggregate MRPK dispersion over time, as plotted in Figure A.5, which reflects the combined result of cohort-neutral year effects and effects of changes in the composition of firm cohorts in a calendar year. To identify age effects, I need to discipline the relative role of year effects and cohort effects in the decrease in aggregate MRPK dispersion over time.

Preferred Approach. My preferred identification approach assumes a linear trend in age effects on MRPK dispersion after age 10. For example, consider the assumption of no trend in the age effects after firm-cohort age 10 as a special case of the linear effects. Then year effects can be identified by following the firm cohorts older than age 10 because all the changes over time are only due to year effects in the absence of cohort and age effects. This assumption is actually also in accord with the empirical findings of Haltiwanger, Jarmin, and Miranda (2013): Mature firms in the US have stable dynamics compared to the younger firms.

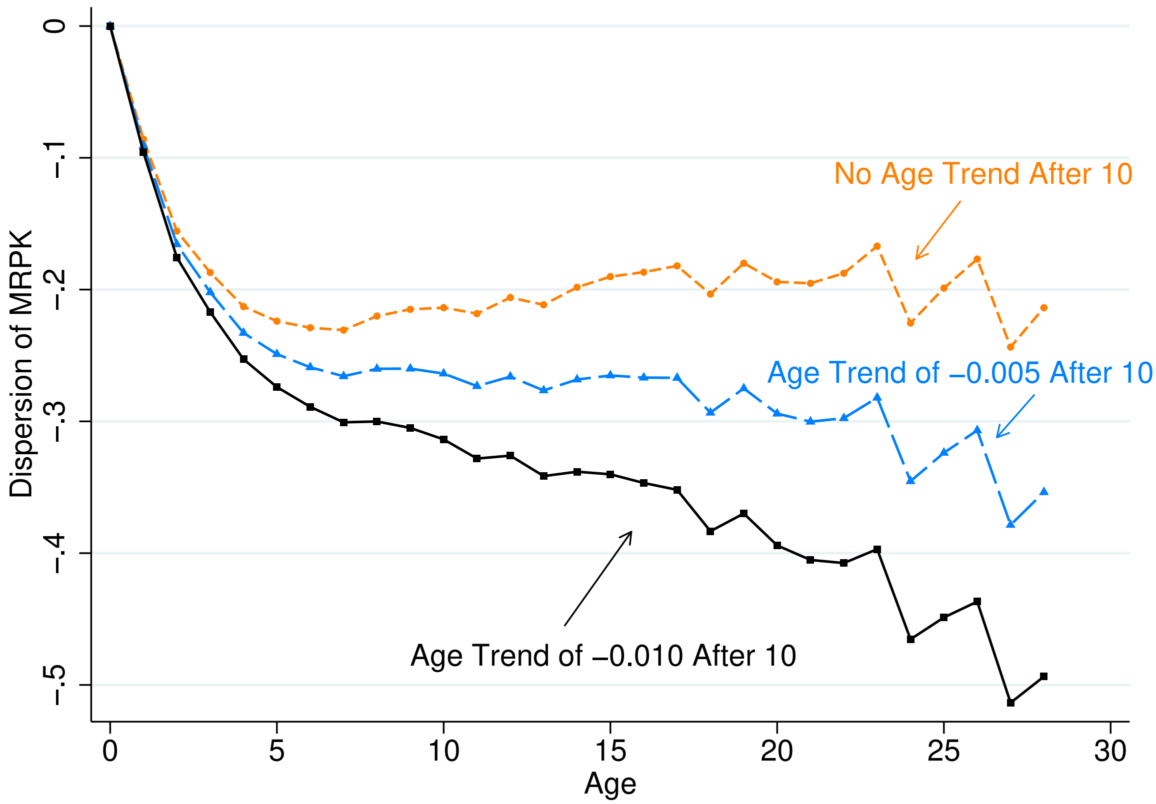
Furthermore, I can also test this assumption, because the second derivatives of age effects, which inform the curvature of age effects, are always identified as shown in McKenzie (2006). I find the age effects are convex for the young firms, meaning MRPK dispersion decreases at a decreasing rate with firm cohort age. In addition, I cannot reject the null hypothesis that the age effects on MRPK dispersion are linear after age four. I describe the econometric details in section A.1. The test results provide econometric foundations for identifying first-order age effects by assuming a linear trend in the age effects at older ages.

Figure 4 plots the estimated profile of MRPK dispersion by firm-cohort age using the second identification approach. In particular, I impose three different plausible magnitudes of the linear trend in the age effects on MRPK dispersion after firm-cohort age 10: (a) no trend in the age effects after age 10 (dashed orange line with circle markers); (b) a small decreasing trend of 0.005 points per age after age 10 (long-dashed blue lines with triangle markers); (c) a moderate decreasing trend of 0.010 points per age after age 10 (solid black lines with square markers). The three different trends imposed on age effects after firm-cohort age 10 all yield a substantial and convex decline of dispersion in MRPK for the young firms. In particular, MRPK dispersion decreases 0.22 to 0.25 point before age five.

Alternative Approach One. Instead of picking a plausible magnitude of the trend in age effects, the alternative identification approach estimate it by imposing the assumption that two consecutive old cohorts are the same, as in Hall (1971). In the context of my sample between 1998 and 2007 in China, this assumption is based on the background that old firm cohorts founded in the late 1970s are similar because they were founded at the beginning of the economic reform and were adapting gradually, whereas the young firm cohorts could be drastically different because they are founded in different years in the fast-changing economy as China deepened its privatization reforms and largely opened up. The assumption in this approach is also a relaxed constraint of the assumption in Hall (1971) (p.248), who assumed all vintages have the same cohort effects to identify the age effects on the prices of used trucks.

This assumption identifies the slope of the linear trend in age effects by observing the old cohorts in the same years. Consider the two consecutive old firm cohorts founded in 1979 and 1980, both observed in the year 1998: one at age 18 and the other at age 19. Because they are observed in the same year, there is no difference in the year effects. The cohorts

Figure 4: Estimated MRPK Dispersion by Firm Age, Preferred Approach



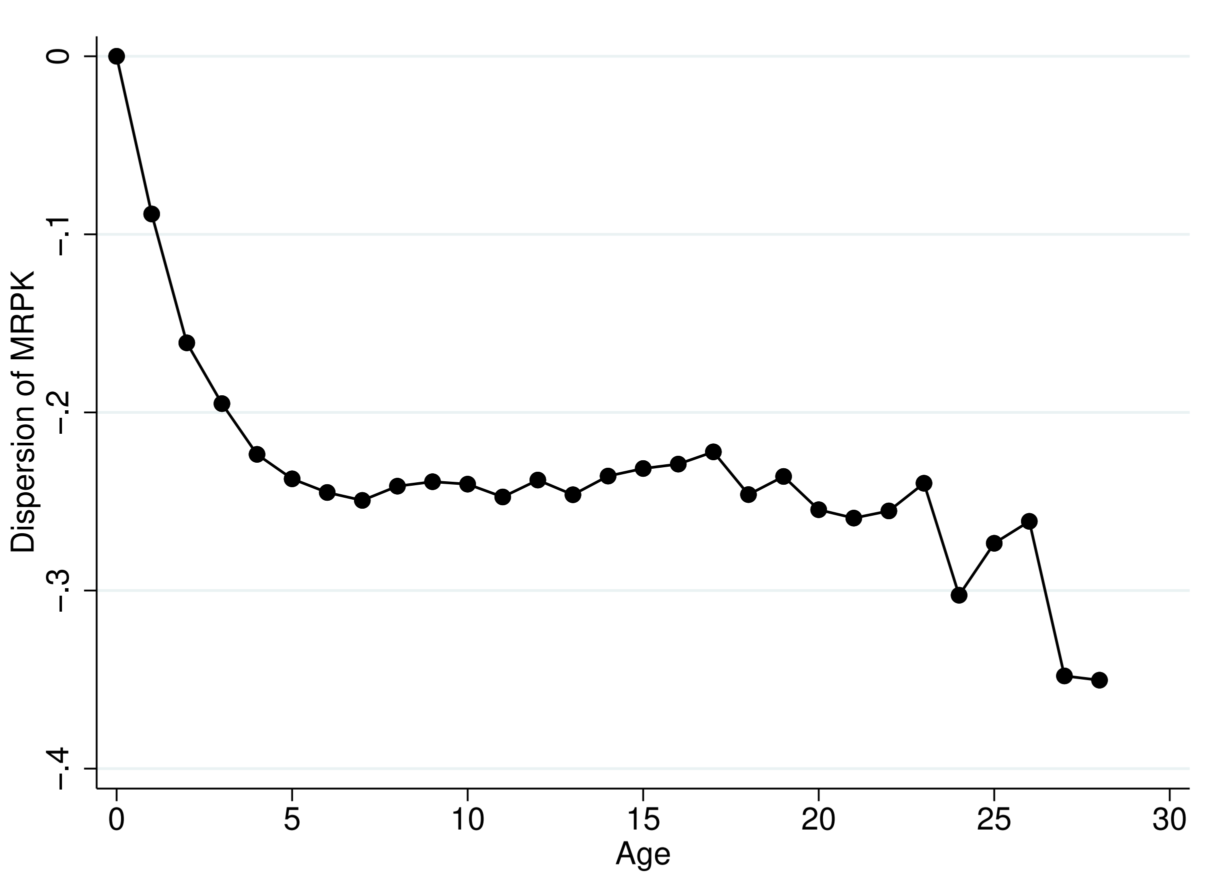
Note: This figure plots the estimated profile of MRPK dispersion by firm-cohort age in equation (6) using the second identification approach, which assumes (a) no trend in the age effects on MRPK dispersion after age 10 (dashed orange line with circle markers); (b) a small decreasing trend of 0.005 points per age after age 10 (long-dashed blue lines with triangle markers); (c) a moderate decreasing trend of 0.01 points per age after firm age 10 (solid black lines with square markers).

effects are the same as well under the assumption; hence, the difference in MRPK dispersion is only due to the different age effects at age 18 and age 19. In total, they are observed for ages 19-28 and 18-27, respectively, during 1998-2007. The average difference across years then gives the least-squares estimate of the age effect per year. In addition, by following all firm cohorts over time, this assumption can now help identify the trend in year effects given the age trend.

I implement this approach by estimating equation (5) in the framework as described in section A.3. In practice, I assume every two adjacent cohorts founded during 1979 to 1983 are the same. They are observed from age 15 to 28. For each of the four pairs of adjacent cohorts, I calculate the difference in MRPK dispersion in each year between 1998 and 2007. Then I take the average of the differences of the four pairs as the trend in age effects after age 15, which turns out to be .009. I force the cohort effects of firms born in 1979 - 1983 to be the same.

Figure 5 plots the estimated profile of MRPK dispersion by firm-cohort age using this approach: MRPK dispersion decreases substantially through the first five years of a firm cohort's life cycle, accumulating to more than -0.24 points. It further decreases after age five, though at a slower rate, accumulating to -0.35 points at age 28 compared to age zero.

Figure 5: Estimated MRPK Dispersion by Firm Age, Alternative Approach One



Note: This figure plots the estimated profile of MRPK dispersion by firm-chort age when assuming the two adjacent firm cohorts founded in 1979 - 1983 have the same cohort effects.

Alternative Approach Two. This approach makes econometric assumptions to split the decreasing trend of dispersion over time between time effects and cohort effects as in Deaton (1997), and does not make assumptions on the curvature of age effects. This approach also illustrates the econometric difficulty in disentangling the three effects.[8]

[8]This methodology is commonly used in the literature on individuals' life-cycle consumption and income dynamics (e.g., Lagakos, Moll, Porzio, Qian, and Schoellman 2018), and was recently used for firms in Moreira (2017) and Argente, Lee, and Moreira (2018).

In practice, I implement two ways to split the decline in aggregate MRPK dispersion over time: One version attributes all the decline to cohort effects, and the other version attributes all the decline to cohort-neutral year effects. I show in Appendix section A.2 that the two restrictions provide the lower and upper bounds of age effects if all three effects of age, year, and cohort on MRPK dispersion have non-positive trends. The condition of a non-negative trend in all three effects is a plausible case because the patterns of MRPK dispersion decrease with firm age, calendar year, and the birthyear of a firm cohort, as shown in section 2.2.

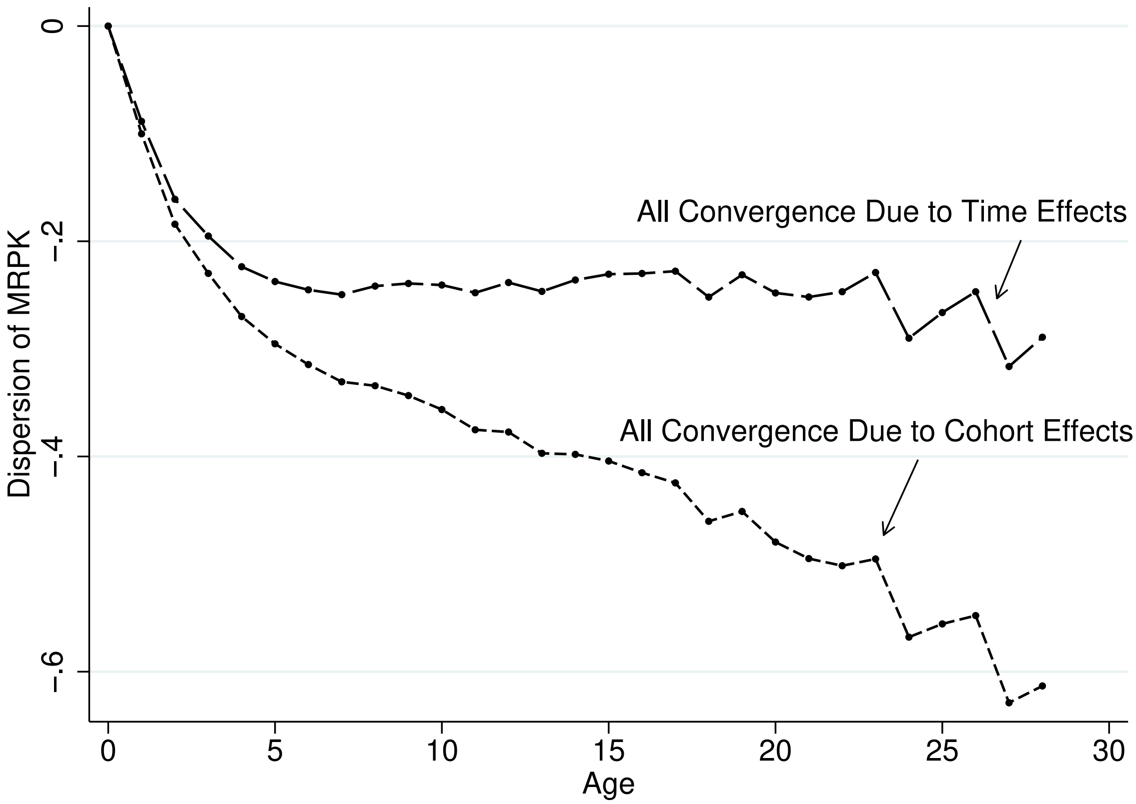
Specifically, I estimate equation (5) under restrictions. The first version attributes all the decline over time to cohort effects. It makes the same assumption as in the original analysis in Deaton (1997) and uses year dummies to capture only cyclical fluctuations. In practice, the first age dummy and the first cohort dummy are omitted as the benchmark reference, and the time dummies are transformed to meet the restriction that the time effects are orthogonal to a time trend. The second version is the opposite extreme case and attributes all decline to time effects. In this version, I assume the cohort effects are orthogonal to a time trend. See Appendix A.3 for a formal description of this approach and the details of implementing it.

The long-dashed line in Figure 6 plots MRPK dispersion with firm-cohort age estimated under the assumption in the first version that all decline is driven by cohort effects. It provides the lower bound of the age effects, if all three effects of age, cohort, and year have non-positive trends, as shown in Appendix section A.2. In this version, MRPK dispersion decreases substantially with firm age, accumulating to a magnitude of 0.6 points at age 28. The dashed line in Figure 6 plots the profile of MRPK dispersion estimated in the second version, where I assume all the decline over time is driven by year effects. In the second version, MRPK dispersion again decreases with firm age; note the decrease is most substantial before age five and flattens out afterward. This version also provides the upper bounds of age effects, as shown in Appendix section A.2.

In conclusion, the first alternative approach shows MRPK dispersion decreases substantially with firm age, both in the lower and upper bounds. In particular, MRPK dispersion within a firm cohort decreases more than 0.04 points per age on average until age five, though the slope of MRPK dispersion at older ages is sensitive to the restrictions used for identification. When I attribute all the decline in MRPK dispersion over time to year effects, the estimated profile of MRPK dispersion closely resembles the results in the preferred approach.

In summary, although identifying the first-order age effects directly without any additional assumptions is impossible, the three identification approaches in this section establish a substantial decrease in MRPK dispersion with firm-cohort age. The estimation result of the three cases in the preferred approach lies between the upper and lower bounds (estimated

Figure 6: Estimated MRPK Dispersion by Firm Age, Alternative Approach One



Note: This figure plots the estimated MRPK dispersion by firm-cohort age using the first approach. The long-dashed line plots MRPK dispersion by firm-cohort age estimated using equation (5), under the assumption that all the decline in MRPK dispersion over time is driven by cohort effects. The dashed line plots MRPK dispersion by firm cohort age estimated using equation (5), under the assumption that all the decline over time is driven by year effects. See Appendix A.3 for a detailed description of implementing this methodology.

in alternative approach two). The result from alternative approach one is also consistent with the upper and lower bounds, with estimates closer to the lower bounds. All three approaches conclude a substantial decrease in MRPK dispersion within the firm cohorts at young ages, accumulating to a magnitude of 0.2 to 0.3 by age five. In addition, the estimated profiles of MRPK dispersion through the first five years of the firm cohort's life cycle are convex, as McKenzie tests predict. The age effects on MRPK dispersion after age five are generally negative across the three approaches, though the magnitude is sensitive to the specific identification assumptions.

3.2 Robustness

This section assesses the robustness of the fact I document that MRPK dispersion decreases substantially with firm-cohort age. In particular, I consider other plausible factors that can affect the profile of life-cycle MRPK dispersion within a firm cohort: exit selection, time-series volatility of productivity, firm ownership, financial frictions, firm size, and measurement errors.

3.2.1 Controlling for the Volatility of Productivity

In this section, I investigate the life-cycle MRPK dispersion after controlling for the volatility of productivity at each age of the firm cohort. Asker, Collard-Wexler, and De Loecker (2015) argue the dispersion in MRPK can largely be explained by capital adjustment costs in an environment with productivity volatility, where firms choose the capital stocks in the current period, taking into consideration that the volatility of productivity in the future, thus resulting in ex-post static MRPK dispersion in the current period. If volatility in productivity decreases as the firm cohort ages and matures, the older firms will tend to have less dispersion in the ex-post MRPK than the younger firms. In this case, MRPK dispersion may decrease over the firms' life cycles due to decreasing productivity volatility with firm-cohort age. Therefore, not controlling for the volatility of productivity could overstate the magnitude of negative age effects.

I define the time-series productivity volatility to be $\sigma_{\Delta z,stj}$, as in Asker, Collard-Wexler, and De Loecker (2015), which measures the standard deviation of productivity changes, $(z_{it} - z_{it-1})$, from one period to the next. The index stj indicates the standard deviation is taken across firms within the same industry-year-age bin. Adding $\sigma_{\Delta z,stj}$ as a control variable in equation (5), I use the second identification approach to estimate

$$\sigma_{mrpk,stj} = \alpha_0 + \alpha_{vol} \cdot \sigma_{\Delta z,stj} + \sum_{j \in J} \phi_j D_j + \theta_s + \chi_c + \psi_t + \epsilon_{stj}.^{9} \tag{6}$$

Consistent with Asker, Collard-Wexler, and De Loecker (2015), I find higher productivity volatility is correlated with a higher level of dispersion in marginal capital products. A one unit increase in the volatility of productivity predicts a 0.32-point increase in the cross-sectional MRPK dispersion, significant at the 1-percent level. The estimated profile of MRPK dispersion, after controlling for volatility of productivity, also decreases with firm

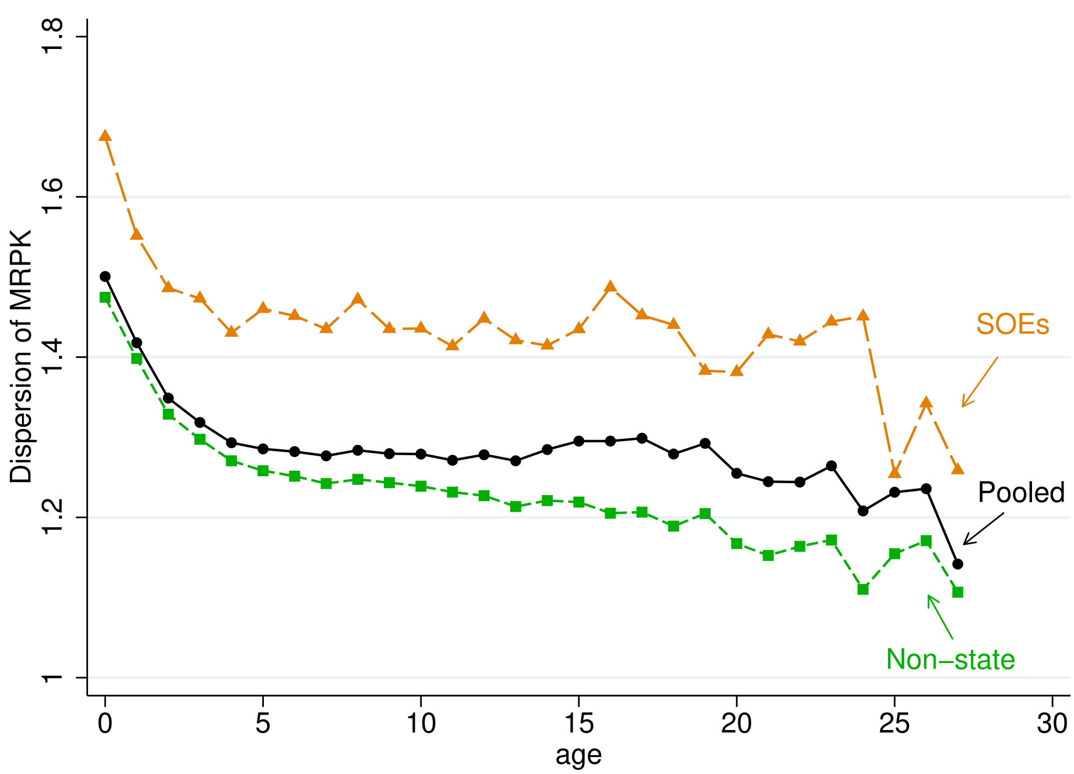
[9] The third approach, which assumes two adjacent old cohorts have the same cohort effects, becomes less straightforward here, because the two cohorts have different volatility of productivity even when observing in the same year.

age. As before, the decrease is most substantial through the first few years of a firm cohort's life cycle, though at a slightly smaller magnitude, accumulating to -0.18 to -0.22 points by age five compared to entry. I conclude that the decrease in MRPK dispersion with firm-cohort age cannot be explained by declining volatility of productivity as a firm cohort ages. See the estimation results plotted in Appendix Figure A.9.

3.2.2 State-owned and Non-state Firms

The misallocation of capital between the state-owned and the non-state-owned firms is a salient feature in the Chinese economy (see, e.g., Bai, Jin, and Lu, 2018; Brandt, Kambourov, and Storesletten, 2016; Brandt, Tombe, and Zhu, 2013). One may worry that the life-cycle production of state-owned enterprises (SOEs) in China responds largely to government policies and do not reflect the market outcomes. The SOEs might drive the pattern of MRPK dispersion over the firm cohort's life cycle. This section reports the life-cycle MRPK dispersion by firm ownership.

Figure 7: MRPK Dispersion by Firm Ownership



Note: This figure plots the weighted average MRPK dispersion for the SOEs, the non-state firms, and the pooled sample.

I define the firm as a SOE if more than half of its assets is owned by the state, and define the firm as a non-state firm otherwise. Figure 7 plots the weighted average MRPK dispersion for the SOEs, the non-state firms, and the pooled aggregate sample. The life-cycle MRPK dispersion of the non-state firms closely resembles that of the pooled sample, which decreases from 1.5 to around 1.2 between entry and age 27. Dispersion in MRPK across SOEs within a cohort also decreases with firm age. In addition, it constantly remains at a higher level than the dispersion among non-state firms, which may reflect larger informational frictions or less learning among SOEs. I conclude that MRPK dispersion robustly decreases with firm-cohort age, both for the SOEs and the non-state firms.

Table 2: Dispersion of MRPK by Firm Age

$\bar{\sigma}_j$	Full Sample	Non-state Firms
Dispersion at Age 1	-.04*** (.01)	-.03** (.01)
Dispersion at Age 5	-.18*** (.01)	-.19*** (.01)
Dispersion at Age 10	-.23*** (.01)	-.25*** (.01)
Dispersion at Age 20	-.30*** (.02)	-.37*** (.02)
Dispersion at Age 27	-.41*** (.04)	-.44*** (.04)

Note: This table reports $\bar{\sigma}_j$, the average standard deviation of MRPK at a given age, in the data compared to entry with the estimate of standard error in parentheses. Row 1 uses the full sample, and Row 2 uses only the non-state firms.

In addition, Table 2 reports the difference in the dispersion of MRPK at an older age relative to entry, for the full sample and for only the non-state firms. The t-test results of the equal means show that all the differences are strongly significant. For the full sample, the dispersion in MRPK decreases by 0.4 points until age 27. This decline through entry to age 27 is slightly larger for the non-state firms. Both for the full sample and for the non-state firms, MRPK dispersion decreases substantially before age five. In particular, it drops by almost 0.2 points until age five compared to entry, which accounts for around half of the decrease in MRPK dispersion during firm entry to age 27.

To further identify age effects, I estimate the dispersion in MRPK with firm-cohort age after restricting the sample to only non-state firms, using the second identification approach and controlling for volatility of productivity. It yields the same coefficient on volatility (0.32), significant at 1%, as the full sample. Dispersion in MRPK across non-state firms decreases 0.20 to 0.25 points by age five compared to age zero, which has a slightly larger magnitude than the estimates using the pooled sample. See the estimation results plotted in Appendix Figure A.10.

3.2.3 Financial Frictions

An alternative explanation for the pattern I document is financial frictions, which could generate MRPK dispersion if they were high for some firms and low for others. Then they could gradually go away for various reasons, such as internally generated funds, or learning by banks. If young firms overcome financial constraints over time, they will start with high marginal product, and then decrease it.

However, financial frictions cannot explain why some firms start out with low marginal product and then raise it. In the data, 56% of the firms have higher MRPK than the previous year. In particular, 67% of the survived firm entrants have higher MRPK at age 1, and 59% of the survived age-one firms have higher MRPK at age 2; this percentage fluctuates between 54% to 56% from age 3. Because financial frictions cannot reconcile MRPK dynamics over time of these many firms in the data, I conclude financial constraints are not the driving force of the decline in MRPK dispersion with firm age.

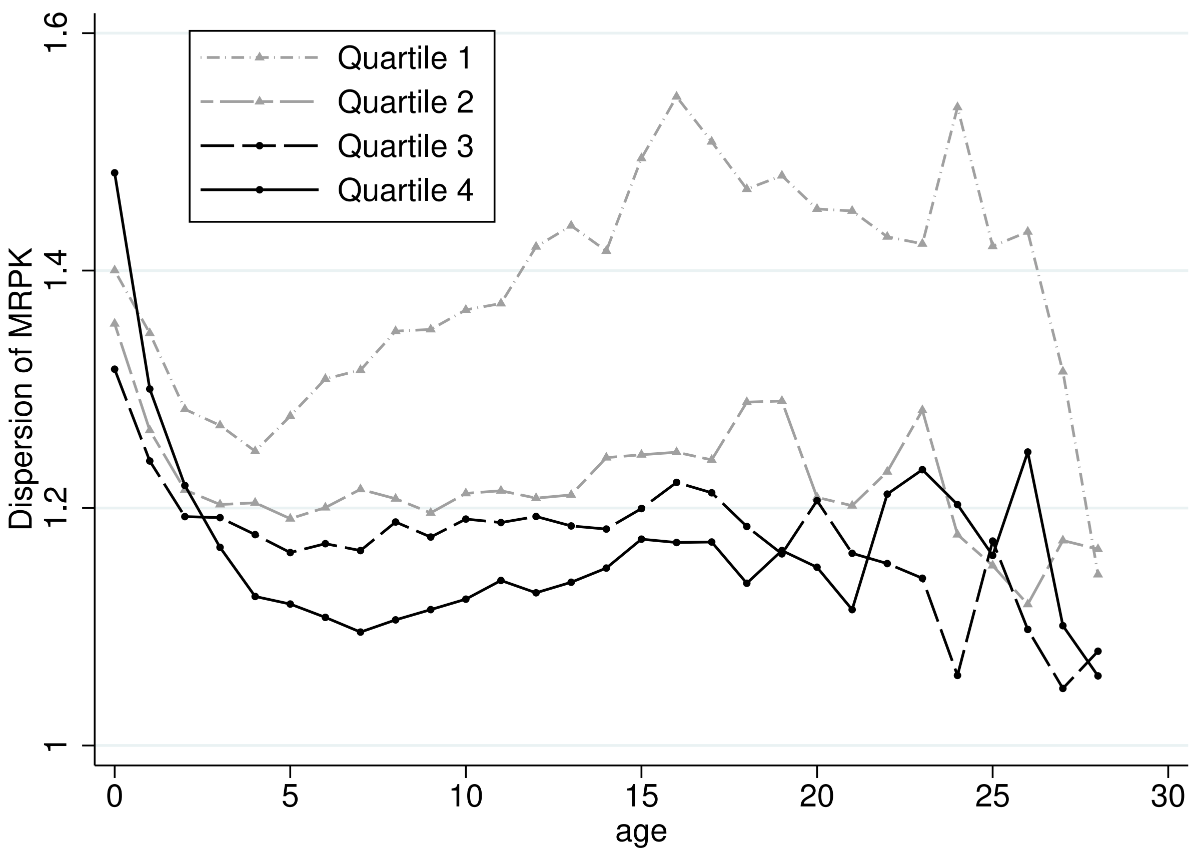
3.2.4 Firm Size and Measurement Error

Because average firm size and firm-cohort age are strongly and positively correlated, one may worry about whether the fact that I document is robust across firm groups with different average sizes. This section assesses the robustness of the decreasing MRPK dispersion with firm-cohort age to firm size.

Figure 8 plots the weighted average MRPK dispersion by firm size. The bottom-quartile firms have around 45 employees on average, and the top quartile firms on average have more than 166 employees. For firms in quartile two and three, and the top-quartile of firm-size distribution, MRPK dispersion decreases substantially with firm age, particularly for young firms. For firms in the bottom quartile, MRPK dispersion within a firm cohort decreases through the first five years but increases afterward. I conclude the age effects, particularly before age five, are robust to firm size. This is also consistent with the finding in Haltiwanger, Jarmin, and Miranda (2013) that effects of firm size become insignificant after controlling for firm age.

Measurement error has been an important and challenging concern for the misallocation lit-

Figure 8: MRPK Dispersion by Firm Size



Note: This figure plots the weighted average MRPK dispersion by firm-size quartile.

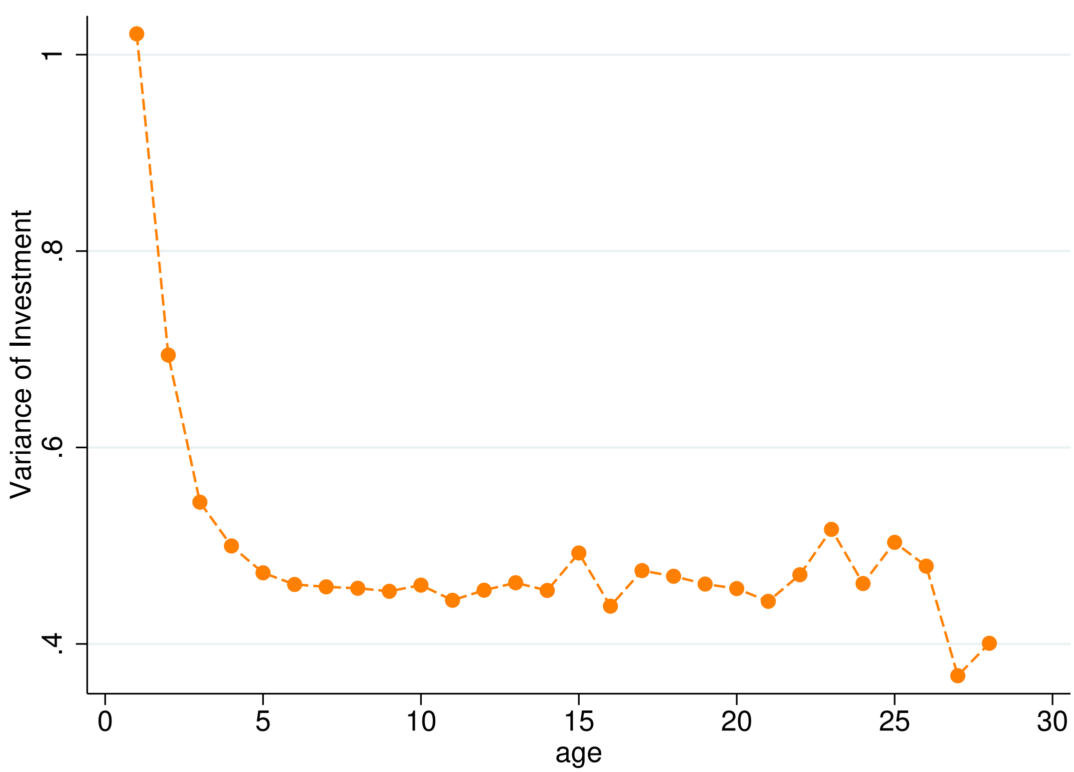
erature and, more broadly, for measuring capital stocks and revenue outputs using firm-level data.[10] For this paper in particular, one may worry that measurement errors are larger for young and small firms because they lack the resources and experiences to report measurements precisely. If one thinks of larger firms as those that are competent in corporate finance and accounting, and thus have relatively small measurement errors in reported revenue output and capital stocks, then there is less concern about measurement errors when we look at MRPK dispersion within large firms. The black lines in Figure 8 show that MRPK dispersion decreases robustly with firm-cohort age for the third-quartile and top-quartile firms, which have an average of 120 and 520 employees, respectively. That MRPK dispersion across large firms decreases over the firm cohort's life cycle provide indirect evidence of the pattern's robustness to measurement error.

[10] Rotemberg and White (2017) argue that the editing strategies used for U.S. Census of Manufactures may largely decrease the measured MRPK dispersion in the cleaned dataset, leading to lower MRPK dispersion in the US than in India. However, because I focus on the firm panel-data within one country, differential data editing strategies across countries is much less concerning.

In addition, I follow the approach in Bils, Klenow, and Ruane (2017) to estimate how much additive measurement errors in the revenue output and capital input for firms at each age accounts for observed MRPK dispersion (σ^2_{mrpk}). This approach essentially involves estimating the following regression: $\Delta log(y_{it}) = \phi_{err} mrpk_{it} + \Psi_{err}\Delta log(k_{it}) - \Psi_{err}(1-\lambda_{err}) mrpk_{it} \cdot \Delta log(k_{it}) + D_{st} + \epsilon_{it}$, where $\Delta log(y_{it})$ and $\Delta log(k_{it})$ denote changes in log revenue output and capital, and D_{st} denotes the industry-year fixed effects. They show that $(1-\lambda_{err})$ represents the contribution of measurement error to observed variance in MRPK under certain assumptions. The estimates of $1-\lambda_{err}$, using samples restricted to firm cohorts at age one to 28, respectively, have an average value of 0.02, which suggests measurement errors contribute to only 2% of the observed MRPK dispersion on average. Regressing the estimated $1-\hat{\lambda}_{err}$ on age j, one will find a positive and insignificant coefficient of 0.15 with a P-value of 0.19, thus suggesting that the additive measurement error does not contribute differently to MRPK dispersion within firm cohorts at different ages.

3.3 Empirical Evidence of Firm Life-Cycle Learning

Figure 9: Variance of Investment by Firm Age



Note: This figure plots the weighted average variance of investments by firm age.

The decline in the variance of firm growth with firm age is evidence of firm life-cycle learning (Evans, 1987). Corresponding to the focus of this paper on MRPK, I use capital investment, the difference in the capital between two consecutive years, as the measure of firm growth. Figure 9 then plots the weighted average variance of investments across industry-year-age bins by firm age. It shows that investment dispersion decreases substantially with firm age, particularly for young firms. This is consistent with the basic Bayesian learning mechanism: firms enter with imprecise beliefs about their true productivity and they learn over time by observing revenue output realizations. Therefore, young firms face larger uncertainty about their productivity and revise their beliefs and investments relatively more, compared to the older firms, who are better informed with more observations. Furthermore, the pattern of decreasing variance of investment with firm age highly resembles that of MRPK dispersion, which implies the decline in MPRK dispersion with firm age is likely to be associated with learning. [11]

4 Model of Firm Life-Cycle Learning

In this section, I develop a general equilibrium model to match the convexly downward sloping profile of MRPK dispersion with firm-cohort age in the data. The model features firm life-cycle learning and endogenous exit as in Jovanovic (1982). By adding capital to his original model, I bring in capital adjustment costs and capital fire-sales upon exiting the market, which are important to match the fact that firms scale down their capital stocks prior to exiting the market. Furthermore, the model with capital input choices in this paper generates losses in aggregate productivity due to informational frictions and capital market frictions over the firms' life cycles.

To conduct quantitative analyses, I build the model with multiple market frictions and firm-level distortions that can contribute to MRPK dispersion. Firms choose inputs facing (i) informational frictions, in the form of imperfect signals about their own fundamental productivity as well as contemporaneous uncertainty due to idiosyncratic shocks in each period, (ii) exit frictions, in the form of discount value from capital fire-sale on exit, (iii) technological frictions, in the form of quadratic capital adjustment costs, and (iv) a generic class of idiosyncratic firm-level distortions as in Hsieh and Klenow (2009). The key mechanism is that as firms learn over time, those with too much or too little capital stock adjust and the less productive firms within a cohort exit over time, leading to decreases in MRPK dispersion over a firm cohort's life cycle.

[11] Using firm-level panel data from Japan, Chen, Senga, Sun, and Zhang (2018) show that older firms make less forecasting errors about their idiosyncratic demand than younger firms, which is more direct evidence of firm learning.

4.1 Environment and Equilibrium

Consider a discrete-time, infinite-horizon economy, populated by a representative household. The household inelastically supplies a fixed amount of labor N and has a preference over consuming the final good. The household discounts time at rate β. I deliberately keep the household side of the economy simple because of its limited role in the analysis.

Distribution of fundamentals. The distribution of firm fundamental productivity x_i is log-normally distributed, that is, $x_i \sim N(\mu_x, \sigma_x^2)$. In each period, with probability $\lambda \in (0,1)$, a firm i carries over the same fundamental to the next period, and with probability $1-\lambda$, the firm exits exogenously.

Production. At the beginning of every period, each firm draws a productivity z_{it}, which combines its fundamental and an idiosyncratic transitory shock $e_{it} \sim N(0, \sigma_e^2)$. That is, $z_{it} = x_i + e_{it}$. I assume the standard Cobb-Douglas production function, where output $y_{it}(k_{it}, n_{it}; z_{it}) = e^{z_{it}} k_{it}^{\alpha_1} n_{it}^{\alpha_2}$ with $\alpha_1 + \alpha_2 < 1$. Recall that k_{it} denotes capital input and n_{it} labor input. Note the assumption of decreasing return to scale is equivalent to an alternative environment in which firms produce differentiated products and face downward-sloping demand curves due to decreasing marginal utility of consumption. In that setup, z_{it} can be interpreted as an idiosyncratic demand shifter. In this paper, I will refer to z_{it} as the productivity specific to firm i at time t.

Learning. Firms learn about their own fundamental productivity by experimenting and observing realized outputs in the previous periods, as in Jovanovic (1982) and Kerr, Nanda, and Rhodes-Kropf (2014). The firms' beliefs about their fundamentals are summarized in expected mean $\hat{x}_{it}$ and expected variance $\hat{\sigma}_{it}^2$. At the beginning of the firm-entry period, where $t = 0$ and no realizations of productivity z_{it} arrive yet, firms have a common prior belief about their fundamental technology as $N(\hat{x}_{i0}, \hat{\sigma}_{i0}^2) \equiv N(\mu_x, \sigma_x^2)$. In every period t, they use the noisy signal z_{it} to update and form a posterior belief about their fundamental productivity x_i as $N(\hat{x}_{i,t+1}, \hat{\sigma}_{i,t+1}^2)$. Bayesian updating is based on the following equations:

$$\hat{x}_{i,t+1} = \frac{\hat{\sigma}_{it}^2 z_{it} + \sigma_e^2 \hat{x}_{it}}{\hat{\sigma}_{it}^2 + \sigma_e^2} \tag{7}$$

$$\hat{\sigma}_{i,t+1}^2 = \frac{\hat{\sigma}_{it}^2 \sigma_e^2}{\hat{\sigma}_{it}^2 + \sigma_e^2}. \tag{8}$$

Fixed and Input Costs. Firms pay a fixed operation cost f_o in every period they produce. Labor is hired period by period in a spot market with the competitive wage w. With capital depreciation rate δ and quadratic adjustment costs parameter ξ, the total cost of choosing capital stock $k_{i,t+1}$ is given by

$$\Phi(k_{it}, k_{i,t+1}) = k_{i,t+1} - (1-\delta)k_{it} + \frac{\xi}{2}\left(\frac{k_{i,t+1}}{k_{it}} - (1-\delta)\right)^2 k_{it}. \tag{9}$$

I also consider other factors that affect capital stock choices in addition to the fundamental productivity or demand. These factors include, for example, government policies, such as size-dependent taxes, or institutional environment, such as legal forms. As in Hsieh and Klenow (2009) and David and Venkateswaran (2017), to capture these factors, I introduce a class of idiosyncratic firm-level "distortions" that appear in the firm's problem as proportional taxes on capital. I leave out the wedges on hiring decisions for simplicity. I allow distortions on capital to covary with contemporaneous productivity, that is, taxes $T^k_{it} = e^{z_{it}\tau_k}$, where τ_k denotes the correlation that determines the extent to which the capital price comoves with the contemporaneous productivity. If τ_k is positive, distortions discourage the investment of firms with stronger fundamentals while protecting those with weaker fundamentals, which is arguably the empirically relevant case (Bento and Restuccia, 2017; Hsieh and Olken, 2014). The opposite incentive is true if τ_k is negative.

Firm's problem. At the beginning of each period t, firms choose whether to exit permanently or continue operating the business, and choose capital stocks $k_{i,t+1}$ if they continue operating. When exiting the market, firms turn to fire sales for their capital stocks and retain discounted values of γk_{it}, as in Ramey and Shapiro (2001). A firm's state variables, or information set, includes the capital stock k_{it}, the observation of a noisy signal in productivity z_{it}, and the belief about the their fundamentals, summarized in $\hat{x}_{it}$ and $\hat{\sigma}^2_{it}$. Because $\hat{\sigma}^2_{it}$ has a deterministic path over firm age j, I make the j an explicit state variable instead of $\hat{\sigma}^2_{it}$. Therefore, the value of an incumbent firm at age j is given by $V(k_{it}, z_{it}, \hat{x}_{it}, j) = \max_{D\in\{0,1\}}\big\{V^E(k_{it}, z_{it}, \hat{x}_{it}, j), V^C(k_{it}, z_{it}, \hat{x}_{it}, j)\big\}$, where the dummy variable D denotes the exit choice, $V^C(k_{it}, z_{it}, \hat{x}_{it}, j)$ denotes the value of continuing operation, and $V^E(k_{it}, z_{it}, \hat{x}_{it}, j) = \gamma k_{it}$ is the value of exit. Writing the value of continuation in the recursive form yields

$$\begin{aligned} V^C(k_{it}, z_{it}, \hat{x}_{it}, j) = \max_{k_{i,t+1}, n_{it}} \{& e^{z_{it}} k_{it}^{\alpha_1} n_{it}^{\alpha_2} - w n_{it} - T^k_{it}\Phi(k_{it}, k_{i,t+1}) - f_o + \\ & \beta\big((1-\lambda)\mathbb{E}V(k_{i,t+1}, z_{i,t+1}, \hat{x}_{i,t+1}, j+1) + \lambda V^E(k_{i,t+1}, z_{i,t+1}, \hat{x}_{i,t+1}, j+1)\big)\big\}, \end{aligned}$$

where $\mathbb{E}$ denotes the firm's expectation of the value in $t+1$ conditional on the current information set $\{k_{it}, z_{it}, \hat{x}_{it}, j\}$. Maximizing over the choice of labor inputs yields $n_{it}(z_{it}, k_{it}) = \big(\alpha_2 \frac{e^{z_{it}} k_{it}^{\alpha_1}}{w}\big)^{\frac{1}{1-\alpha_2}}$. After substituting the optimal choice of labor inputs, the value of continuation becomes

$$V^C(k_{it}, z_{it}, \hat{x}_{it}, j) = \max_{k_{i,t+1}}\{GAk_{it}^{\alpha} - T_k\Gamma(k_{it}, k_{i,t+1}) - f_o + \beta\lambda\gamma k_{i,t+1} + \beta(1-\lambda)V(k_{i,t+1}, z_{i,t+1}, \hat{x}_{i,t+1}, j+1)\}, \tag{10}$$

where $G \equiv (1-\alpha_2)\left(\frac{\alpha_2}{w}\right)^{\frac{\alpha_2}{1-\alpha_2}}$, $A = e^{z\frac{1}{1-\alpha_2}}$, and $\alpha \equiv \frac{\alpha_1}{1-\alpha_2}$ is the curvature of revenues net of wages.

Stationary equilibrium. We can now define a stationary equilibrium as follows: (i) a wage w; (ii) a set of value and policy functions of the firm: $V(k_{it}, z_{it}, \hat{x}_{it}, j)$, $D(k_{it}, z_{it}, \hat{x}_{it}, j)$, and $k_{i,t+1}(k_{it}, z_{it}, \hat{x}_{it}, j)$; and (iii) a joint distribution of $\Omega(k_{it}, z_{it}, \hat{x}_{it}, j)$ such that (a) taking wages and the law of motion for information set as given, the value and policy functions solve the firm's optimization problem, (b) the labor market clears as labor demand equals labor supply: $\int n_{it}(z_{it}, k_{it})d\Omega(k_{it}, z_{it}, \hat{x}_{it}, j) = N$, and (c) the joint distribution is the fixed point through time.

4.2 Intuitions of the Firm's Problem

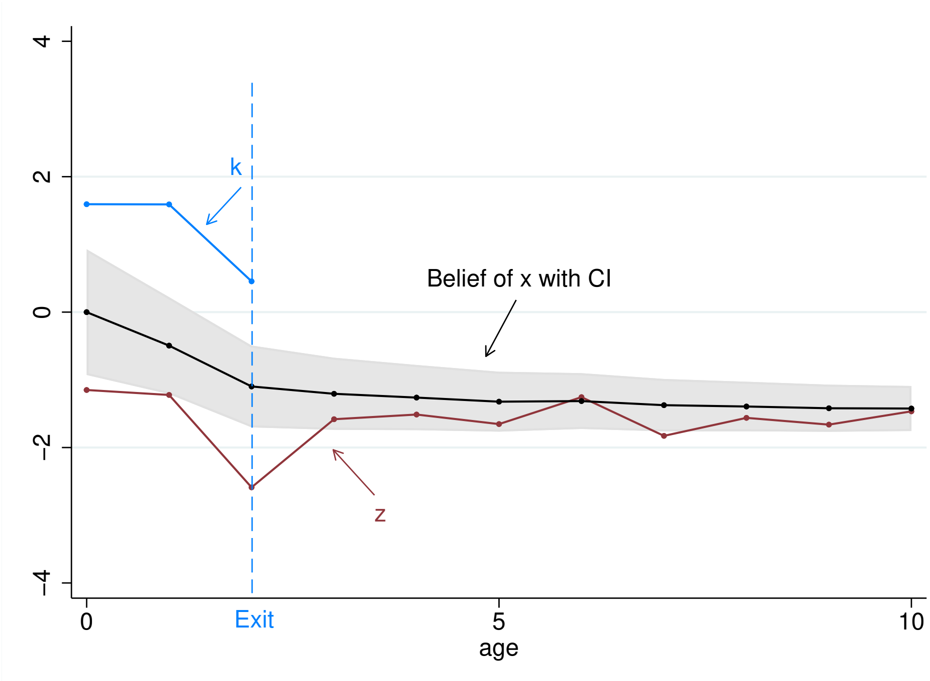
Intuitively, without distortions (i.e., τ_k=0), the choice of the next period's capital $k_{i,t+1}$ should be weakly increasing in the three state variables k_{it}, z_{it}, and $\hat{x}_{it}$ at any age j. However, sufficiently large distortions, which disincentivize investment of more productive firms, may lead to less investment of more productive firms. Although $k_{i,t+1}$ is always weakly increasing in k_{it} given the other state variables, it is not necessarily increasing $\hat{x}_{it}$ given the other state variables. In section 5, I discuss the relevant case of distortions and investment decisions.

Figure 10 plots two examples of one firm's state variables over the firm's life cycle, using the same parameter values as in section 5. The left panel plots a firm with a low fundamental, in which case the firm chooses to downsize its capital stock in the next period after updating its belief of x_i from zero to negative at age two. When a large negative shock arrives at age three, this firm chooses to exit. The right panel plots a firm with a high fundamental, in which case the firm updates its belief of x_i upward smoothly and accumulates the capital stably through the first 10 years of its life cycle. This figure shows that less productive firms endogenously exit the market over time, whereas more productive firms stay and grow larger. Because initial capital stocks at entry may not match the firms' fundamentals for various reasons, including imprecise priors and large shocks, MRPK dispersion is large within the firm cohort at entry. This dispersion decreases over time as firms learn over time, adjust their capital stocks, and some firms exit the market.

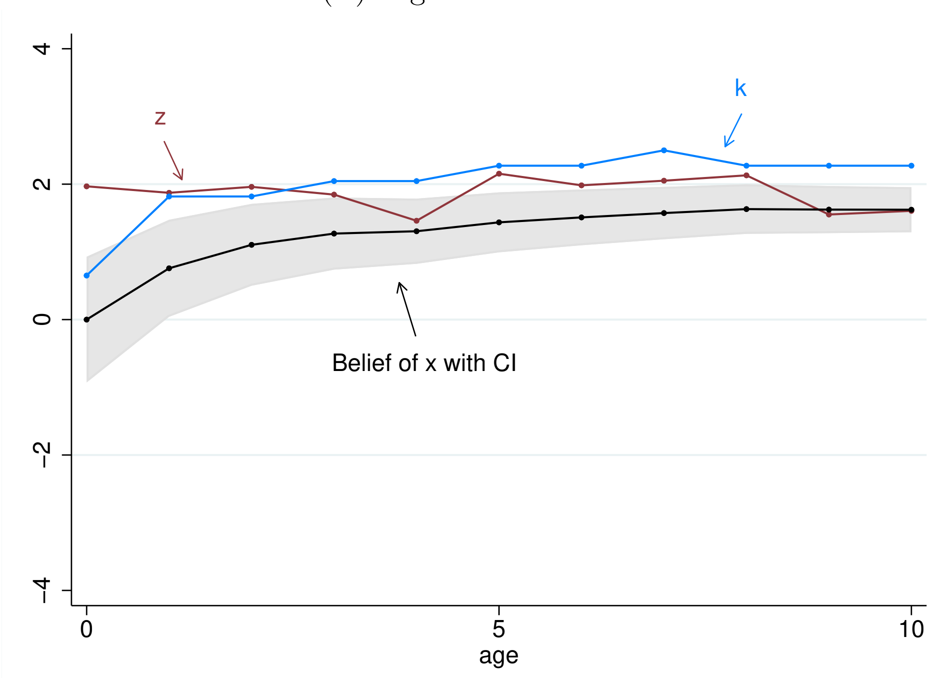
Now I turn to a formal expression of computing the effects of MRPK dispersion on aggregate

Figure 10: Examples of One Firm's Life Cycle in the Model

(A) Low Fundamental



(B) High Fundamental



Note: This figure plots two examples of a firm's state variables over the firm's life cycle. The maroon line plots realizations of productivity z_{it} over time; the black line plots corresponding beliefs of the fundamental $\hat{x}_{i,t+1}$ with a 95% confidence interval based on $\hat{\sigma}^2_{i,t+1}$; and the blue line plots capital stock k_{it}.

productivity. As shown in Appendix A.4, combining the firm's optimal labor choice with the labor market and capital market clearing condition gives the expression of aggregate productivity as

$$z = z^* - \frac{1}{2}\frac{\alpha_1(1-\alpha_2)}{(1-\alpha_1-\alpha_2)}\sigma^2_{mrpk}, \tag{11}$$

where z^* is the TFP in the frictionless and undistorted economy without any dispersion in MRPK, and σ_{mrpk} is the aggregate standard deviation of MRPK. Taking the partial derivative of equation (11) reveals the relationship between MRPK dispersion (σ^2_{mrpk}) and productivity losses ($z - z^*$):

$$\frac{dz}{d\sigma^2_{mrpk}} = -\frac{1}{2}\frac{\alpha_1(1-\alpha_2)}{(1-\alpha_1-\alpha_2)}.$$

This expression provides a natural way to quantify the effects of changes in σ^2_{mrpk} on aggregate productivity. In Section 5.2, I use this strategy to decompose the quantitative contribution of each factor to MRPK dispersion and TFP losses.

5 Quantitative Analysis

Throughout the analysis, I focus on dynamics of MRPK dispersion over the firm cohort's life cycle. Consider an economy with exogenous firm entry: Every period, the firm cohort with a joint distribution over $\{k_{i0}, x_i, e_{i0}\}$ enters. The model can predict the firm cohort's distribution over its life cycle by solving the firms' optimization problems. The joint distribution over $\{\hat{x}_{it}, z_{it}, k_{it}\}$ is then fixed over time for any given firm-cohort age j. Therefore, stationary equilibrium must exist given the distribution of the firm cohort at entry, as long as exogenous exit rate λ is positive.

5.1 Parameterization

I begin by directly assigning parameter values in the production function based on aggregate moments in the Chinese economy. I set the capital share α_1 to 0.28, which is the weighted average capital share in the manufacturing sector, and set the labor share to 0.53 as in the Annual National Accounts. These two numbers lead to decreasing returns to scale as $\alpha_1 + \alpha_2 = 0.82$, which is in line with the standard value in the literature. The discount factor is set to 0.97 based on an interest rate of 10-year China government bonds of 3% during my sample period. I set the discount rate of capital fire-sale upon exit to 0.5, as used in Ramey and Shapiro (2001). I set the depreciation rate to 0.1, which is close to the median ratio of reported current-year depreciation value to capital stock. I use an exogenous firm exit rate of 0.04, which is close to the average exit rate of old firms in the US. I normalize μ_x, the mean of the firm cohort's fundamentals at entry, to be zero, which is also the common initial belief of expected fundamentals. Regarding the productivity process in the model with time-invariant fundamental x_i, the dispersion of fundamentals and transitory shocks, σ_x^2 and σ_e^2, are exactly identified by the TFPR variance of the entrants, $Var(z_{it}|j=0)$, and the variance of time-series TFPR changes, $Var(z_{i,t+1} - z_{it})$.

Treating entry as exogenous in the model, I take directly the joint distribution of capital stocks and TFPR of all 60,972 age-zero firms in my sample as the initial distribution of $\{k_{i0}, z_{i0}\}$ among the firm cohort. I back out the fundamentals $x_i = z_{i0} - e_{i0}$, using randomly generated $e_{i0} \sim N(0, \sigma_e^2)$. Now, given the initial joint distribution of $\{k_{i0}, x_{i0}, e_{i0}\}$, a unique stationary equilibrium always exists.

I calibrate the remaining three parameters to jointly match three key moments in the data. The three parameters are the correlated distortion τ_k, the fixed operating cost every period f_0, and the parameter in quadratic adjustment cost ξ. Let capital investment be $i_{it} = k_{i,t+1} - k_{it}$. The three moments are the exit rate of the firm entrants (11%), the autocorrelation of firm investments (-0.21), and the correlation of investment and productivity (0.17).[12]

Table 3 reports each parameter I used in the calibration. In particular, the calibrated value of correlated distortion τ_k is 0.5. The positive value is consistent with the positive correlation between distortion and fundamental in the literature (David and Venkateswaran, 2017; Yang, 2016). In addition, Bento and Restuccia (2017) and Fattal-Jaef (2018) show evidence of stronger correlation in poorer countries than in richer countries. Because a large correlation can potentially offset the positive correlation between capital investment and productivity, it is helpful to get a sense of the magnitude of τ_k in the calibration. Appendix Figure A.11 plots the policy functions of $k_{i,t+1}$ in the calibrated model, which shows $k_{i,t+1}$ is always increasing in k_{it}, as I discussed earlier. In addition, the intuition that firms with lower capital stock and lower idiosyncratic productivity are more likely to exit carries to the calibrated model with distortions. However, firms with the strongest beliefs of fundamentals and the highest contemporary productivities choose smaller capital stocks in the next period

[12] Because the China Annual Industry Surveys keep the non-state firms only if their revenues are above 5 million RMB, exit in the survey does not necessarily mean the firm goes out of business. To get a more precise measure of firm exit rates, I searched the operating status of a random sample of firms that exited from the survey during my sample period on the "National Enterprise Credit Information Pulicity System". Among the 528 firms I did find a record, 58% of the firms did shut down and unregistered. Therefore, I calibrate the model to target the adjusted exit rate of 11% rather than the 19% attrition rate for the firm entrants in the survey. If I nonetheless targets an exit rate of 19% in the calibration, the model then over-explains 10% of the data dynamics of MRPK dispersion.

Table 3: Parameter Values

Parameter	**Value**
Panel A: Pre-assigned Parameters	
α_1 - Capital share	0.28
α_2 - Labor share	0.53
β - Discount factor	0.97
δ - Depreciation rate	0.1
γ - Exit discount in capital	0.5
λ - Exogenous Exit Rate	0.04
μ_x - Mean of fundamentals	0
Panel B: Exactly Matched Parameters	
σ_x^2 - Dispersion of fundamentals	0.70
σ_e^2 - Dispersion of transitory shocks	0.33
Panel C: Calibrated Parameters	
τ_k - Correlated distortion	0.50
f_0 - Fixed operating costs	0.41
ξ - Adjustment cost	7.20

than firms with weaker beliefs and fundamentals due to severe distortions, as shown in the bottom-right panel in Figure A.11. Therefore, correlated distortions in the model calibration under $\tau_k = 0.5$ are substantial, which strongly disincentivise more productive firms.

Table 4: Moments Targeted in the Model and Data

Moments	**Target**	**Model**
$\rho(i, z)$	0.17	0.17
Exit rate of the entrants	11.0	12.1
$\rho(i, i')$	-0.21	-0.18

Table 4 reports the targeted moments in the data and model, which match decently. Although the three parameters are disciplined jointly by three moments, some useful intuitions apply. As in standard firm models, fixed operation costs positively relate to exit rates; and capital adjustment costs are most informative about the autocorrelation of investments. The correlated distortions captured in τ_k are informative about the correlation between capital investments and productivity. This is because a larger τ_k mitigate the investment responses to the productivity signals for the young firms, but asymptotically $\rho(i, z)$ always goes to 1 for the mature firms, independent of the value of τ_k. Hence, τ_k is negatively associated $\rho(i, z)$.

5.2 Quantitative Predictions

I take the initial distribution of the firm cohort at entry as given in the data and report predictions of the calibrated model on the dynamics of MRPK dispersion over the firm cohort's life cycle. This section focuses on the model predictions over the first 10 years of the firm cohort's life cycle, where the data MRPK dispersion decreases robustly with firm-cohort age. In addition, the model MRPK dispersion stabilizes after age 10.

Figure 11: MRPK Dispersion (σ_{mrpk}) in the Model and Data

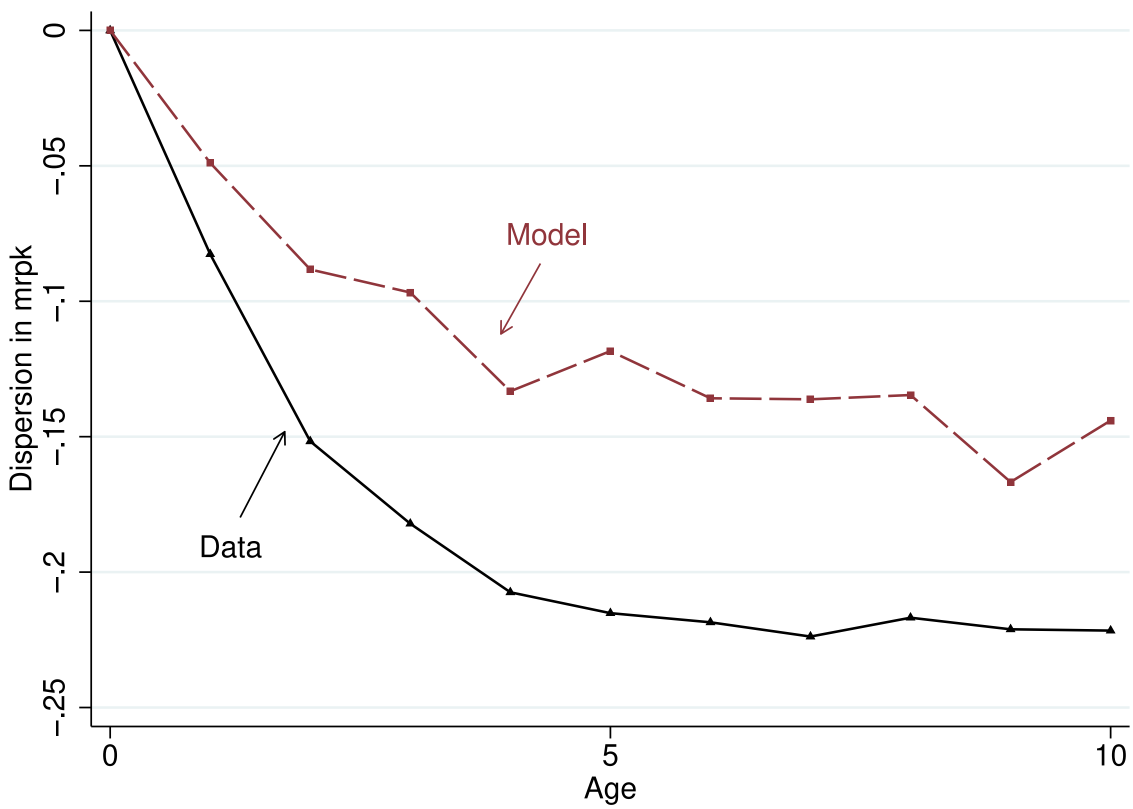


Figure 11 plots MRPK dispersion by firm-cohort age in the model and data. As the cohort of firms learn over time and adjust their capital stocks, the model predicts a decrease in MRPK dispersion by 0.15 points until age 10, compared to 0.22 in the data. Hence, the decrease in the model accounts for around two thirds of the magnitude in the data. Accordingly, within the firm cohort, σ^2_{mrpk} decreases by 0.43 (that is, $1.50^2 - 1.35^2$) from age zero to age 10, corresponding to 15% TFP gains, based on equation (11). The sizable TFP gains over the firms' life cycles suggest considerable improvements in how efficiently resources are allocated across firms within the firm cohort.

Table 5: Second Derivatives $\tilde{\phi}_j$ in the Model and Data

	Age 0	Age 1	Average of Age 2 - 9
Data 95% CI	(0.005, 0.05)	(0.02,0.05)	(-0.01, 0.02)
Model	0.01	0.03	-0.005

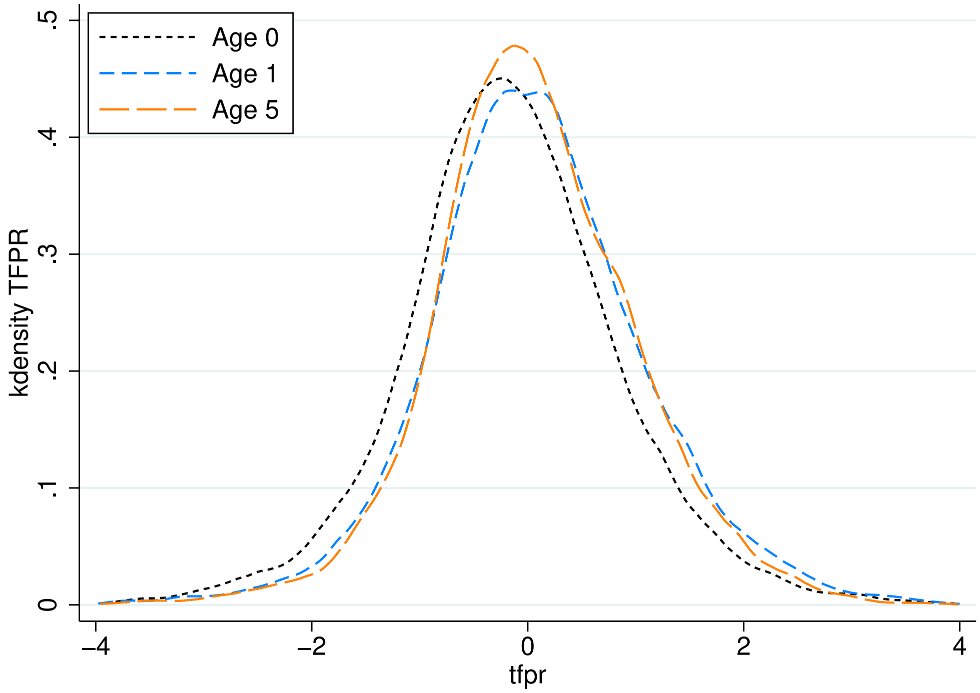
Furthermore, the model correctly predicts the convex relationship between MRPK dispersion and firm-cohort age, matching the curvature in the data without targeting it directly. Table 5 reports second derivatives of the age effects in the model, which are 0.01 at age zero and 0.03 at age one, respectively. These estimates fall right in the confidence interval of second derivatives in the data, as plotted in Figure A.6. The average second derivative for firms between two and nine years of age is close to zero in the model, consistent with the insignificant values in the data. That MRPK dispersion decreases at a decreasing rate with firm age both in the model and data is consistent with the theory of firm life-cycle learning. For young firms, the number of observations is small, which limits the precision of firm priors. Hence, marginal gains of learning are larger at younger ages, which leads to larger decreases in MRPK dispersion.

To emphasize the selection in exit over the firms' life cycles, Figure 12 plots the distribution of firm productivity at age 0, 1 and 5. As in the data, the model predicts that the productivity distribution shifts to the right (i.e., the average productivity increases) as less productive firms exit over time. The growth rate of average productivity from age zero to age one is around 5.7% in the model, which matches the growth rate of 5.4% in the data, without targeting it directly.

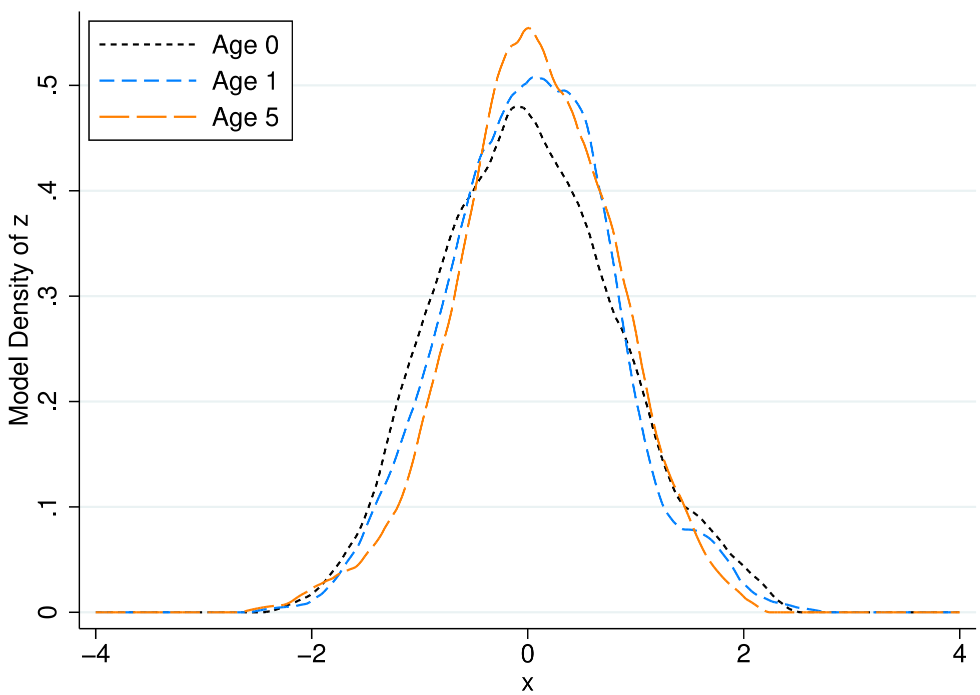
In order to understand the quantitative contribution of each mechanism in the decrease in MRPK dispersion, I simulate how MRPK dispersion changes with firm-cohort age by sequentially adding mechanisms in the model. In the basic version, I shut down the exit channel by setting the fixed operation cost and capital fire-sale value to zero, and shut down the learning channel by solving the optimization problem when the firms never updated their beliefs. I find the dispersion barely decreases over age in this scenario. Specifically, σ_{mrpk} decreases 0.007 points by age 10 in a model without firm life-cycle learning and an endogenous exit option, which is only 5% of the 0.15-point decrease in the benchmark model. Next, I add Bayesian updating to firms' beliefs about their fundamental productivity but

Figure 12: Distributions of Productivity in Model and Data

a) Life-Cycle Productivity in the Data



b) Life-Cycle Productivity in the Model



still do not allow endogenous exits. The model then predicts a decrease of 0.08 points in MRPK dispersion by age 10, which accounts for as much as 54% of the decrease in the benchmark model. Further adding endogenous exit brings the model back to the benchmark version and accounts for the remaining half of the life-cycle decrease in MRPK dispersion as plotted in Figure 11.[13]

What are then the consequences of life-cycle learning for aggregate TFP rather than for one firm cohort? I begin with a hypothetical baseline in which all firms have completed their life-cycle learning. In particular, I assume MRPK dispersion within each firm cohort remains constant after age 10 in the stationary equilibrium. This assumption is consistent with quantitative predictions in the calibrated model. In effect, I regard the firm cohorts age 10 and older as having learned sufficiently about their fundamental productivities that they cannot reduce their levels of MRPK dispersion by further learning.

Consider the model predictions on two moments: the age distribution of firms, and MRPK dispersion at each age. The aggregate MRPK dispersion is given by the average MRPK dispersion across all firm ages weighted by the number of firms at each age in the equilibrium, that is, 1.46 in the model. Meanwhile, aggregate MRPK dispersion in the hypothetical baseline is calculated by replacing the model MRPK dispersion across firms at ages zero to nine with the dispersion of age-10 firms, while keeping the age distribution of firms the same as in the benchmark model predictions. Mechanically, the aggregate MRPK dispersion is lower in the hypothetical baseline than in the model, because of the absence of firm life-cycle adjustments. I can use equation (11) to compute the implied TFP losses, Δz, for any given model prediction on σ_{mrpk} relative to the hypothetical baseline.

In the first column of Table 6, I report the differences in aggregate MRPK dispersion and in log TFP between predictions of the benchmark model and its hypothetical baseline. Aggregate MRPK dispersion in the hypothetical baseline is 0.11 points lower. This difference shows firm life-cycle adjustments accounts for 7% of MRPK dispersion across firms in the economy, which lead to a 10 percent loss in TFP.

To consider the consequences of firm-level distortions for aggregate TFP, I conduct the counterfactual experiment of removing firm-specific distortions by setting τ_k to 0 in the benchmark model. The standard deviation of MRPK (σ_{mrpk}) across firms at age 10 becomes 1.25, which is 0.1 points smaller than in the benchmark model. As reported in the second column of Table 6, in the corresponding hypothetical baseline, which removes both distor-

[13] If I consider the decomposition of aggregate capital stock within the firm cohort by age, from age zero to age 10, as in Olley and Pakes (1996), the covariance between capital stock and market share (defined by revenue output share) increases by 72%, from 0.11 to 0.19. This increase in covariance with firm age is consistent with the theory of firm life-cycle learning, but unlike the quantitative analysis of my model, it cannot estimate the contribution of life-cycle capital adjustments separately from learning.

Table 6: Consequences of Firm Life-Cycle Learning in the Model

	Learning	Distortions + Learning
$\Delta\sigma_{mrpk}$	0.11	0.21
$\frac{\Delta\sigma_{mrpk}}{\sigma_{mrpk}}$	7%	14%
Δz	0.10	0.19

tions and firm life-cycle adjustments, the aggregate MRPK dispersion would decrease 0.21 points, from 1.46 to 1.25. Hence, distortions and firm life-cycle learning together account for 14% of MRPK dispersion in the economy, which leads to a 19 percent loss in TFP. Omitting learning over the firm cohort's life cycle will attribute all changes in MRPK dispersion in the hypothetical baseline to distortions, which causes more than half of the TFP losses to be incorrectly attributed to distortions.

I conclude that the model featuring firm life-cycle learning explains around two thirds of the life-cycle MRPK dispersion. Without targeting the curvature of age effects and the productivity growth over the firm cohort's life cycle directly, the model correctly matches these moments in the data. Through the lens of this model, omitting firm life-cycle learning leads to a sizable overestimation of TFP losses from misallocation.

6 Evidence from Colombia and Chile

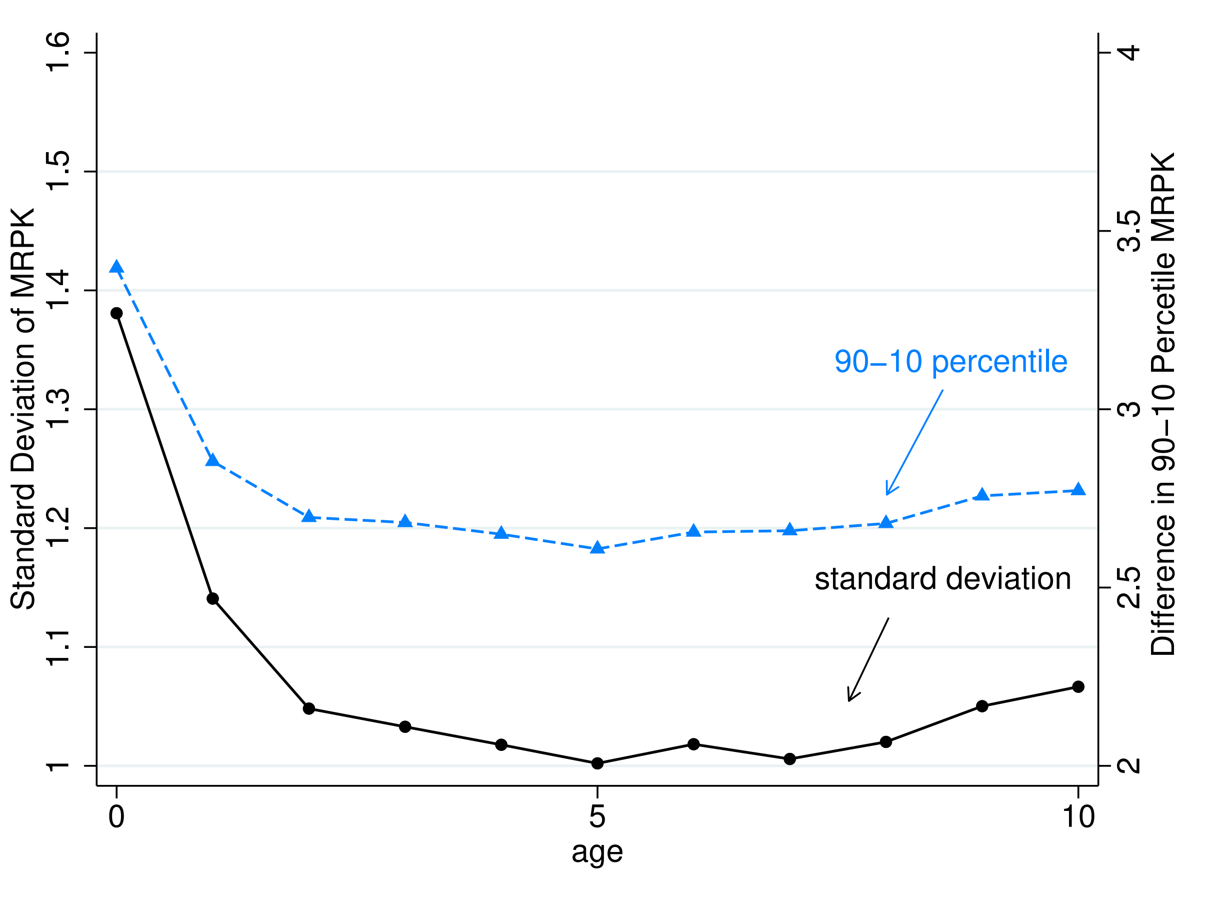
In this section, I report patterns of MRPK dispersion over the firms' life cycles using older data from the manufacturing sectors in Colombia and Chile.[14]

The Colombia Industrial Surveys during the period 1977 - 1991 cover around 6,600 plants per year on average. I measure the capital stock (k_{it}) as the book value of fixed assets, and measure revenue output (y_{it}) as value added constructed by subtracting intermediate inputs from the sum of the value of production, inventory changes, and sales tax (Roberts and Tybout, 1996). Again, I use the industry-level capital share from the NBER-CES database and equation (3) to calculate MRPK in log terms. To keep sufficient observations to measure dispersion, I calculate the standard deviation of MRPK across plants within the same year-age bins, rather than the same industry-year-age bins.

Figure 13 reports the average MRPK dispersion with firm cohorts measured in two ways as firms age from zero to 10 in Colombia. The standard deviation of MRPK decreases from almost 1.4 to 1 by age five, and remains below 1.1 until age 10. The log difference of MRPK

[14] I thank Mark Roberts for sharing his data with me.

Figure 13: MRPK Dispersion by Firm Age, Colombia



between the 90th and the 10th percentile plant decrease from 3.4 to around 2.6 by age five and stays at around 2.8 till age 10. That is, the ratio of the 90th to the 10th percentile MRPK drops one half, from 30 to around 15, during the first five years of the firm cohort's life cycle. In addition, similar to the convex age effects estimated using Chinese data, both measures of MRPK dispersion in Colombia decrease at a decreasing rate before age 10.

The data I have on the manufacturing sector in Chile cover plants with at least 10 employees during the period 1979 - 1986. Though the year of plant entry is not reported in the survey, based on the panel structure of the data, I can identify the year of plant entry t if one plant does not have a record in year $t-1$ but shows up in year t. Hence, the oldest plant cohort with a well-defined plant age is established in 1980 and can be observed until age five. The final sample size grows from 226 plants in 1980 to 1,037 plants in 1986.

Using the older and much smaller dataset from Chile, I measure revenue output (y_{it}) as value added, and measure capital stocks (k_{it}) by summing up the annual investments in buildings, machinery, and vehicles net of depreciation since birth year (Roberts and Tybout, 1996). Then I calculate MRPK in log terms using equation (3) as before, and I calculate the standard deviation of MRPK across plants within the same year-age bins. I find that, between firm-cohort age zero and five, the average standard deviation of MRPK in Chile decreases from 1.7 to less than 1.2, and the average log difference of MRPK between the 90th and the 10th percentile plant decrease from 4.5 to around 2.5. Note the decrease in MRPK dispersion in the Chilean data is larger than that in China and Colombia during the same age range. Because of the large confidence intervals due to the small number of firms in Chile, I report the t-test results of the differences of average MRPK dispersion between age 0-1 and age 2-5 firms. Table 7 shows that both the standard deviation and 90-10 percentile difference of MRPK are significantly larger for young firms in Chile.

Table 7: Chile: MRPK Dispersion by Firm Age Group

	Age 0-1	Age 2-5	Difference
Average $\sigma_{mrpk,tj}$	1.60	1.48	-0.12***
Average 90-10	4.00	3.62	-0.38***
Obs. of firms	1,935	989	

I conclude that evidence from Colombia and Chile is in accord with my finding using Chinese data that MRPK dispersion decreases over the firm cohort's life cycle. In Colombia, MRPK dispersion decreases substantially before age five and at a decreasing rate. As in the Chinese data, this pattern is consistent with the theory of firm life-cycle learning, which has larger impacts at younger ages.

7 Conclusion

This paper provides a new interpretation of MRPK dispersion as firm life-cycle learning. I draw on the panel firm-level data in China to document substantial decreases in MRPK dispersion with firm-cohort age. In addition, for young firm cohorts, MRPK dispersion decreases substantially and at a decreasing rate. The pattern also holds broadly for data on the manufacturing sectors in Colombia and Chile. Building on the new facts, I develop a dynamic model featuring informational frictions over the firm cohort's life cycle as the firms learn about their own fundamental productivity. The model predicts that as firms learn over time and adjust their capital stocks, possibly through endogenously exiting the market, MRPK dispersion decreases over their life cycles. I highlight the importance of firm life-cycle learning to ex-post aggregate MRPK dispersion. Quantitative analysis suggests that omitting this dimension leads to sizable overestimation of the TFP losses due to misallocation. In addition, TFP losses resulting from firm life-cycle learning to overcome informational frictions is an optimal constrained equilibrium, which may not be fixed by policy interventions.

Though direct measurements of firm- or individual-level information learning is scarce, Tanaka, Bloom, David, and Koga (2018) provide empirical evidence that more productive Japanese firms make more accurate forecasts about the macro economy. Their findings suggest learning may be endogenous: firms can pay costs to learn better information. Although the learning process in this paper is essentially mechanical and homogeneous across firms, I leave the discussion of richer learning models to future research.

This paper shows that data from developing countries generally show decreasing MRPK dispersion over the firm cohort's life cycle. Further exploration of the profiles of life-cycle MRPK dispersion in developed countries would be worthwhile. Comparing economies at different income levels can potentially shed light on the theory of cross-country TFP.

References

Akcigit, U., H. Alp, and M. Peters (2016): "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries," Unpublished Working Paper, University of Chicago.

Antras, P. (2015): *Global Production: Firms, Contracts, and Trade Structure*. Princeton University Press.

Argente, D., M. Lee, and S. Moreira (2018): "Innovation and product reallocation in the great recession," *Journal of Monetary Economics*, 93, 1 – 20, Carnegie-Rochester-NYU Conference on Public Policy held at the Stern School of Business at New York University.

Asker, J., A. Collard-Wexler, and J. De Loecker (2015): "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 105(1), 131–171.

Asturias, J., S. Hur, T. J. Kehoe, and K. J. Ruhl (2017): "Firm Entry and Exit and Aggregate Growth," Working Paper 23202, National Bureau of Economic Research.

Bachas, P., R. N. Fattal-Jaef, and A. Jensen (2018): "Size-dependent tax enforcement and compliance: global evidence and aggregate implications," Unpublished Working Paper, The World Bank.

Bai, Y., K. Jin, and D. Lu (2018): "Misallocation Under Trade Liberalization," Unpublished Working Paper, University of Rochester.

Bento, P., and D. Restuccia (2017): “Misallocation, Establishment Size, and Productivity,” *American Economic Journal: Macroeconomics*, 9(3), 267–303.

Bils, M., P. J. Klenow, and C. Ruane (2017): “Misallocation or Mismeasurement?,” Unpublished Working Paper, Stanford University.

Brandt, L., G. Kambourov, and K. Storesletten (2016): “Firm Entry and Regional Growth Disparities: the Effect of SOEs in China,” *University of Toronto mimeo.*

Brandt, L., T. Tombe, and X. Zhu (2013): “Factor market distortions across time, space and sectors in China,” *Review of Economic Dynamics*, 16(1), 39–58.

Brandt, L., J. Van Biesebroeck, and Y. Zhang (2012): “Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing,” *Journal of Development Economics*, 97(2), 339–351.

Buera, F. J., J. P. Kaboski, and Y. Shin (2011): “Finance and development: A tale of two sectors,” *American Economic Review*, 101(5), 1964–2002.

Calligaris, S., M. Del Gatto, F. Hassan, G. I. Ottaviano, and F. Schivardi (2018): “The productivity puzzle and misallocation: an Italian perspective,” Forthcoming, Journal of Population Economics.

Caselli, F. (2005): “Accounting for Cross-Country Income Differences,” in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf.

Chen, C., A. Habib, and X. Zhu (2018): “Contracting Frictions with Managers, Financial Frictions, and Misallocation,” Unpublished Working Paper.

Chen, C., T. Senga, C. Sun, and H. Zhang (2018): “Uncertainty, Imperfect Information, and Learning in the International Market,” Working paper, University of Hong Kong.

David, J., H. Hopenhayn, and V. Venkateswaran (2016): “Information, Misallocation and Aggregate Productivity,” *Quarterly Journal of Economics*, 131, 943–1005.

David, J. M., and V. Venkateswaran (2017): “The Sources of Capital Misallocation,” Working Paper 23129, National Bureau of Economic Research.

Deaton, A. (1997): *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. The World Bank.

Evans, D. S. (1987): “The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries,” *Journal of Industrial Economics*, pp. 567–581.

Fattal-Jaef, R. N. (2018): "Entry and Exit, Multiproduct Firms, and Allocative Distortions," *American Economic Journal: Macroeconomics*, 10(2), 86–112.

Guner, N., G. Ventura, and Y. Xu (2008): "Macroeconomic implications of size-dependent policies," *Review of Economic Dynamics*, 11(4), 721–744.

Hall, R. E. (1971): "The measurement of quality change from vintage price data," in *Price indexes and quality change*, ed. by Z. Griliches, chap. 8, pp. 240–271. Cambridge: Harvard University Press.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013): "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 95(2), 347–361.

Hamilton, J. D. (2017): "Why you should never use the Hodrick-Prescott filter," *Review of Economics and Statistics*, (0).

Ho, S.-J., and D. Ruzic (2018): "Returns to Scale, Productivity Measurement, and Trends in US Manufacturing Misallocation," in *2018 Meeting Papers*, no. 119. Society for Economic Dynamics.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, 1403–1448.

——— (2014): "The Life Cycle of Plants in India and Mexico *," *The Quarterly Journal of Economics*, 129(3), 1035–1084.

Hsieh, C.-T., and B. A. Olken (2014): "The Missing Missing Middle," Working Paper 19966, National Bureau of Economic Research.

Hsieh, C.-T., and Z. M. Song (2015): "Grasp the large, let go of the small: the transformation of the state sector in China," Discussion paper, NBER Working Paper No. 21006.

Jovanovic, B. (1982): "Selection and the Evolution of Industry," *Econometrica*, 50(3), 649–670.

Kehrig, M., and N. Vincent (2017): "Do firms mitigate or magnify capital misallocation? Evidence from plant-level data," US Census Bureau Center for Economic Studies Paper No. CES-WP-17-14.

Kerr, W. R., R. Nanda, and M. Rhodes-Kropf (2014): "Entrepreneurship as Experimentation," *Journal of Economic Perspectives*, 28(3), 25–48.

Lagakos, D., B. Moll, T. Porzio, N. Qian, and T. Schoellman (2018): "Life cycle wage growth across countries," *Journal of Political Economy*, 126(2), 797–849.

Li, H. (2015): "Leverage and productivity," *Unpublished manuscript: Stanford University.*

McKenzie, D. J. (2006): "Disentangling age, cohort and time effects in the additive model," *Oxford bulletin of economics and statistics*, 68(4), 473–495.

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

Moreira, S. (2017): "Firm Dynamics, Persistent Effects of Entry Conditions, and Business Cycles.," U.S. Census Bureau Center for Economic Studies Working Paper No. CES-WP-17-29.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297.

Ramey, V. A., and M. D. Shapiro (2001): "Displaced capital: A study of aerospace plant closings," *Journal of political Economy*, 109(5), 958–992.

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

——— (2017): "The causes and costs of misallocation," *Journal of Economic Perspectives*, 31(3), 151–74.

Roberts, M. J., and J. R. Tybout (1996): *Industrial evolution in developing countries: Micro patterns of turnover, productivity, and market structure*. Oxford University Press.

Rotemberg, M., and T. K. White (2017): "Measuring Cross-Country Differences in Misallocation," Discussion paper, Working Paper.

Senga, T. (2015): "A new look at uncertainty shocks: Imperfect information and misallocation," Discussion paper, Working Paper, School of Economics and Finance, Queen Mary University of London.

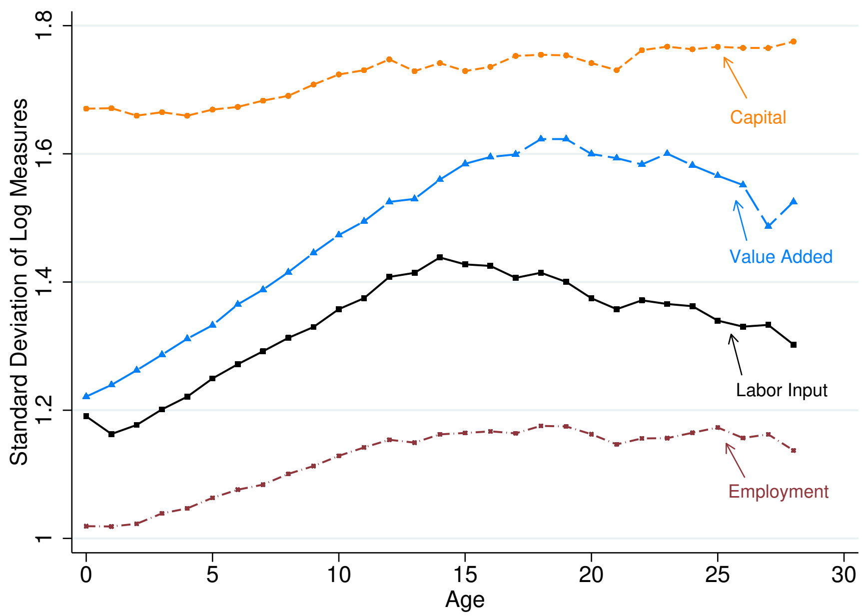
Song, Z., and G. L. Wu (2015): "Identifying capital misallocation," .

Tanaka, M., N. Bloom, J. M. David, and M. Koga (2018): "Firm Performance and Macro Forecast Accuracy," Working Paper 24776, National Bureau of Economic Research.

Yang, M.-J. (2016): "Micro-level Misallocation and Entry Selection," Unpublished Working Paper, University of Washington.

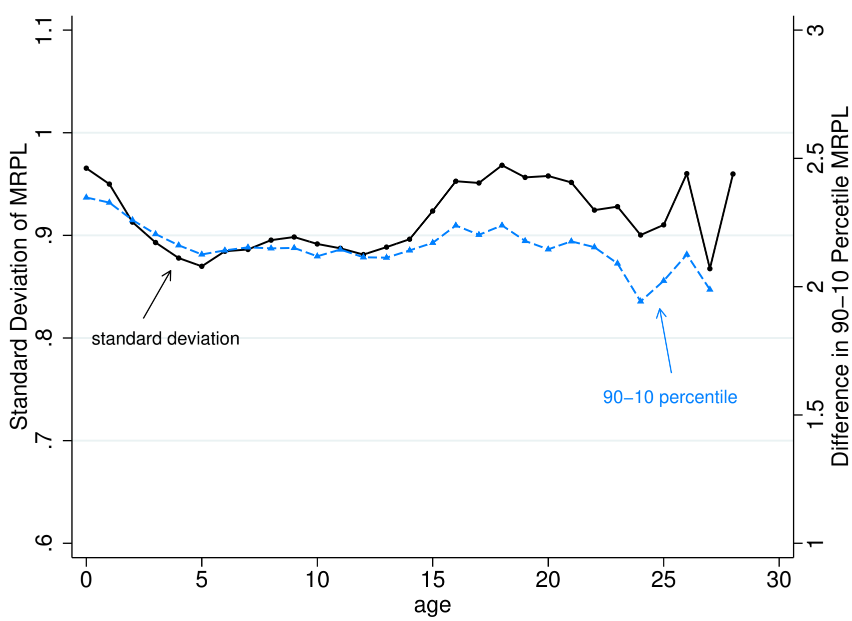
Appendices

Figure A.1: Dispersion of Key Variables by Firm Age



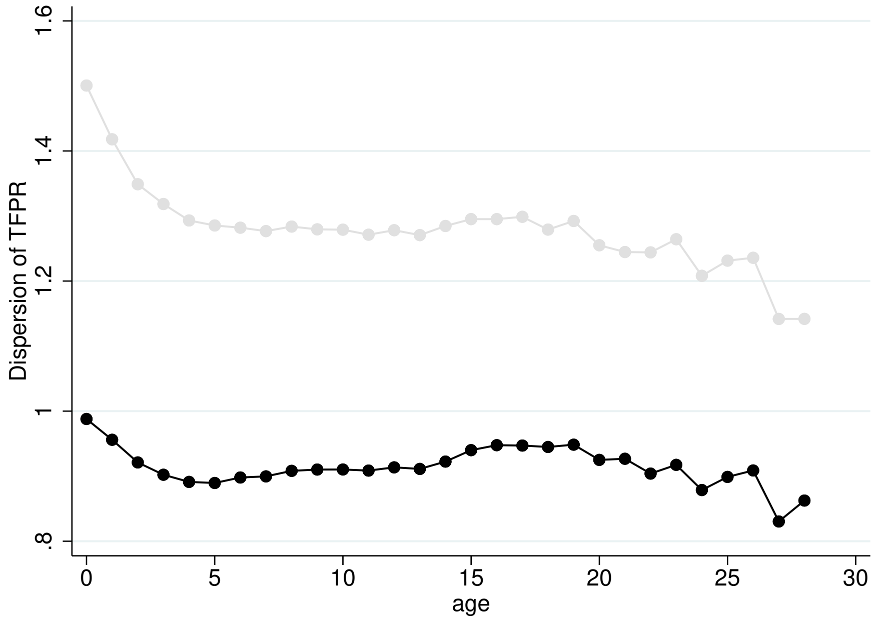
Note: This figure plots the standard deviation of log value-added (y_{it}), log capital input (k_{it}), log labor input (n_{it}), and log employment by firm age.

Figure A.2: Dispersion of MRPL by Firm Age



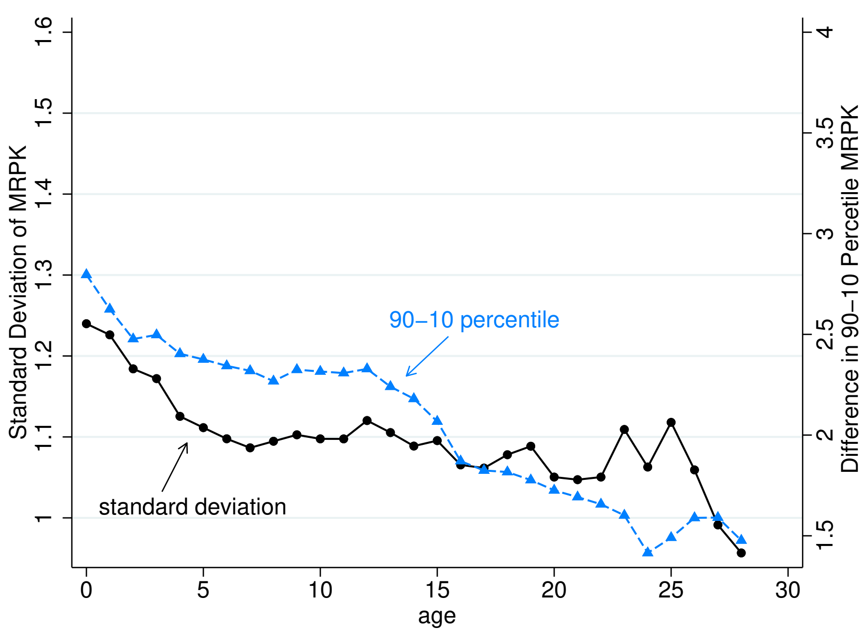
Note: This figure plots the weighted average standard deviation of MRPL and the weighted average value of the 90th minus the 10th percentile MRPL by firm age.

Figure A.3: Dispersion of TFPR ($\bar{\sigma}_{tfpr,j}$) by Firm Age



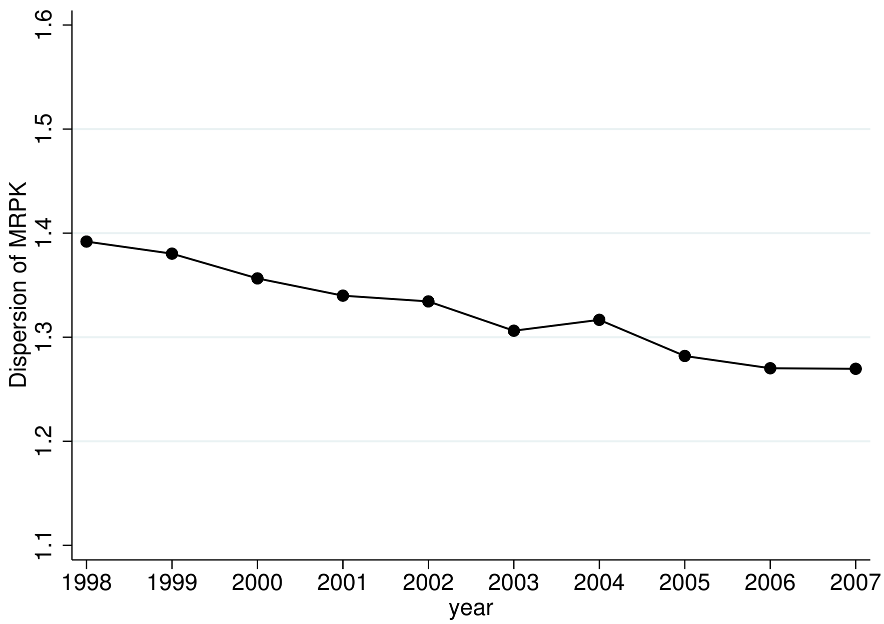
Note: This figure plots the average standard deviation of TFPR ($\bar{\sigma}_{tfpr,j}$) over age, weighted by the number of firms in industry-year-age bins. The gray line replicates the dispersion in MRPK as Figure 1 for reference.

Figure A.4: Dispersion of MRPK by Firm Age, Balanced Panel



Note: This figure plots the weighted average standard deviation of MRPK ($\bar{\sigma}_j$) and the weighted average value of the 90th minus the 10th percentile ($\bar{D}_j^{90-10}$) by firm-cohort age, for the firms are recorded every year during the sample period 1998 - 2007.

Figure A.5: Dispersion of MRPK by Year



Note: This figure plots the weighted average $\sigma_{mrpk,stj}$ during 1998 - 2007.

A.1 Curvature of the Age Effects

I use the second derivatives of age effects to test the curvature of age effects. Though none of the first-order effects of age, cohort, or time can be identified separately, their second derivatives are always identified (McKenzie, 2006). Recall that the cohort of firms aged j in time period t is denoted as c_{t-j}. Consider equation (5) for the cohort c_{t-j} observed in year t and $t+1$. To eliminate cohort effects, taking the first difference yields the sum of the first-order age effect between j and $j+1$ and the year effect between t and $t+1$: $\Delta\sigma_{mrpk,stj} \equiv \sigma_{mrpk,s,t+1,j+1} - \sigma_{mrpk,stj} = (\phi_{j+1} - \phi_j) + (\psi_{t+1} - \psi_t) + \Delta_c\epsilon_{stj}$, where $\Delta_c\epsilon_{stj} \equiv \epsilon_{s,t+1,j+1} - \epsilon_{stj}$. Consider an older firm cohort c_{t-j-1}, observed at age $j+1$ and $j+2$ in the same year t and $t+1$. Again, we can identify the sum of the first-order age effect and year effect: $(\phi_{j+2} - \phi_{j+1}) + (\psi_{t+1} - \psi_t)$. Now taking the difference of the two first-order effects gives the second derivative of age effects:

$$\tilde{\phi}_j \equiv (\phi_{j+2} - \phi_{j+1}) - (\phi_{j+1} - \phi_j).$$

The second derivative of age effects $\tilde{\phi}_j$ is the difference between two slopes: one slope of MRPK dispersion between age $j+2$ and age $j+1$, and the other slope between age $j+1$ and age j. If $\tilde{\phi}_j = 0$, that is, if the two slopes are the same, the age effects between age j and $j+2$ are linear. If $\tilde{\phi}_j > 0$, the profile of MRPK dispersion is convex between age j and $j+2$. Therefore, I can estimate $\tilde{\phi}_j$ to inform the curvature of age effects.

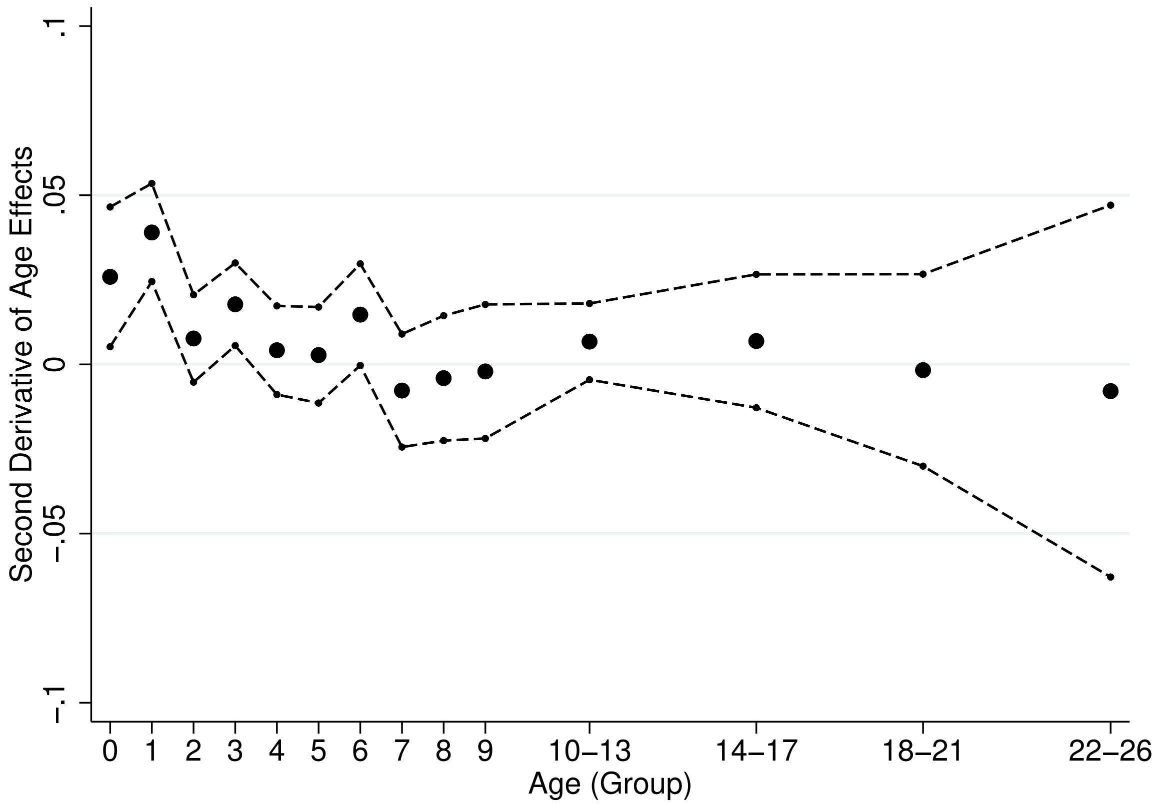
I estimate $\tilde{\phi}_j$ for each age j between zero and nine, and put the older ages into groups for tighter confidence intervals. Figure A.6 plots the second derivatives of age effects with 95% confidence intervals. It shows the second derivatives are significantly positive at age zero, one, and three, and become near zero and insignificant after age five. Based on the point estimates, firm age has convex effects on MRPK dispersion through the first five years of the firm cohort's life cycle; that is, MRPK dispersion decreases at a decreasing rate before age five.

Table A.1: McKenzie Test of Linear Age Effects

H_0: Linear Range	Age 0-5	Age 5-10	Age 10-28
P-value	0.00	0.30	0.15
H_0: Linear Range	Age 5-28	Age 4-28	Age 3-28
P-value	0.12	0.14	0.03

Note: This table reports the p-value of the McKenzie test of linear age effects over several age ranges.

Figure A.6: Second Derivatives of Age Effects



Note: This figure plots the estimates of the second derivatives of age effects $\tilde{\phi}_j$ with 95% confidence intervals. It is estimated for each age between zero and nine, and for each four-age group afterward in order to get tighter confidence intervals.

Table A.1 further reports p-values of the Mckenzie tests on linear age effects. It is essentially a formal Wald test for the null hypothesis H_0 of $\tilde{\phi}_j$ being jointly zero for a set of j values. Jointly zero second derivatives imply the corresponding age effects are linear. The tests in the first row strongly reject the hypothesis that the age effects are linear between age zero and age five, but cannot reject they are linear between age 5 to 10 or 10 to 28. The tests in the second row show one cannot reject the null hypothesis of linear age effects between age four or five and age 28. But the McKenzie test rejects the linear age effects with a p-value of 0.03 if one extends the age range to between three and 28. Based on these results, I will assume a linear trend in age effects after age 10 in the second alternative approach to identify the first-order age effects.

A.2 Lower and Upper Bounds of Age Effects

In this section, I show the two restrictions that I impose in the second alternative approach provide the upper and lower bounds of age effects if all three effects of age, year, and cohort on MRPK dispersion have non-positive trends.

Consider the case of a linear trend in the three effects of age, year, and cohort: $\phi_j = g_\phi j + u_{\psi,j}$, $\psi_t = g_\psi t + u_{\psi,t}$, and $\chi_c = g_\chi c + u_{\chi,c}$. The condition that all three effects of age, year, and cohort have non-positive trends on MRPK dispersion gives $g_\phi, g_\psi, g_\chi \leq 0$. I show below that (i) $g_\psi = 0$ (attributing the entire decline in MRPK dispersion over time to year effects) yields the upper bounder of g_ϕ, and (ii) $g_\chi = 0$ (attributing the entire decline in MRPK dispersion over time to cohort effects) yields the lower bounder of g_ϕ.

Substituting the identity of cohort birth year $c = t - j$ into the observed result, which is the sum of three effects:

$$\phi_j j + \psi_t t + \chi_c c = (g_\phi - g_\chi)j + (g_\psi + g_\chi)t + u,$$

where $u = u_{\psi,j} + u_{\psi,t} + u_{\chi,c}$. Denote $g_{M^*} = g_\phi - g_\chi$ and $g_M = g_\psi + g_\chi$. The unobserved negative trend in age effects g_ϕ can be expressed as $g_{M^*} + g_\chi$. Note that g_M is negative by definition; thus, the trend in cohort effects satisfies $g_\chi \in [g_M, 0]$, given the condition of three non-positive trends. Therefore, g_ϕ is bounded between $g_{M^*} + g_M$ and g_{M^*}.

The first restriction, which attributes the entire deline in MRPK dispersion over time to cohort effects, $g_\chi = 0$ is now equivalent to $g_\chi = g_M$. Hence, it gives the lower bound of the negative g_ϕ, that is, $g_{M^*} + g_M$. Similarly, the second restriction, $g_\chi = 0$, yields the upper bound of the negative g_ϕ, that is, g_{M^*}. Figure 6 shows the first restriction indeed yields a much steeper profile of MRPK dispersion with firm-cohort age.

A.3 Details of Alternative Approach Two

Here I explain the details of estimating equation (5) under the framework of imposing one additional linear restriction as in Deaton (1997). In particular, I describe the two different linear restrictions I impose for results in section 3.1 and how to implement them in practice.

To derive the restrictions, consider the weighted average dispersion of marginal products in year t:

$$SD_t = \sum_{c \in C_t} \omega_{stj} \cdot SD_{stj}(mrpk_{it}),$$

where ω_{stj} is a weight defined as the number of firms in an industry-age-year bin divided by the total number of firms. Let CIC denote the set of all 4-digit industry codes. Substituting in $SD_{stj}(mrpk_{it})$ from equation (5), it is easily shown that the weighted average can be written as

$$\begin{aligned} SD_t &= \alpha + \psi_t + \bar{X}_t + \bar{\Phi}_t \\ \bar{X}_t &= \sum_{c \in C_t} \frac{\Phi_{ct}}{\bar{\Phi}_t} \chi_c \\ \bar{\Phi}_t &= \sum_{c \in C_t} \Phi_{ct}, \text{and } \Phi_{ct} = \sum_{s \in CIC} \omega_{stj}(\phi_j D_j + \epsilon_{stj}). \end{aligned} \tag{A.1}$$

We see in Figure A.5 that the weighted average dispersion of marginal products of capital (or SD_t) declines from one year to the next. equation (A.1) shows clearly that the decline of dispersion has three sources: the decline due to the time effects ψ_t, the decline due to the aggregate cohort effects captured in $\bar{X}_t$, and the decline due to composition of firms at different ages captured in $\bar{\Phi}_t$. The restrictions will be imposed on the term

$$\Omega_t = \alpha + \psi_t + \bar{X}_t. \tag{A.2}$$

This term Ω_t captures the year-specific aggregate effects. It changes over time as a result of two effects: (i) cohort-neutral effects captured in ψ_t, and (ii) effects due to the changes in the composition of active cohorts operating, captured in $\bar{X}_t$. For example, if younger cohorts are born with a small dispersion of marginal products, the observed aggregate dispersion can decease over time only because young cohorts enter and older cohorts exit the market.

The basic idea of this approach is to decompose the time series of Ω_t into a trend component and a cyclical component. To identify cohort and year effects in addition to age effects, this approach makes assumptions on the relative role of time and cohort effects in the trend component.

In practice, the first step of implementation is to transform the time dummies as equation (2.94) in Deaton (1997) such that two restrictions are satisfied: (i) the year dummies add to zero: $\sum_{t=0}^{T} t = 0$, and (ii) the normalization of all year effects adding up to zero: $\frac{1}{T}\sum_{t=0}^{T} \psi_t = 0$. I also want to normalize the cohort effects $\bar{X}_t$ such that $\frac{1}{T}\sum_{t=0}^{T} \bar{X}_t = 0$. I do so by appropriately choosing the constant term α in equation (A.2). Second, the time series of ψ_t and $\bar{X}_t$ can be decomposed into a trend component and a cyclical component:

$$\psi_t = g_\psi t + u_{\psi,t}, \;\; \bar{X}_t = g_\chi t + u_{\chi,t}, \tag{A.3}$$

where $g_\psi = \frac{\sum_{t=0}^T \psi_t t}{\sum_{t=0}^T t^2}$ and $g_\chi = \frac{\sum_{t=0}^T \bar{X}_t t}{\sum_{t=0}^T t^2}$. Intuitively, the estimates are simply regressing ψ_t and $\bar{X}_t$ on time, thereby decomposing each time series into a trend component and the cyclical component orthogonal to time. It is the same method as proposed in Hamilton (2017). Finally, substituting equation (A.3) into equation (A.2) gives

$$\Omega_t = \alpha + g_M t + u_{M,t},$$

where $u_{M,t} = u_{\psi,t} + u_{\chi,t}$ and recall that $g_M = g_\psi + g_\chi$. The restrictions I used in Section 3.1 simply make assumptions on how g_M is split between g_ψ and g_χ.

I can also use the McKenzie test, as described in section A.1, to test the linearity restriction on the series of ψ_t. In practice, I first take the difference of MRPK dispersion of the same cohort observed in the two adjacent years to eliminate the cohort effects $\Delta_c \mathrm{SD}_{stj}$. Then I take the second difference for observations of the same age but in two adjacent years: $\Delta_c \Delta_a SD_{stj} = \Delta_c \mathrm{SD}_{stj} - \Delta_c \mathrm{SD}_{st'j}$. Therefore, I can test the hypothesis that the second derivative of time effects, $(\psi_{t+2} - \psi_{t+1}) - (\psi_{t+1} - \psi_t)$, is zero. As a result, I cannot reject the linear hypothesis except for t equal to 2003 and 2004, meaning linear specifications are good enough to estimate the time effects at all other sample years. This McKenzie test result is intuitive by looking at Figure A.5. We cannot reject the linear hypothesis at year 2003 and 2004, because the MRPK dispersion deviates from the linear fit in 2004, thus decreasing relatively significantly between year 2004 and 2005. Actually, even at this outlier, the deviation from the linear trend is only around 0.01 point in A.5. This deviation from the linear trend is relatively small compared to the age or time effects I estimated, which have magnitudes around 10 times larger. So I conclude that the linear restriction in the first approach is a reasonable approximation.

Specifically, the two restrictions I use to get the results in Figure 6 are the following:

Restriction 1 (All Decline due to Cohort Effects):

$$g_\psi = 0, \;\; g_\chi = g_M$$

By the definition of g_ψ, this restriction implies $\sum_{t=0}^T \psi_t = 0$, meaning that the year effects g_ψ only capture the cyclical variations and are orthogonal to the time trend. This restriction is the same as illustrated by (Deaton, 1997, pp. 123 - 127).

Restriction 2 (All Decline due to Time Effects):

$$g_\psi = g_M, \;\; g_\chi = 0$$

This restriction actually implies the linear restriction $\sum_{t=0}^{T} \bar{X}_t t = 0$, or

$$\sum_{t=0}^{T} \sum_{c \in C_t} \frac{\Phi_{ct}}{\bar{\Phi}_t} \chi_c t = 0.$$

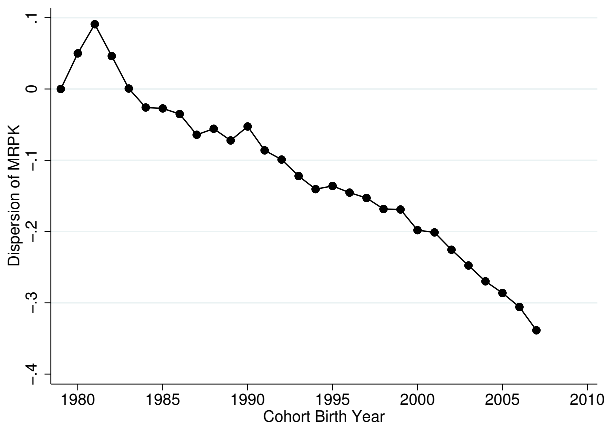
Note the term Φ_{ct} enters this restriction, which requires estimating equation (5). In practice, I use an iterative algorithm to meet this restriction.

Figure A.7 plots the estimates of the cohort and time effects under the two restrictions above. The top-panel results impose Restriction 1, so we see a declining trend in the cohort effects, but the time effects are relatively flat. The bottom-panel results impose Restriction 2, so the cohort effects are relatively flat but the time effects have a declining trend. Note the cohort and time variations are large, with the largest magnitudes at -0.3 for cohort effects and -0.1 for the year effects.

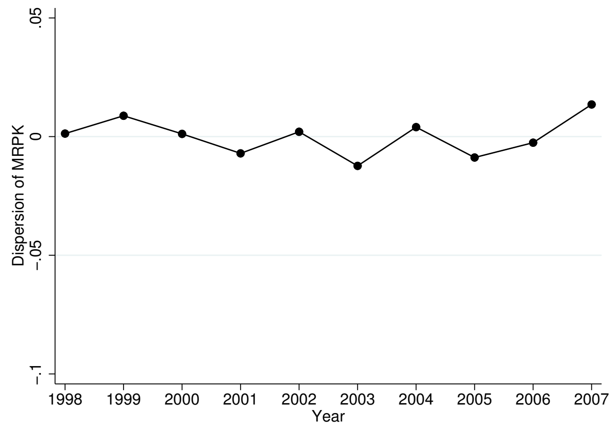
In addition, Figure A.8 plots the exit rates by firm-cohort age of the same firm cohorts in Figure 2 after removing zero-sum year effects using this methodology.

Figure A.7: MRPK Dispersion by Cohort and Year in Alternative Approach One

(a) All MRPK Decline Driven by Cohort Effects

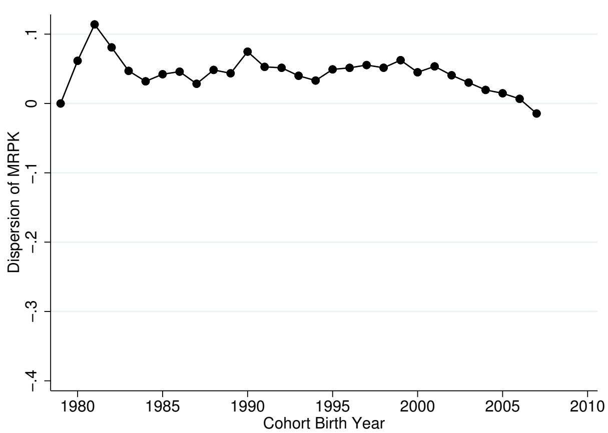


Cohort Profile χ_c

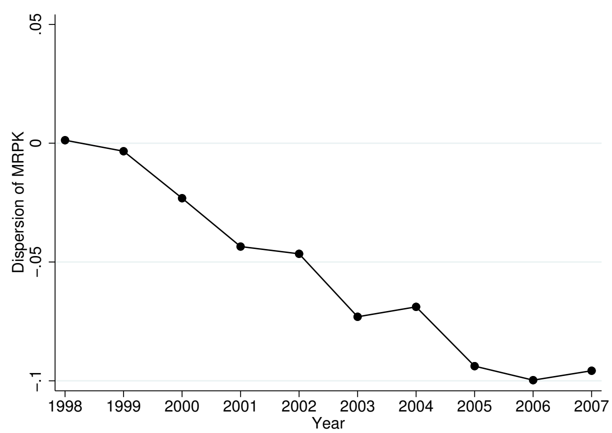


Year Profile ψ_t

(b) All MRPK Decline Driven by Time Effects



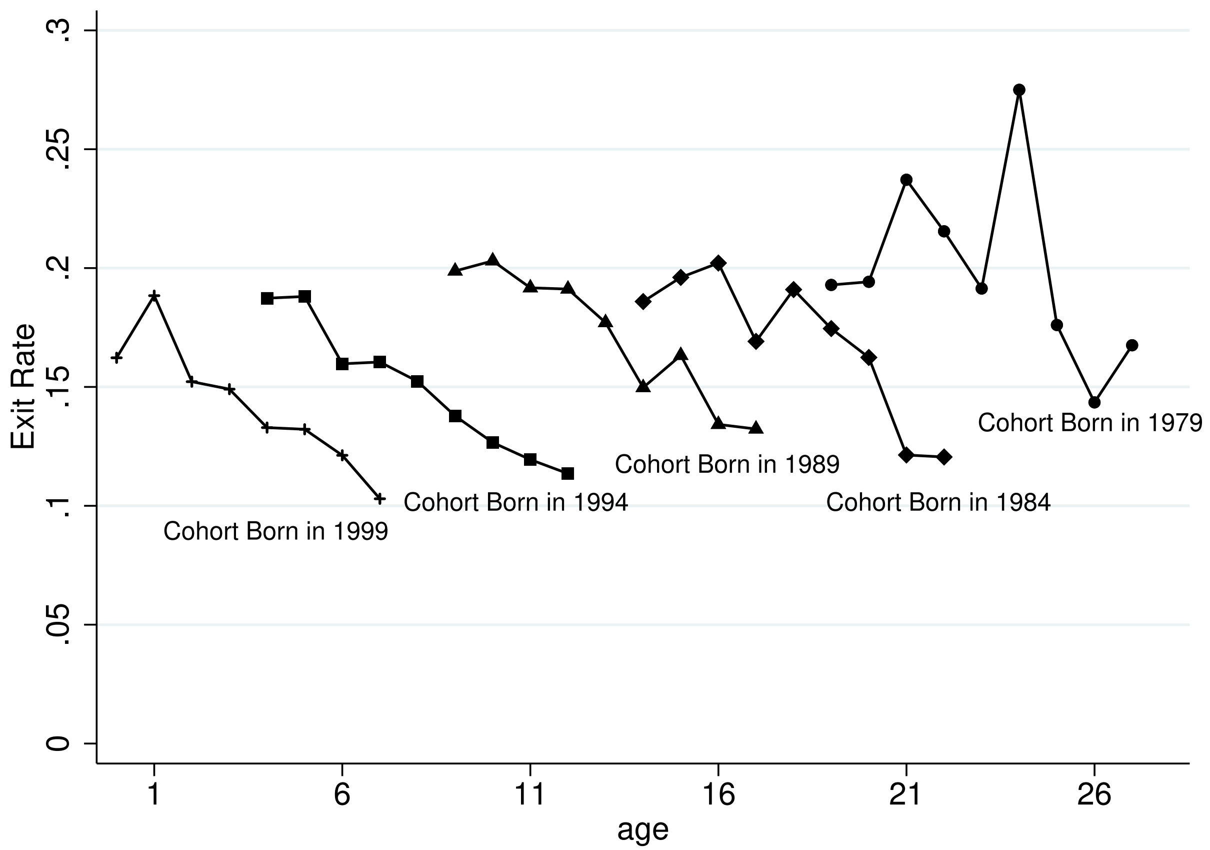
Cohort Profile χ_c



Year Profile ψ_t

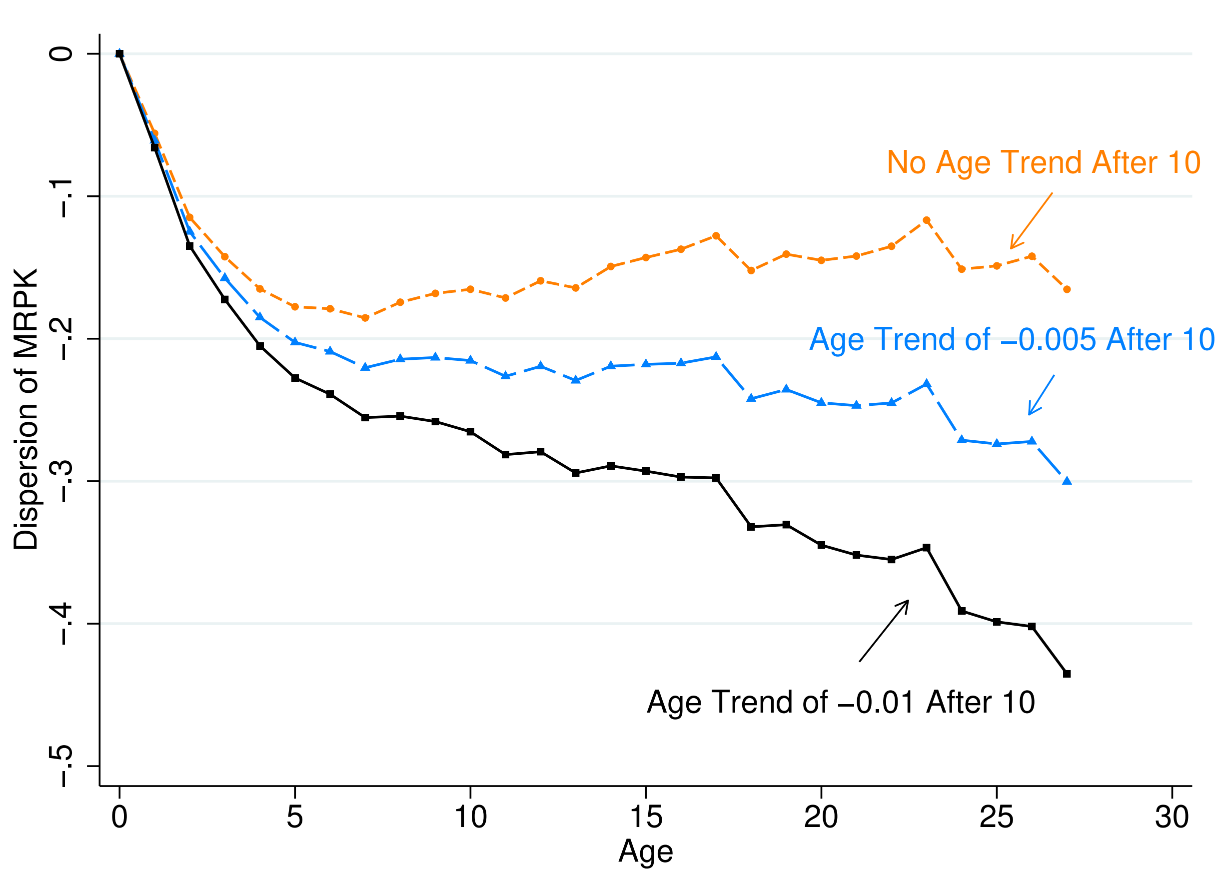
Note: This figure plots the MRPK dispersion by birth year of the firm cohorts and by calendar year estimated using the first alternative approach. The top panel shows the dispersion-cohort and dispersion-year profiles estimated in equation (5) using Restriction 1: $g_\psi = 0$. The bottom panel shows the dispersion-cohort and dispersion-year profiles estimated in equation (5) using Restriction 2: $g_\chi = 0$.

Figure A.8: Exit Rates by Cohort



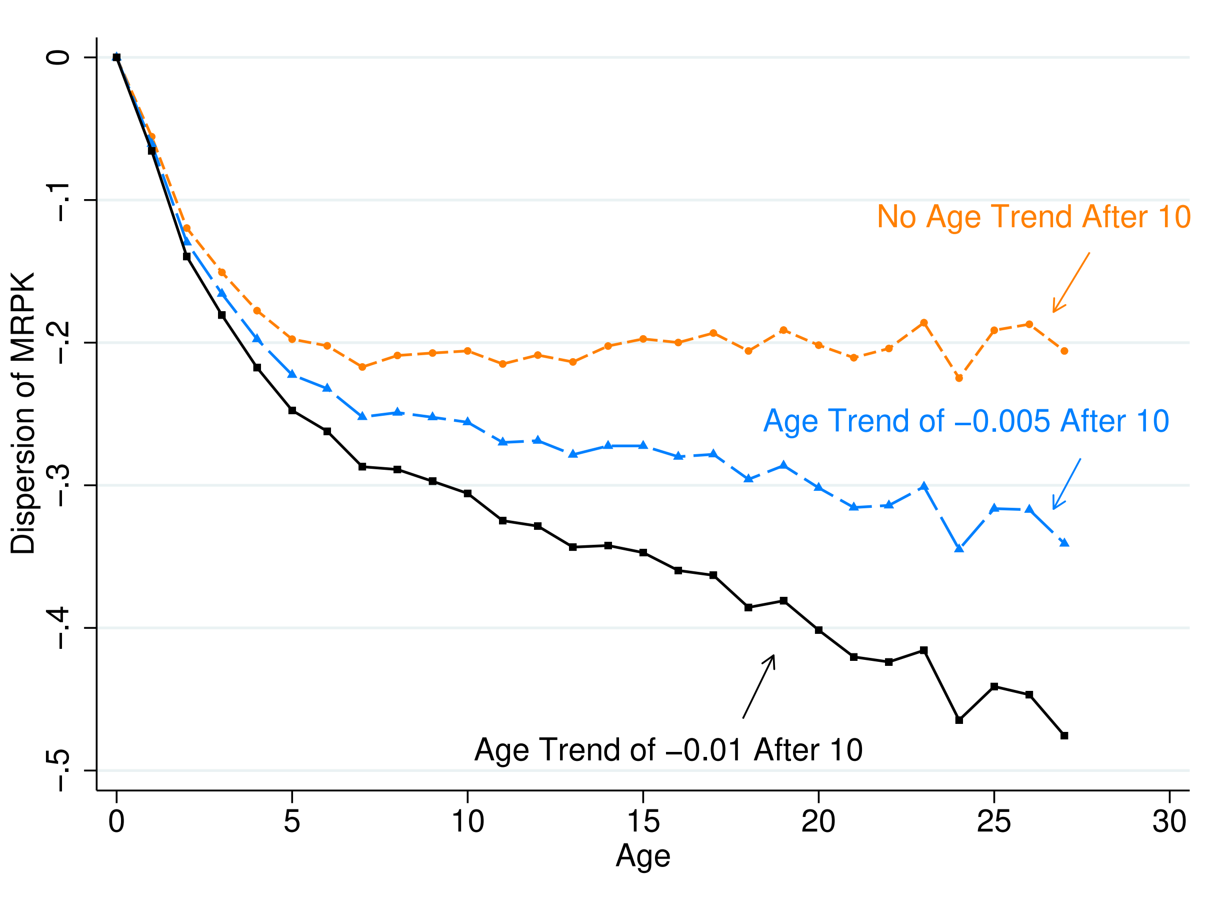
Note: This figure plots the exit rates by cohorts born in 1979, 1984, 1989, 1994, and 1999, respectively, after removing the zero-sum year effects following Deaton (1997).

Figure A.9: Dispersion Profiles over Age, Robustness with Volatility of Productivity



Note: This figure plots the estimated profile of MRPK dispersion by firm-cohort age in equation (6) using the second approach, which assumes (a) no trend in the age effects on MRPK dispersion after age 10 (dashed orange line with circle markers), (b) a small decreasing trend of 0.005 point per age after age 10 (long-dashed blue lines with triangle markers), (c) a moderate decreasing trend of 0.01 points per age after age 10 (solid black lines with square markers).

Figure A.10: Dispersion Profiles over Age, Robustness with Only Non-state Firms



Note: This figure plots the estimated profile of MRPK dispersion by firm-cohort age in equation (6) using the second approach when restricting the sample to only non-state firms. It plots the estimation results assuming (a) no trend in the age effects on MRPK dispersion after age 10 (dashed orange line with circle markers), (b) a small decreasing trend of 0.005 point after age 10 (long-dashed blue lines with triangle markers), (c) a moderate decreasing trend of 0.005 point after age 10 (solid black lines with square markers).

A.4 Aggregate Productivity and MRPK Dispersion

Substituting the optimal labor choice of $n_{it}(z_{it}, k_{it}) = \left(\alpha_2 \frac{e^{z_{it}} k_{it}^{\alpha_1}}{w}\right)^{\frac{1}{1-\alpha_2}}$ into the production function gives

$$y_{it} = \left(\frac{\alpha_2}{w}\right)^{\frac{\alpha_2}{1-\alpha_2}} e^{z_{it}\frac{1}{1-\alpha_2}} k_{it}^{\frac{\alpha_1}{1-\alpha_2}}. \tag{A.4}$$

Meanwhile, the labor market clearing condition requires that the fixed labor supply equals the aggregate labor demand $N = \int n_{it} di = \left(\frac{\alpha_2}{w}\right)^{\frac{1}{1-\alpha_2}} \int \left(e^{z_{it}} k_{it}^{\alpha_1}\right)^{\frac{1}{1-\alpha_2}} di$, so that $\left(\frac{\alpha_2}{w}\right)^{\frac{1}{1-\alpha_2}} = \frac{N}{\int \left(e^{z_{it}} k_{it}^{\alpha_1}\right)^{\frac{1}{1-\alpha_2}} di}$. Substituting this expression in y_{it} gives

$$y_{it} = \frac{e^{z_{it}\frac{1}{1-\alpha_2}} k_{it}^{\frac{\alpha_1}{1-\alpha_2}} N^{\alpha_2}}{\left(\int \left(e^{z_{it}} k_{it}^{\alpha_1}\right)^{\frac{1}{1-\alpha_2}} di\right)^{\alpha_2}}.$$

Further taking derivative with respect to k_{it} yields $MRPK_{it} = \frac{\alpha_1}{1-\alpha_2} \frac{e^{z_{it}\frac{1}{1-\alpha_2}} k_{it}^{\frac{\alpha_1+\alpha_2-1}{1-\alpha_2}} N^{\alpha_2}}{\left(\int \left(e^{z_{it}} k_{it}^{\alpha_1}\right)^{\frac{1}{1-\alpha_2}} di\right)^{\alpha_2}}$, which can be rearranged to express k_{it} in terms of $MRPK_{it}$:

$$k_{it} = \left(\frac{\frac{\alpha_1}{1-\alpha_2} e^{z_{it}\frac{1}{1-\alpha_2}}}{MRPK_{it}}\right)^{\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}} \cdot \left(\frac{N}{\int \left(e^{z_{it}} k_{it}^{\alpha_1}\right)^{\frac{1}{1-\alpha_2}} di}\right)^{\frac{\alpha_2(1-\alpha_2)}{1-\alpha_1-\alpha_2}}.$$

Meanwhile, capital market clearing condition implies

$$K = \int k_{it} di = \left(\frac{\alpha_1}{1-\alpha_2}\right)^{\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}} \left(\frac{N}{\int \left(e^{z_{it}} k_{it}^{\alpha_1}\right)^{\frac{1}{1-\alpha_2}} di}\right)^{\frac{\alpha_2(1-\alpha_2)}{1-\alpha_1-\alpha_2}} \int \left(\frac{e^{z_{it}\frac{1}{1-\alpha_2}}}{MRPK_{it}}\right)^{\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}} di.$$

Cancelling out the term with N in the last two expressions yields

$$k_{it} = \frac{\left(\frac{e^{z_{it}\frac{1}{1-\alpha_2}}}{MRPK_{it}}\right)^{\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}}}{\int \left(\frac{e^{z_{it}\frac{1}{1-\alpha_2}}}{MRPK_{it}}\right)^{\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}} di} K.$$

Now substituting k_{it} in terms of K into the expression of y_{it} and rearranging gives

$$y_{it} = \frac{\frac{e^{z_{it}\frac{1}{1-\alpha_1-\alpha_2}} MRPK_{it}^{-\frac{\alpha_1}{1-\alpha_1-\alpha_2}}}{\left(\int \left(\frac{e^{z_{it}\frac{1}{1-\alpha_2}}}{MRPK_{it}}\right)^{\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}} di\right)^{\frac{\alpha_1}{1-\alpha_2}}}}{\left(\frac{\int e^{z_{it}\frac{1}{1-\alpha_1-\alpha_2}} MRPK_{it}^{-\frac{\alpha_1}{1-\alpha_1-\alpha_2}} di}{\left(\int \left(\frac{e^{z_{it}\frac{1}{1-\alpha_2}}}{MRPK_{it}}\right)^{\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}} di\right)^{\frac{\alpha_1}{1-\alpha_2}}}\right)^{\alpha_2}} K^{\alpha_1} N^{\alpha_2}.$$

Finally, aggregating the revenue output y_{it} gives $Y = \int y_{it} di = Z K^{\alpha_1} N^{\alpha_2}$, where the aggregate productivity is

$$Z = \left(\frac{\int e^{z_{it}\frac{1}{1-\alpha_1-\alpha_2}} MRPK_{it}^{-\frac{\alpha_1}{1-\alpha_1-\alpha_2}} di}{\left(\int \left(\frac{e^{z_{it}\frac{1}{1-\alpha_2}}}{MRPK_{it}}\right)^{\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}} di\right)^{\frac{\alpha_1}{1-\alpha_2}}}\right)^{1-\alpha_2}.$$

Taking the log of the expression above gives

$$z = (1-\alpha_2)\left[\ln\left(\int e^{z_{it}\frac{1}{1-\alpha_1-\alpha_2}} MRPK_{it}^{-\frac{\alpha_1}{1-\alpha_1-\alpha_2}} di\right) - \frac{\alpha_1}{1-\alpha_2}\ln\left(\int e^{z_{it}\frac{1}{1-\alpha_1-\alpha_2}} MRPK_{it}^{-\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}} di\right)\right].$$

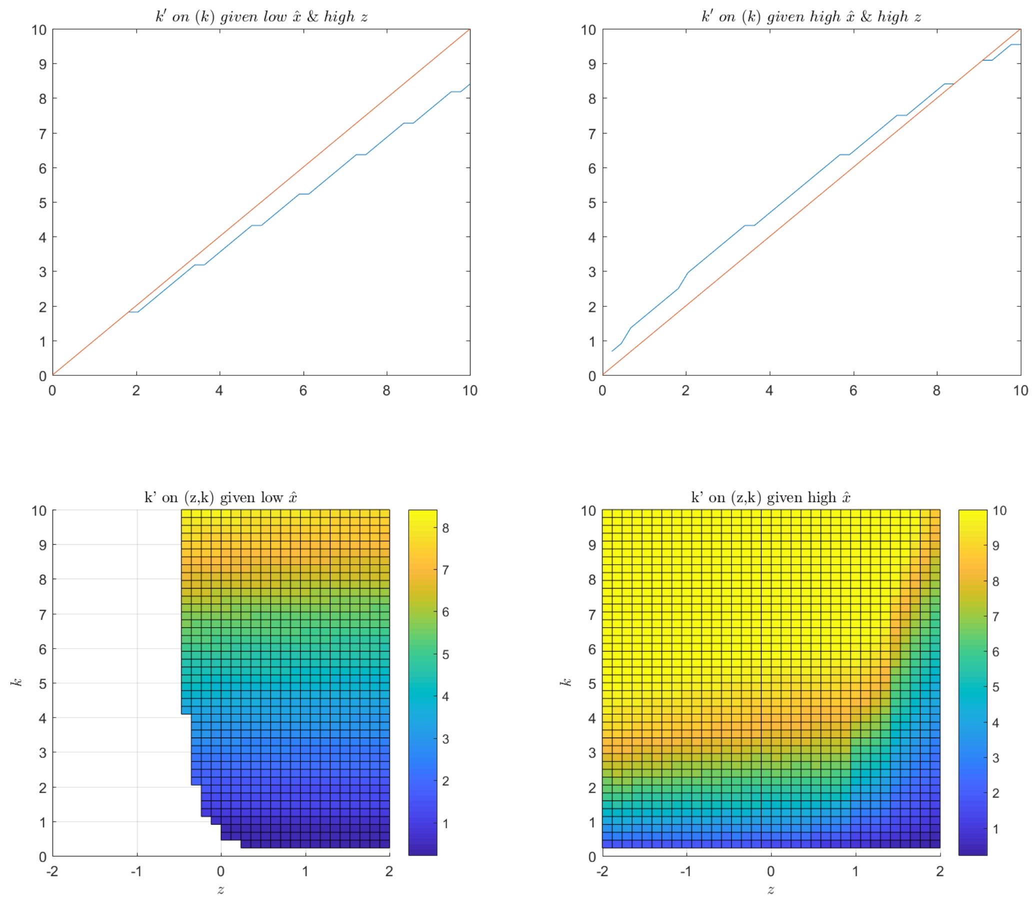
Expanding the two terms in the brackets respectively,

$$\ln\left(\int e^{z_{it}\frac{1}{1-\alpha_1-\alpha_2}} MRPK_{it}^{-\frac{\alpha_1}{1-\alpha_1-\alpha_2}} di\right) = \frac{\bar{z} - \alpha_1 \overline{mrpk}}{1-\alpha_1-\alpha_2} + \frac{\sigma_z^2 + \alpha_1^2 \sigma_{mrpk}^2 - 2\alpha_1 \sigma_{mrpk,z}}{2(1-\alpha_1-\alpha_2)^2},$$
$$\ln\left(\int e^{z_{it}\frac{1}{1-\alpha_1-\alpha_2}} MRPK_{it}^{-\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}} di\right) = \frac{\bar{z} - (1-\alpha_2)\overline{mrpk}}{1-\alpha_1-\alpha_2} + \frac{\sigma_z^2 + (1-\alpha_2)^2 \sigma_{mrpk}^2 - 2(1-\alpha_2)\sigma_{mrpk,z}}{2(1-\alpha_1-\alpha_2)^2}.$$

Finally, combining them into the expression of z reveals the relationship between the productivity loss ($z - z^*$) and dispersion in MRPK (σ_{mrpk}^2):

$$\begin{aligned} z &= (1-\alpha_2)\left[\bar{z} + \frac{\sigma_z^2 - \alpha_1 \sigma_{mrpk}^2}{2(1-\alpha_1-\alpha_2)}\right] \\ &= z^* - \frac{\alpha_1(1-\alpha_2)}{2(1-\alpha_1-\alpha_2)}\sigma_{mrpk}^2. \end{aligned}$$

Figure A.11: Policy Function Given Low and High Beliefs in Model



Note: This figure plots the policy function of $k_{i,t+1}$ for age-one firms. The top two panels fix the state variable z_{it} at a relatively high level and plot the firm choice of $k_{i,t+1}$ against k_{it} with low belief (left) and high belief (right); the bottom panels plot $k_{i,t+1}$ on the space of (k_{it}, z_{it}) with low belief (left) and high belief (right). Blank space in the top- and bottom-left two panels represent missing $k_{i,t+1}$ values when the firm chooses to exit the market.