Learning and the Capital Age Premium

Kai Li, Chi-Yang Tsou, and Chenjie Xu*

August 17, 2019

Abstract

We introduce imperfect information and parameter learning into a production-based asset pricing model. Our model features slow learning about firms' exposure to aggregate productivity shocks over time. In contrast to a full information case, our framework provides a unified explanation for the stylized empirical features of the cross-section of stocks that differ in capital age: old capital firms (1) have higher capital allocation efficiency; (2) are more exposed to aggregate productivity shocks and, hence, earn higher expected returns, which we refer to as capital age premium; and (3) have shorter cash-flow duration, when compared with young capital firms.

JEL Codes: E2, E3, G12
Keywords: parameter learning; capital age; cross-section of expected returns; capital misallocation; cash flow duration

---

*Kai Li (kaili@ust.hk) is an assistant professor of finance at Hong Kong University of Science and Technology; Chi-Yang Tsou (bentsou@ust.hk) is a research fellow of finance at Hong Kong University of Science and Technology; and Chenjie Xu (cxuab@ust.hk) is a PhD student of finance at Hong Kong University of Science and Technology. We are grateful for helpful comments from Hengjie Ai, Utpal Bhattacharya, Pierre Collin-Dufresne, Naveen Gondhi, Po-Hsuan Hsu, Yan Ji, Xiaoji Lin, Dimitris Papanikolaou, Jincheng Tong, Fan Yang and Chu Zhang. We also thank the seminar participants at Hong Kong University of Science and Technology, University of Technology Sydney, the Econometric Society European Winter Meeting, the NCTU International Finance Conference, the annual meeting of MFA, the Econometric Society Australian Meeting, the 4th SDU Finance Workshop, and the Taiwan Economics Research. We acknowledge the General Research Fund of the Research Grants Council of Hong Kong (Project Number: 16504618) for financial support.

1 Introduction

Parameter uncertainty is ubiquitous in economics and finance. In this paper, we study the role of parameter learning in a general equilibrium asset pricing model on cross-section of stock returns. In a large class of models that links production and investment to the cross-section of expected returns, we typically assume that market participants can directly observe firm-specific productivity and can distinguish its systematic component from the idiosyncratic component. We refer to this assumption as the full information paradigm. While this paradigm offers a natural starting place and an important analytic benchmark, we consider an alternative but more realistic specification in this paper: that in which individual firms' managers have imperfect information about their productivity; in particular, firms with newly installed capitals (i.e. young capital vintages) have limited information about their exposure to aggregate productivity shocks, but still receive noisy signals from which they learn over time. We show that this single deviation from the full information case coherently explains a wide set of empirical facts about the links between capital age, resource allocation efficiency, the timing of cash flows and, most importantly, expected returns in the cross-section of stocks.

Information about the separation between the systematic and idiosyncratic components of firm-level productivity is likely to be limited. Indeed, for newly installed capital, there is not enough historical data for firm managers to estimate its firm-specific exposure to aggregate productivity. Thus, the full information assumption may overestimate the amount of information that agents have. In this paper, we propose an alternative imperfect information structure: that individual firms' managers have imperfect information about their capital's productivity and, in turn, face a signal extraction problem. In particular, we assume young capital vintage firms [1] have less precise signals about their firm-specific exposure with respect to the aggregate productivity growth than old vintage firms.

This assumption is directly motivated and strongly supported by our empirical evidence in Section 2. Specifically, among U.S. public listed firms contained in the Compustat database, the dispersion of static marginal product of capital (MPK), which has been interpreted as a form of capital misallocation by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), monotonically declines with firms' capital age. We interpret this evidence by using a parameter learning channel, and it is consistent with our model prediction as summarized in Lemma 2 of Section 3. The intuition is that adolescent firms have less precise information

[1] In this paper, we focus on firm's capital age, rather than its founding year age. Although in the data there is a positive correlation between these two age measures, they may still carry different information. For instance, an old founding year age firm that has a lot of newly installed capital could have a relatively young capital age.

about their firm-specific productivity to facilitate resource allocations, which leads to higher capital misallocations. Feng (2019) documents a consistent negative relation between MPK dispersion and firm age by using firm-level panel data from China, Columbia, and Chile, and provides a firm life-cycle learning model to account for this empirical pattern, in a similar spirit to ours. Different from Feng (2019), we use such an evidence to motivate our assumption on the information heterogeneity among firms with different capital age, and focus on its asset pricing implications.

In this paper, we propose a novel general equilibrium production-based model. In the model, capital vintages differ in the information precision about their capital specific productivities, and firm managers have to learn about it over time. In our setting, the endogenous responses of firms' investment and payout to news about future productivity can provide a coherent explanation on the relationship among firm-level capital age, the cross-section of stock returns, and cash-flow duration. First and mostly importantly, the model generates a positive capital age and expected return relation in the cross section. Second, it also simultaneously produces a negative link between capital age and the duration of cash flow. In contrast, in the full information case, both the expected returns and cash flow durations are completely flat with capital age.

The intuition for these key model implications is as follows. As compared with young capital firms, old firms have more precise information about their exposure to aggregate productivity shocks, and are therefore more capable of taking advantage of aggregate technological progress. A higher information precision acts like a risk exposure amplification mechanism, and makes old capital firms more sensitive to aggregate productivity shocks, which implies a high average return. In other words, heterogeneity in information translates into heterogeneity in risk exposures across different capital vintages. For similar reasons, young capital firms have lower resource reallocation efficiency, are less able to take advantage of aggregate technology growth, and pay lower payouts than when they get older. Therefore, young capital firms feature longer cash flow duration, as their future cash flows obtain more weight. Old capital firms, in contrast, have shorter cash flow duration.

These model implications are strongly supported by empirical evidence. In order to investigate the empirical link between capital age, duration of cash flow, and expected returns in the cross-section, we first construct the capital age measure. Following Salvanes and Tveteras (2004), Ai, Croce, and Li (2013) and Lin, Palazzo, and Yang (2019), we measure a firm's capital age of U.S. public listed companies as a weighted average of the age of each capital vintage. Please refer to Appendix D.2 for details about capital age construction. We then implement the standard procedure and sort firms into quintile portfolios based on these firms' capital ages within Fama-French 30 industries. As we report in Table 2, the average equity excess return for firms with old capital age (Quintile O) is 5.79% higher on an annualized basis than that of young capital firms (Quintile Y). We call the return spread of a long-short old-minus-young (OMY) strategy the capital age premium. The return difference is statistically significant with a t-value of 2.91, and its Sharpe ratio is 0.44. The evidence on the capital age premium is consistent with the empirical finding in Lin, Palazzo, and Yang (2019). Specifically, the positive capital age-expected return relation is consistent with the first model implication that more precise information of old capital firms translates into higher risk exposure, and in turn, leads to a higher average return than that of young capital firms.

Next, we document a negative correlation between capital age and cash flow duration in Table 7, in which the cash flow duration is defined by Dechow, Sloan, and Soliman (2004), consistent with the second implication of the model we previously discussed that explores the relation between the timing of cash flow and capital age.

Guided by our theory, we also provide extensive empirical evidence that directly supports the key learning mechanism. First, we document that the productivity of new vintages of capital is less sensitive to aggregate productivity shocks than that of older vintages, consistent with our model implication. Second, we find that young capital firms have a lower learning rate about their exposure than old capital firms, consistent with the assumption about heterogenous information across capital vintages. Last but not least, we also empirically document that the normalized payouts for young capital firms have a lower exposure to both long-run and short-run aggregate productivity shocks than those of old capital firms, which strongly supports our model.

We also examine empirical evidence that differentiates our explanation from other alternative economic channels for the capital age premium. Lin, Palazzo, and Yang (2019) demonstrates that technology frontier shock (TFS) offers an additional source of risk that drives the capital age premium. We document that, even within industries that has low exposure to TFS shocks, as measured by a low rolling-window correlation between the aggregate TFS shock and various leads and lags of output (sales) growth at an industry level, the capital age premium is still present and significant, though its magnitude of the return spread is lower than that among high TFS exposure industries. This empirical evidence suggests that both channels may co-exist and significantly determine the capital age premium. Moreover, the learning mechanism in this paper also coherently derives the negative relationships between capital age, capital misallocation, and cash flow duration simultaneously.

In our quantitative analysis, we show that our model, when calibrated to match both conventional macroeconomic quantity dynamics and asset pricing moments, generates a significant capital age premium and a negative capital age and cash flow duration relation. As the data confirm, firms with older capital age have a higher average return and a shorter cash flow duration. Quantitatively, our model reproduces the joint empirical relationships between capital age, expected returns, and cash flow duration in the data quite well.

Literature review Our paper builds on the literature that studies the implication of learning on asset market valuations. Pástor and Veronesi (2009b) provide a comprehensive review of learning models in finance. David (1997), Veronesi (2000), and Ai (2010) study how learning and imperfect information affect both asset valuations and the risk premium on the aggregate equity market, while our work studies the implication of parameter learning on the cross-section of expected returns. Pástor and Veronesi (2009a) present a model in which learning impacts the life-cycle dynamics of firms and their exposure to aggregate risks. Their model implication is that young firms are less exposed to aggregate shock than older firms, which is consistent with our model; however, we provide a production-based general equilibrium model that generates rich implications with respect to links between capital misallocation, expected returns, cash flow durations and capital age. Croce, Lettau, and Ludvigson (2014) studies the role that learning plays with respect to the long-run cash-risk on the term structure of equity returns. Collin-Dufresne, Johannes, and Lochstoer (2016) studies the role that parameter learning plays with respect to the dynamics of long-run consumption risks and their implications for the aggregate stock market. Gondhi (2017) studies a general equilibrium model with endogenous learning and attention constraint and its implications on misallocation and asset prices. Our paper studies the role that parameter learning plays with respect to firm-level productivity and focuses on the cross-section of stock returns, at which none of the above-mentioned papers target.

The theoretical approach in this paper is connected to the investment-based asset pricing literature by endogenizing investment and linking it to the cross-section of expected returns.[2] Zhang (2005) provides an investment-based explanation for the value premium. Chan, Lakonishok, and Sougiannis (2001) and Lin (2012) focus on the relationship between R&D investment and stock returns. Eisfeldt and Papanikolaou (2013) develops a model of organizational capital and expected returns. As far as we are aware, the closest related work to our paper is Lin, Palazzo, and Yang (2019). That paper identifies a new source of risk for households: systematic fluctuations in the adoption of the latest technology, which is embodied in new capital to reach the stochastic technology frontier across firms. From the

[2]There is a vast literature on investment-base asset pricing models, but we do not attempt to summarize it here. A partial list includes Cochrane (1991), Jermann (1998), Li, Livdan, and Zhang (2009), Tuzel (2010), Lin (2012), Belo and Lin (2012), Eisfeldt and Papanikolaou (2013), Belo, Lin, and Bazdresch (2014a), Belo, Lin, and Vitorino (2014b), İmrohoroğlu and Tüzel (2014), Favilukis and Lin (2015), Belo, Li, Lin, and Zhao (2017), Belo, Lin, and Yang (2018), among others.

perspective of households, investing in old capital firms is risky. Old capital firms, unlike their younger counterparts, are more likely to upgrade their capital in the near future; as a result, their cash flows face higher exposure to technology frontier shocks. Even though both Lin, Palazzo, and Yang (2019) and our work document the capital age premium, the underlying mechanism is significantly different from that in Lin, Palazzo, and Yang (2019). The primary difference is that our general equilibrium framework with a vital learning mechanism endogenously generates the asymmetric exposures between old and young capital firms to the conventional aggregate productivity shocks. Furthermore, the learning mechanism in our paper also coherently drives the negative relationships between capital age, capital misallocation, and cash flow duration.

Moreover, our paper is related to the production-based asset pricing literature, for which Kogan and Papanikolaou (2012) provide an excellent literature survey. Compared with those in the prior literature, our model departs from existing models in two significant aspects. First, although we address the equity premium puzzle in ways that are similar to those in other papers in the literature, our model generates heterogeneity in asset risk premia through different exposures to the aggregate risk between old and young capital firms. Second, we nest a tractable vintage capital model into a general equilibrium in which individual firms have imperfect information about their productivity and must learn their productivity over time; in contrast, most previous papers assume perfect information in the model economy. Ai, Croce, and Li (2013) and Ai, Croce, Diercks, and Li (2018) also emphasize a vintage capital channel and account for the value premium and the term structure of aggregate dividend strips. In contrast, we focus on the parameter learning channel and coherently explain the link between capital age, capital misallocation, cash-flow durations, and stock returns in the cross-section.

Our work is related to a growing field of work that explores the impact of learning on capital misallocation. Using firm-level panel data from China, Columbia, and Chile, Feng (2019) finds that misallocation decreases with firm age, and provides a firm life-cycle learning mechanism to interpret the empirical finding. Also, using the U.S. firm-level data, we find a consistent negative capital age and MPK dispersion relation and use this empirical pattern as supporting evidence for our key model assumption of heterogenous information across firms with different capital ages to study the general equilibrium asset pricing implications of parameter learning. David, Hopenhayn, and Venkateswaran (2016) examines and quantifies the effects of resource misallocation across firms through information frictions. David, Schmid, and Zeke (2018) develops a theory to link misallocation to systematic investment risks, but without a learning channel. In contrast, we focus on asset pricing in the cross-section. In particular, we show that, through a learning channel, old capital firms are subject to less capital misallocation and feature higher integration with aggregate fluctuations. Therefore, old capital firms face higher aggregate risk exposure and are thus risky than those of young capital firms.

The rest of the paper is organized as follows. We summarize our motivating empirical facts on the relationship between capital age, capital misallocations, and expected returns in Section 2. We describe a general equilibrium model with production and parameter learning and analyze its quantitative asset pricing implications in Section 3. In Section 4, we provide direct supporting evidence on the learning mechanism and discuss the calibration of key learning parameters. In Section 5, we provide a quantitative analysis of our model. We present some additional testable implications are presented in Section 6. We then conclude this paper in Section 7.

2 Empirical facts

In this section, we present several empirical facts that motivate our interest in studying the link between imperfect information, parameter learning, and the cross-section of expected returns sorted on capital age. Here, we provide a brief description of the evidence, but will reserve details of the data construction to Appendix D.

First, we investigate the empirical link between capital age and capital misallocation.[3] Following Salvanes and Tveteras (2004), Ai, Croce, and Li (2013) and Lin, Palazzo, and Yang (2019), we measure a firm's capital age of the U.S. public listed companies as a weighted average of the age of each capital vintage. Details of capital age construction refer to Appendix D.2 . We then implement a standard procedure and sort firms into quintile portfolios based on these firms' capital ages within Fama-French 30 industries. In Table 1, we report the cross-sectional dispersion of the marginal product of capital (MPK, hereafter) within each quintile portfolio as a measure of capital misallocation, following Hsieh and Klenow (2009). Within each quintile portfolio of firms, we first calculate the MPK dispersion within narrowly defined industries, either at the Fama-French 30 industries level or at a more refined SIC two-digit level, and then average the dispersion across the industries.

[Place Table 1 about here]

From Table 1, we observe a salient inverse relationship between capital age and capital misallocation. That is, portfolios with higher capital age present lower capital misallocation,

[3] "Misallocation" is somewhat of a misnomer in our environment, as firms are acting optimally given the information at hand. We follow the literature and use the term to refer broadly to deviations from marginal product equalization.

ranging from 1.13 in the young capital age quintile to 0.77 in the old capital age quintile. Such a downward sloping pattern of misallocation across portfolios sorted by capital age are robust not only to different industry classifications but also to different measures of MPK dispersion, as used in Chen and Song (2013) or in David, Schmid, and Zeke (2018), respectively.[4]

The negative relationship between capital age and capital misallocation provides suggestive evidence to support our key model assumption that young capital firms contain less information about their exposure to common productivity shocks than old capital firms. Consistent with David et al. (2016), less information leads to lower resource reallocation efficiency. Our model predicts a negative capital age and capital misallocation relation, as summarized in Lemma 2 of Section 3, strongly supported by the evidence in Table 1.[5]

[Place Table 2 about here]

Next, we present the evidence on the cross-section of stock returns based on portfolios sorted by capital age. In Table 2, we report average annualized excess returns, t-statistics, and Sharpe ratios of the five portfolios sorted by capital age. The average equity excess return for firms with old capital age (Quintile O) is 5.79% higher on an annualized basis than that of young capital firms (Quintile Y). We call the return spread of a long-short, old-minus-young (OMY) strategy the capital age premium. The return difference is statistically significant with a t-value of 2.91, and its Sharpe ratio is 0.44, which is almost comparable to that of the aggregate stock market index (around 0.5). The evidence on the capital age premium is consistent with the empirical finding in Lin, Palazzo, and Yang (2019); however, we sort portfolios within industries to control for industry heterogeneity, while Lin, Palazzo, and Yang (2019) does not. In this paper, we propose a learning mechanism that emphasizes firms are uncertain about their firm-specific exposure. Therefore, our theory guides us to compare firms within the same industry, as we presume that firms in the same industry share the same industry exposure to aggregate productivity.

In sum, we suggest that the learning mechanism proposed in this paper provides a potential coherent explanation. On the one hand, old firms contain more precise information about their exposure to common productivity shocks, and hence have lower capital mis-

[4] The average MPK dispersion over the whole sample in our calculation is broadly consistent with that in Chen and Song (2013) and David, Schmid, and Zeke (2018), although the latter papers did not calculate the dispersions across capital-age-sorted portfolios.

[5] In our model, we assume old capital firms contain full information about their exposure, therefore, Lemma 2 suggests the group of old capital firms should not have capital misallocation. However, the evidence shows that this firm group (i.e. portfolio O) still displays a positive MPK dispersion. This may be attributable to other factors, such as, adjustment costs, financial constraints, taxes, and regulations that are not included in our model.

allocations; on the other hand, old firms with more information take better advantage of aggregate technological progress, and they therefore feature a high exposure to aggregate shocks, which they expect to result in higher average returns. In the next section, we develop a production-based general equilibrium model with a parameter learning mechanism to formalize our preceding intuition and to quantitatively account for the capital age premium.

3 Model setup

The key mechanism of our model is that firms learn about their exposure to aggregate productivity shocks over time. In this section, to illustrate the basic premise of our theory, we first describe the learning mechanism in a static framework. Then we incorporate learning into a dynamic general equilibrium model with different capital vintages to formalize our intuition and study its implications for the cross-section of expected returns.

3.1 Aggregation with learning

3.1.1 Static problem

We start with a static setup similar to that of Melitz (2003) and Hsieh and Klenow (2009). Consider a group of infinitesimal firm units that produce intermediate inputs, y_i. There is also a single final good Y produced by a representative producer in a perfectly competitive final output market. This final good producer combines intermediate inputs using a CES production function:

$$Y = \left[\int y_i^{\nu} di\right]^{\frac{1}{\nu}}, \tag{1}$$

in which the parameter ν controls for the elasticity of substitution between intermediate inputs. Firms use capital and labor to produce intermediate goods through the production function:

$$y_i = k_i^{\alpha}(A_i n_i)^{1-\alpha} \tag{2}$$

We assume that $A_i = e^{m+\beta_i \Delta a}$, in which Δa is a common shock that affects the productivity of all firms, β_i is the firm-specific exposure to the common shock Δa, and m is a constant. We assume the common productivity shock follows a Gamma distribution $\Delta a \sim \mathcal{G}(k_a, \theta_a)$ [6].

[6] The results of this section hold with any non-negative distributions. We restricts the common shock to be in the positive region because we focus on the implications of firms learning their exposure to technological progress. As for technological regress, our learning mechanism does not directly apply because the information about the exposures to old technologies is already there. Our setup is also flexible enough to account for the patterns of total factor productivity (TFP) growth. In the data, TFP growth is positive over 90% of the time. To generate the 10% negative TFP growth, we can set the constant parameter m to be a

We assume that firm managers do not know exactly their exposure, β_i, and have to make production decisions based on their interference on β_i. To facilitate a closed-form solution, we assume that conditioning on the common shock Δa, β_i has a prior of a normal distribution $N(\mu, \frac{1}{\Delta a}\sigma^2)$[7]. Before making the production decision, each firm receives a noisy signal of its own exposure:

$$s_i = \beta_i + \epsilon_i, \tag{3}$$

in which $\epsilon_i \sim i.i.d. N(0, \frac{1}{\Delta a}\tau^2)$, conditioning on Δa. The parameter τ determines the level of noise with respect to firm signals. When $\tau = 0$, firms have perfect information about their exposure to common shocks. As τ increases, firms are less certain about their exposure to common shocks, and input choices are less efficient. In the extreme case of $\tau \to \infty$, signals are not informative at all.

Each firm unit chooses capital and labor inputs to maximize expected profit under its information set:

$$\max_{k_i, n_i} E_s[k_i^\alpha (A_i n_i)^{1-\alpha} p_i] - Rk_i - Wn_i, \tag{4}$$

in which R is the capital rent and W is the wage rate, and E_s denotes the belief given signal s_i, which explicitly emphasizes that a firm takes its signal into consideration when making the production decision. p_i is the market price of the intermediate good j, which can be determined as the marginal product of intermediate input $\frac{\partial Y}{\partial y_i}$.

In this economy, we adopt the Dixit-Stiglitz aggregate production function among intermediate inputs with imperfect substitution; however, instead of using the monopolistic competition setup as in Hsieh and Klenow (2009), we assume that the intermediate good producers are perfectly competitive. This assumption allows us to focus on imperfect information as the only source of resource reallocation inefficiency in our dynamic setup[8]. Under this assumption, all the intermediate firms take price p_i as given when solving their maximization problems, and each individual firm's production quantity has no impact on the market price of the intermediate good, p_i.

negative number so that when the common shock Δa is small enough, TFP growth is negative.

[7] Hsieh and Klenow (2009) assume the prior distribution of exposure β_i to be $N(\mu, \sigma^2)$. Their setup implies that beta dispersion is constant over the business cycle while ours implies a counter-cyclical beta dispersion.

[8] This perfect competitive market setup promises that, under the current imperfect information structure, the allocation is Pareto optimal so that in the dynamic model we may solve the problem with a social planner. Note, however, comparing to the perfect information economy, there is a deadweight loss. Hsieh and Klenow (2009) assume monopolistic competition, which does not distort the allocation of capital across firms in their static model, but will lead to inefficiency in capital accumulation in our dynamic setup.

We define the aggregate production function of the firm group as

$$F(K,N) \equiv \left[\int \left(k_i^{\alpha}\left(A_i n_i\right)^{1-\alpha}\right)^{\nu} di\right]^{\frac{1}{\nu}}$$
$$\text{subject to: } \int k_i di = K, \tag{5}$$
$$\int n_i di = N,$$

in which for each i, the choices of k_i, n_i must be measurable with respect to firm i's information. That is, k_i and n_i can only be functions of the signal s_i. In Appendix A.1 , we prove that the optimality of resource allocation implies that the aggregate production of the firm group can be written as $Y = K^{\alpha}(\mathbf{A}N)^{1-\alpha}$, in which

$$\mathbf{A} = \left[\int E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di\right]^{\frac{1-\nu}{(1-\alpha)\nu}}. \tag{6}$$

For the sake of simplicity, we impose the normalization condition $\mu = 1 - \frac{1}{2}(1-\alpha)\nu\sigma^2$ throughout the paper.[9] As we show below, this normalization assumption implies that the exposure to the aggregate productivity Δa is 1 in the case of no information ($\tau = \infty$). The functional form of the group level production function is given by the following lemma:

Lemma 1. *The aggregate production function of the firm group is given by*

$$F(K,N) = K^{\alpha}(\mathbf{A}N)^{1-\alpha}, \tag{7}$$

in which log productivity is given by $\ln \mathbf{A} = m + \lambda(\tau^2)\Delta a$, *and* $\lambda(\tau^2)$ *is defined as*

$$\lambda(\tau^2) = \left[1 + \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{\sigma^4}{\sigma^2+\tau^2}\right]. \tag{8}$$

In the no private information case,

$$\lim_{\tau^2\to\infty} \lambda(\tau^2) = 1,$$

[9] This normalization condition is only for simplification purpose. We can always recover the original economy by simultaneously scaling up/down other parameters. For example, conditional on Δa, an economy with parameters and conditioning variable $(\mu, \sigma^2, \tau^2, \Delta a)$ can be represented by a normalized version $(1, \frac{\sigma^2}{\mu}, \frac{\tau^2}{\mu}, \mu\Delta a)$.

and in the full information case,

$$\lambda^* = \lim_{\tau^2 \to 0} \lambda(\tau^2) = 1 + \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\sigma^2. \tag{9}$$

Proof. See Appendix A.1 . □

The realized log marginal product of capital dispersion is given by the following lemma:

Lemma 2. *The realized log MPK dispersion (cross-sectional variance) follows:*

$$Var\left[\log(MPK_i) - \log(MPK)\right] = (1-\alpha)^2\nu^2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\Delta a. \tag{10}$$

Proof. See Appendix A.1 . □

Our paper studies the implications of different information precision for the cross-section of firms. Therefore, we do a comparative statics with respect to signal noise τ. We have three intuitive implications from the above two lemmas. First, as firms acquire better information about their productivity, they can better allocate capital and labor across each other. From Lemma 2 we can see that the realized MPK dispersion across firms decrease with respect to information precision. Apparently, as the signal noise decreases, capital misallocation decreases. When the firms receive perfect information, $\tau = 0$, the deadweight loss goes to zero. In Section 2, we document a negative empirical correlation between capital age and capital misallocation. This evidence is consistent with the theoretical prediction in Lemma 2, and supports our key model assumption that young capital firms contain less information about their exposure to common productivity shocks than old capital firms.

Second, better information leads to higher level of firm group productivity, exactly due to lower level of capital misallocation. If we compute the productivity gap of the firm group with signal noise τ with respect to the full information benchmark, we have

$$\begin{aligned}
\ln A(0) - \ln A(\tau^2) &= (\lambda^* - \lambda(\tau^2))\Delta a \\
&= \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\Delta a \\
&= \frac{1}{2}\frac{1}{(1-\alpha)(1-\nu)}Var\left[\log(MPK_i) - \log(MPK)\right].
\end{aligned}$$

In the static model setup, the productivity gap is proportional to the difference in capital misallocation of these two economies.

Third, and most importantly, better allocation induced by higher information precision also acts as a risk exposure amplification mechanism. To see this, let us inspect the firm group productivity expression $lnA = m + \lambda(\tau^2)\Delta a$: 1 unit increase in common productivity shock Δa corresponds to $\lambda(\tau^2)$ units increase in firm group productivity. Because $\lambda(\tau^2)$ is a decreasing function of τ, the firm group's exposure to common productivity shocks increases with information precision. The upper bound on the exposure is attained under the full information and denoted by λ^* The intuition is that, when firms are uncertain about their exposure to common productivity shocks, more information allows them to take better advantage of aggregate technological progress, and therefore they feature a higher exposure to aggregate shocks.

3.1.2 Dynamic perpetual learning

Now we extend our preceding setup to a dynamic setting. Firm i's productivity follows the following stochastic growth process:

$$A_{i,t} \exp\left(\sum_{u=0}^{t} m + \beta_{i,u}\Delta a_u\right), \tag{11}$$

in which $\{\Delta a_u\}_{u=0}^{t}$ is a sequence of non-negative common productivity shocks. For $u = 0, 1, ..., t$, $\beta_{i,u}$ is the exposure of firm i's productivity with respect to the common shock Δa_u. We assume that $\{\beta_{i,u}\}_{u=0}^{t}$ is i.i.d across firm i and over t, and has a prior distribution of $N(\mu, \frac{1}{\Delta a_u}\sigma^2)$ as in the static setup.

We allow for perpetual learning in our dynamic setting; that is, we allow firms to receive new signals about the entire history of their exposure coefficients in every period t. Note that the object that firm i manager tries to learn is the entire sequence of exposure parameters $\{\beta_{i,u}\}_{u=0}^{t}$. The exposure parameters are firm specific and time specific. Therefore, even with perpetual learning, managers will not obtain full information about their productivity. We describe the specific signal arrival process as follows. For a typical firm i starting its operation at time 0, its productivity $A_{i,t}$ follows equation (11). Due to perpetual learning, at period t, firm i will receive a sequence of new signals $\{s_{i,u,t}\}_{u=0}^{t}$, and each element in this sequence, $s_{i,u,t}(u \leq t)$, is a signal received by the firm manager i at period t that could be used to infer the exposure coefficient $\beta_{i,u}$. In other words, for a signal, $s_{i,u,t}(u \leq t)$, the second subscript u indexes the exposure coefficient $\beta_{i,u}$ that the signal is used for inference, and the third subscript t denotes the arrival time of this signal. At the next period $t+1$, firm i's manager will receive a new signal sequence $\{s_{i,u,t+1}\}_{u=0}^{t+1}$ to learn exposure coefficients $\{\beta_{i,u}\}_{u=0}^{t+1}$. Concretely, the signal processes at time t follows:

$$
\begin{aligned}
s_{i,0,t} &= \beta_{i,0} + \epsilon_{0,t} \text{ with } \epsilon_{0,t} \sim N\left(0, \frac{1}{\Delta a_0}\tau_{0,t}^2\right); \\
s_{i,1,t} &= \beta_{i,1} + \epsilon_{1,t} \text{ with } \epsilon_{1,t} \sim N\left(0, \frac{1}{\Delta a_1}\tau_{1,t}^2\right); \\
&\vdots \\
s_{i,u,t} &= \beta_{i,u} + \epsilon_{u,t} \text{ with } \epsilon_{u,t} \sim N\left(0, \frac{1}{\Delta a_u}\tau_{u,t}^2\right); \\
&\vdots \\
s_{i,t,t} &= \beta_{i,t} + \epsilon_{t,t} \text{ with } \epsilon_{t,t} \sim N\left(0, \frac{1}{\Delta a_t}\tau_{t,t}^2\right).
\end{aligned} \tag{12}
$$

In Appendix A.2 , we provide a concrete example and describe the information update process for for typical firm i in generation n. At the micro-level, the parameter σ and the sequence of noise parameters $\{\{\tau_{u,t}\}_{u=0}^{t}\}_{t=0}^{\infty}$ are the primitive parameters of our learning model. The parameter σ captures the dispersion of firms' exposure to the aggregate productivity shocks, which is the source of the reallocation benefit. And signal noise parameters $\{\{\tau_{u,t}\}_{u=0}^{t}\}_{t=0}^{\infty}$ determines how well firms can reallocate resources. Intuitively, lower signal noises imply a higher efficiency in resource allocation.

3.1.3 Capital vintages, information structure, and the aggregation

The main purpose of this paper is to study the implications of learning on the cross-section of firms. We choose a setup that allows us to study the link between capital age and expected returns, while we can avoid keeping track of the distribution of firms with heterogenous information. In particular, we assume that firms can be divided into $\bar{n}$ generations with the generation index n. In our quantitative model, we set $\bar{n} = 5$, corresponding to 5 portfolios sorted by capital age in the empirical section. We use $n = 1$ to denote the youngest generation, and use $n = \bar{n}$ to denote mature firms. Within each generation, there is a continuum of firms indexed with i that produce intermediate inputs, y_i. These outputs can be transformed into the group-level output Y_n using a CES production function in the same fashion as in (1).

The sole distinction across firms in different generations is different levels of information precision. Motivated by the suggestive evidence in Section 2 that captures a negative relation between capital age and capital misallocation, together with Lemma 2, we make a key assumption that more senior generation firms have more information about their exposure coefficients with respect to aggregate productivity. In our model, higher information precision is corresponding to a lower noise of signal, τ^2. In the extreme case, we assume that mature firms, i.e. generation $\bar{n}$ firms, know the exact values of $\{\beta_{i,u}\}_{u=0}^{t}$; that is, mature firms have full information about their exposures.

The interpretation of the learning process of our model is that, when a new project is launched, managers do not exactly know its specific exposure with respect to aggregate productivity. Over time, managers improve their ability to obtain more precise information about the project and make better resource allocation decisions. We do not emphasize the learning of the exposure that is common among similar technologies or similar business models, which can potentially be learned from other firms or past experiences. Instead, we emphasize the imperfect information about firm-specific exposure that can only be learned through its own operation.

The learning parameter at the micro-level $\{\{\tau_{u,t}\}_{u=0}^{t}\}_{t=0}^{\infty}$ is an infinite-dimensional object, and is not directly observable in the data. In the Appendix A.2 , equations (A14) and (A15) show that we can specify the sequence of $\{\{\tau_{u,t}\}_{u=0}^{t}\}_{t=0}^{\infty}$ as functions of parameters λ_n and ρ_n for $n = 1, 2, ..., \bar{n}$, and we further develop a recursive representation of group-level productivity growth as in following Lemma 3:

Lemma 3. *The aggregate output of firm group n is*

$$Y_{n,t} = K_{n,t}^{\alpha}(\mathbf{A_{n,t}} N_{n,t})^{1-\alpha}. \tag{13}$$

The productivity of mature firm group ($n = \bar{n}$) is

$$\mathbf{A_{\bar{n},t}} = \exp\left[\sum_{u=0}^{t} m + \lambda^* \Delta a_u\right]. \tag{14}$$

If the sequence of noise parameters $\{\{\tau_{u,t}\}_{u=0}^{t}\}_{t=0}^{\infty}$ is specified as functions of λ_n and ρ_n for $n = 1, 2, ..., \bar{n}$ as equations (A14) and (A15) in Appendix A.2 , then the ratio between the productivity of young firms ($n < \bar{n}$) and that of the mature firms ($n = \bar{n}$), $\chi_{n,t}$, is stationary and follows the following AR(1) process:

$$\chi_{n,t+1} = \ln \mathbf{A_{\bar{n},t+1}} - \ln \mathbf{A_{n,t+1}} = \rho_n \chi_{n,t} + (\lambda^* - \lambda_n)\Delta a_{t+1}. \tag{15}$$

In addition, the law of motion of the productivity for mature firms ($n = \bar{n}$) follows:

$$\ln \mathbf{A_{\bar{n},t+1}} - \ln \mathbf{A_{\bar{n},t}} = m + \lambda^* \Delta a_{t+1}, \tag{16}$$

and, with $\chi_{n,t}$ as the state variable, the law of motion for the productivity of young firms

($n < \bar{n}$) follows:

$$\ln \mathbf{A_{n,t+1}} - \ln \mathbf{A_{n,t}} = m + (1-\rho_n)\chi_{n,t} + \lambda_n \Delta a_{t+1}. \tag{17}$$

Proof. See Appendix A.2 . □

We make several comments about Lemma 3. First, young firms with imperfect information about their exposures will clearly be less productive than mature firms, on average. In order to guarantee balanced growth, we keep the specification of productivity in equation (11), and allow for perpetual learning; that is, we allow firms to receive new signals about the entire history of their exposure coefficients in every period t, for which the arrival of signals is described in (12). Lemma 3 shows that, thanks to perpetual learning, young firms can eventually obtain full information about their exposures, which rules out permanent productivity gaps between young and mature firms and guarantees balanced growth. In particular, equations (15), (16), and (17) fully specify the aggregate productivity of young and mature firms, that features the balanced growth.

Second, the parameter λ_n characterizes young firms' contemporaneous exposure to common shocks. As we show in equation (A14), λ_n is decreasing in the noise of the signal, τ^2. This is consistent with Lemma 1 that firms with less information precision are less sensitive to aggregate productivity shocks. Based on equation (A15), $1-\rho_n$ can be interpreted as the learning rate about productivity. ρ_n is increasing in the variances of the signals. Intuitively, higher values of τ^2 imply that young firms' information are less precise and, as a result, the productivity gap between young and mature firms can persist for many periods.

Third, in our quantitative analysis, we do not directly specify the micro learning parameters σ and $\{\{\tau_{u,t}\}_{u=0}^{t}\}_{t=0}^{\infty}$. Rather, we calibrate macro parameters, including the contemporaneous exposure λ_n, and the learning rate ρ_n, for each firm generation n. We empirically estimate these parameters from the difference in the exposure of young and old firms with respect to aggregate productivity shocks. As we will detail in Section 4, the empirical evidence shows that $\lambda_{n+1} > \lambda_n$ and $\rho_n > \rho_{n+1}$; that is, younger firms have a lower contemporaneous exposure to common productivity shocks and feature a lower learning rate. The evidence strongly supports our assumption that young firms have less information with respect to their firm-specific exposure to common productivity shocks than mature firms.

In the dynamic setup, generation n firms' log MPK dispersion, a capital misallocation measure, is directly linked to the productivity gap $\chi_{n,t}$, as summarized by Lemma 4 below. This highlights the fact that, in our model, information induced misallocation is responsible for the productivity difference for different capital vintages.

Lemma 4. *In the dynamic setup, the realized log MPK dispersion (cross-sectional variance)* *in generation n follows*

$$Var\left[\log(MPK_i) - \log(MPK)\right] = 2(1-\alpha)(1-\nu)\chi_{n,t}.$$

Proof. See Appendix A.2 . ☐

3.2 The full model

3.2.1 Preferences

Time is discrete and infinite, and indexed by t. In this economy, there is a representative agent with Kreps and Porteus (1978) preferences, as in Epstein and Zin (1989):

$$V_t = \{(1-\beta)C_t^{1-\frac{1}{\psi}} + \beta(E_t[V_{t+1}^{1-\gamma}])^{\frac{1-1/\psi}{1-\gamma}}\}^{\frac{1}{1-1/\psi}} \tag{18}$$

in which C_t denotes the aggregate consumption at time t. For the sake of model parsimony, we do not consider the dis-utility from labor, and hence, the labor supply is inelastic.

3.2.2 Output producers

We specify the stochastic process for the common productivity growth as

$$\begin{aligned} \Delta a_{t+1} &= \mu + x_t + e^{\sigma_a}\varepsilon_{a,t+1} \\ x_{t+1} &= \rho_x x_t + e^{\sigma_x}\varepsilon_{x,t+1} \\ \begin{bmatrix} \varepsilon_{a,t+1} \\ \varepsilon_{x,t+1} \end{bmatrix} &\sim i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \end{aligned} \tag{19}$$

This specification follows Croce (2014) and captures long-run productivity risks.[10] In particular, the common productivity growth has two components: short-run productivity shock $\varepsilon_{a,t+1}$ and long-run shock $\varepsilon_{x,t+1}$. Short-run shocks affect contemporaneous output directly but have no effect on future productivity growth. Long-run shocks do not affect current output but carry news about future productivity growth rates. We set the log standard deviations of both shocks, σ_a and σ_x, to be constant over time.

[10]In the full model, we specify a standard Gaussian long-run risk process for aggregate productivity growth, rather than the process that includes a non-negative Gamma process plus a negative constant term m used in our static/dynamic learning model. We choose to use a more standard productivity process to avoid the concern that the return pattern is driven by the specification of non-negative shocks itself. Based on our unreported results, we have checked that these two alternative appropriately calibrated productivity growth processes make little difference in model's quantitative asset pricing implications.

We assume that there are $\bar{n}$ generations of firms based on their information precision. We denote $\mathbf{A}_t$ and $\mathbf{K}_t$ as the vectors of firm generation-wide productivities and capital stocks; that is, $\mathbf{A}_t = \{\mathbf{A_{n,t}}\}_{n=1}^{\bar{n}}$ and $\mathbf{K}_t = \{K_{n,t}\}_{n=1}^{\bar{n}}$. The aggregate production can be specified as the solution to the following optimal resource allocation problem:

$$
\begin{aligned}
F(\mathbf{A}_t, \mathbf{K}_t) = \max_{N_{1,t}, N_{2,t}, \cdots, N_{\bar{n},t}} \sum_{n=1}^{\bar{n}} K_{n,t}^{\alpha} (\mathbf{A_{n,t}} N_{n,t})^{1-\alpha} \\
\text{subject to } \sum_{n=1}^{\bar{n}} N_{n,t} = 1
\end{aligned}
\tag{20}
$$

Despite featuring substantial heterogeneity across firms, the production side of our model can be summarized by a representative firm with the production function $Y_t = F(\mathbf{A}_t, \mathbf{K}_t)$, for which the law of motion for productivity is characterized by equations (15)-(17), and the dynamics of capital stocks for firm generations are given by equations (23)-(21), which we will discuss in the next subsection.

3.2.3 Firm dynamics

New firms are created by investment. Upon creation, they belong to the youngest generation ($n = 1$). In each period, firms will exit with probability δ. The surviving firms of generation n ($n < \bar{n}$), $(1-\delta)K_{n,t}$, transit to the next generation, $n+1$, with probability ϕ. For the sake of model parsimony, we assume that all firms are subject to the same exit rate and transition rate. Under this assumption, the capital dynamics of the youngest firm generation follow:

$$K_{1,t+1} = (1-\delta)(1-\phi)K_{1,t} + I_t, \tag{21}$$

in which I_t is investment. Equation (21) says that the next period young firm capitals come from the existing young firm capitals that haven't transit to the next generation and new investment. The capital dynamics for middle firm generations ($1 < n < \bar{n}$) follow:

$$K_{n,t+1} = (1-\delta)(1-\phi)K_{n,t} + (1-\delta)\phi K_{n-1,t}, \tag{22}$$

It is noteworthy that, on the right hand side of equation (22), the inflow to the middle firm generation capitals is only the capitals transited from the previous generation. because old capitals can not be directly created through investment. Lastly, the law of motion for the mass of mature firms, $K_{\bar{n}}$, is:

$$K_{\bar{n},t+1} = (1-\delta)K_{\bar{n},t} + (1-\delta)\phi K_{\bar{n}-1,t}, \tag{23}$$

Note that, in our model, there is no adjustment cost of investment. In standard RBC models, we need a high adjustment cost to generate a high equity premium, like in Croce (2014). However, with the vintage capital and learning mechanism, our model is able to achieve a reasonable equity premium even without adjustment cost, which we will come back in the quantitative section 5.

To complete the model, we also have the market clearing condition, i.e., consumption plus investment equals total output:

$$C_t + I_t = Y_t. \tag{24}$$

3.2.4 Equilibrium conditions

As standard welfare theorems apply in our economy, we can construct equilibrium prices and quantities from the solution to a planner's problem. We provide detailed derivations of the equilibrium conditions in appendix B. The stochastic discount factor $\Lambda_{t,t+1}$ can be written as:

$$\Lambda_{t,t+1} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\psi}} \left(\frac{V_{t+1}}{E_t[V_{t+1}^{1-\gamma}]^{\frac{1}{1-\gamma}}}\right)^{\frac{1}{\psi}-\gamma}. \tag{25}$$

Given the equilibrium quantities, we can show that the cum-dividend price of mature firms, $p_{K_{\bar{n}},t}$, satisfies:

$$p_{K_{\bar{n}},t} = \alpha A_{\bar{n},t}^{1-\alpha} \left(\frac{K_{\bar{n},t}}{N_{\bar{n},t}}\right)^{\alpha-1} + (1-\delta)E_t[\Lambda_{t,t+1}p_{K_{\bar{n}},t+1}]. \tag{26}$$

And the cum-dividend price of adolescent firms $n < \bar{n}$, $p_{K_n,t}$ satisfies:

$$p_{K_n,t} = \alpha A_{n,t}^{1-\alpha} \left(\frac{K_{n,t}}{N_{n,t}}\right)^{\alpha-1} + (1-\delta)\{(1-\phi)E_t[\Lambda_{t,t+1}p_{K_n,t+1}] + \phi E_t[\Lambda_{t,t+1}p_{K_{n+1},t+1}]\} \tag{27}$$

Equation (26) implies that the cum-dividends marginal value of mature firms equals the current period marginal product of capital, $A_{\bar{n},t}^{1-\alpha}\left(\frac{K_{\bar{n},t}}{N_{\bar{n},t}}\right)^{\alpha-1}$, and the expected continuation value of future payoffs, $p_{K_{\bar{n}},t+1}$, adjusted for the survival probability $1-\delta$.

According to equation (27), the value of adolescent firms (generation $n < \bar{n}$) is determined by the marginal product of its capital in the current period, $\alpha A_{n,t}^{1-\alpha}\left(\frac{K_{n,t}}{N_{n,t}}\right)^{\alpha-1}$, plus the continuation value of their future payoffs. Conditional on surviving to the next period with probability $1-\delta$, generation n firms become next generation with probability ϕ and pay $p_{K_{n+1},t+1}$ going forward. With a probability of $1-\phi$, they remain in the same generation and pay the continuation value of $p_{K_n,t+1}$.

The optimal investment should satisfy the Euler equation:

$$1 = E_t[\Lambda_{t,t+1} p_{K_1,t+1}] \tag{28}$$

Equation (28) equates the marginal cost of investment on the left hand side, which is equal to 1 due to zero adjustment cost, to the marginal benefit of investment on the right hand side.

3.3 Asset returns

With the equilibrium conditions, we can compute the asset returns for each firm group. We denote $q_{K_n,t}$ as the ex-dividend price of $K_{n,t}$, that satisfies:

$$q_{K_n,t} = E_t[\Lambda_{t,t+1} p_{K_n,t+1}], for\ n = 1, 2, ..., \bar{n}. \tag{29}$$

The return of capital takes the form:

$$R_{K_n,t+1} = \frac{A_{n,t}^{1-\alpha}\left(\frac{K_{n,t}}{N_{n,t}}\right)^{\alpha-1} + (1-\delta)(1-\phi)q_{K_n,t+1} + (1-\delta)\phi q_{K_{n+1},t+1}}{q_{K_n,t}},$$

$$R_{K_{\bar{n}},t+1} = \frac{\alpha A_{\bar{n},t}^{1-\alpha}\left(\frac{K_{\bar{n},t}}{N_{\bar{n},t}}\right)^{\alpha-1} + (1-\delta)q_{K_{\bar{n}},t+1}}{q_{K_{\bar{n}},t}}.$$

The key mechanism that generates the return spread between old versus young capital is the difference in the marginal production of capital's exposure to the common productivity shock. As we discussed previously, mature firms with more information can take better advantage of aggregate technological progress, and their productivity features a higher exposure to aggregate shocks. Therefore, old firms' marginal products of capital, $\alpha A_{\bar{n},t}^{1-\alpha}\left(\frac{K_{\bar{n},t}}{N_{\bar{n},t}}\right)^{\alpha-1}$, are more sensitive to common productivity shocks, and thus earn higher expected return.

The market return can be computed as a weighted average of the returns on different capital vintages in this economy:

$$R_{m,t+1} = \sum_{n=1}^{\bar{n}} \frac{q_{K_n,t} K_{n,t}}{\sum_{n=1}^{\bar{n}} q_{K_n,t} K_{n,t}} R_{K_n,t+1}$$

3.4 Cash flow duration

Parameter learning also affects the timing of cash flow, which is measured by cash flow duration. Mature firms have already achieved full information and will not increase their productivity through learning in the future, so their cash flow is evenly distributed across time. In contrast, young firms have low cash flow in the beginning due to less efficient resource allocation, but their cash flow will increase in the future as they transit into next generation when they have more precise information and can better take advantage of the technological progress. Thus, young firms have higher cash flow duration than mature firms.

In our model, the Macaulay duration $MD_{\bar{n}}$ of the mature generation $K_{\bar{n}}$ is defined as:

$$MD_{\bar{n},t}q_{K_{\bar{n}},t} = E_t\left[\sum_{j=1}^{\infty} j\Lambda_{t,t+j}D_{\bar{n},t+j}\right].$$

We can rewrite the duration formula recursively as:

$$\begin{aligned} MD_{\bar{n},t}q_{K_{\bar{n}},t} &= E_t[\Lambda_{t,t+1}D_{\bar{n},t+1}] + E_t\left[E_{t+1}\left[\sum_{j=2}^{\infty} j\Lambda_{t,t+j}D_{\bar{n},t+j}\right]\right], \\ &= E_t[\Lambda_{t,t+1}(D_{\bar{n},t+1} + (1-\delta)q_{K_{\bar{n}},t+1} + (1-\delta)MD_{\bar{n},t+1}q_{K_{\bar{n}},t+1})]. \end{aligned}$$

Similarly, for $n < \bar{n}$, the Macaulay duration MD_n can be written as:

$$MD_{n,t}q_{K_n,t} = E_t[\Lambda_{t,t+1}(D_{n,t+1}+(1-\delta)((1-\phi)(MD_{n,t+1}+1)q_{K_n,t+1}+\phi(MD_{n+1,t+1}+1)q_{K_{n+1},t+1}))].$$

4 Empirical evidence on the learning mechanism

In our quantitative analysis, we do not directly specify the micro parameters σ and $\{\{\tau_{u,t}\}_{u=0}^{t}\}_{t=0}^{\infty}$. Rather, we calibrate macro parameters, that is, the contemporaneous exposure λ_n and the learning rate ρ_n, for each firm generation n. In this section, we provide the empirical procedure we use to estimate these parameters from the difference in the exposure of young and old firms with respect to aggregate productivity shocks. The empirical evidence not only directly supports the learning mechanism that we propose, but also helps us to identify the key learning parameters for a quantitative study of the model.

4.1 Firms' exposure to aggregate shocks

To identify the contemporaneous exposure λ_n, we conduct a two-step empirical procedure. First, we estimate the firm-level productivity of public traded companies on U.S. stock exchange, following Ai, Croce, and Li (2013). Please refer to Appendix D.5 for estimation details. Second, we estimate the exposure of firms' productivity with respect to aggregate productivity by different capital age groups ($n = 1, 2, ..., \bar{n}$) by using the following regression:

$$\Delta \ln A_{i,j,t} = \xi_{0,i} + \xi_n \Delta \ln A_t + X_{i,j,t} + \varepsilon_{i,j,t}, \text{ for firms in age group n}$$

in which $A_{i,j,t}$ denotes the productivity of firm i in industry j at time t, and $\xi_{0,i}$ controls for the firm-specific fixed effect, and $\Delta \ln A_t$ is the growth rate of aggregate productivity as measured by the U.S. Bureau of Labor Statistics (BLS). $X_{i,j,t}$ includes a list of control variables for firms' fundamentals.

The key coefficient of our interest, ξ_n captures a capital age group specific exposure with respect to aggregate productivity shocks. Based on Lemma 1, our learning model predicts the exposure ξ_n increases with capital age n. In Table 3 Panel A, we show the regression result within different capital age groups. The productivity exposure increases with capital age, which strongly supports our learning mechanism.

[Place Table 3 about here]

To directly map data to our model parameters, we normalize each group-specific exposure by the whole sample exposure. The model counter-part is the generation-specific exposure that is normalized by the steady state capital share weighted exposure $\frac{\lambda_n}{(\sum_{n=1}^{\bar{n}} K_{n,ss}\lambda_n)/K_{ss}}$.

For the sake of parsimony, we assume λ_n follows a exponentially increasing pattern from young to mature generation: $\lambda_n = \lambda^{n-1}, n = 1, 2, ..., \bar{n}$. In the model, we normalize the exposure of the youngest group to be 1, and thus $\lambda_1 = 1$. We also calibrate $\lambda = 1.18$ such that the normalized exposure implied by the model broadly matches the pattern in the data. In Panel A of Table 3, we shows that the normalized exposure in the model and data are closely aligned.

4.2 Estimation of the learning rate

In our model, the learning rate parameter ρ_n is the persistence of the co-integration residual $\chi_{n,t}$. Thus, we identify these parameters through the autocorrelation of the log productivity differences between the old capital age group $\bar{n}$ and the young age group $n < \bar{n}$. Specifically, we define:

$$\chi_{n,t} = \ln \mathbf{A}_{\bar{\mathbf{n}},\mathbf{t}} - \ln \mathbf{A}_{\mathbf{n},\mathbf{t}}, \ n = 1, ..., \bar{n} - 1,$$

in which n indicates the capital age sorted group, and $\mathbf{A}_{\mathbf{n},\mathbf{t}}$ is empirically measured by the value weighted average productivity of firms in group n. For each n, we estimate the autocorrelation ρ_n by running a AR(1) regression of $\chi_{n,t}$.

Since $1 - \rho_n$ is the learning rate on exposure in the past, our learning model predicts that mature generations learn faster due to their information advantage, thus it implies ρ_n to decrease from the young to mature generations. In Panel B of Table 3, we report the estimated autocorrelation of capital age groups 1-4. Indeed, there is a decreasing pattern from young to old firms, consistent with our model prediction. This directly supports the learning mechanism of our model. Again, for parsimony's sake, we assume that ρ_n follows a exponentially decay pattern from a young to mature generation: $\rho_n = \rho^n$ (in the model, the oldest generation has perfect information on its exposure, and thus thus, $\rho_{\bar{n}} = 0$). We calibrate the quarterly learning parameter $\rho = 0.96$ such that the annual autocorrelation implied by the model matches the pattern in the data.

5 Quantitative model implications

In this section, we calibrate our model at the quarterly frequency and evaluate its ability to replicate key moments of both macroeconomic quantities and asset prices at the aggregate level. More importantly, we investigate its performance in terms of quantitatively accounting for capital age premium in the cross-section. For macroeconomic quantities, we focus on a long sample of U.S. annual data from 1930 to 2016. All macroeconomic variables are real and per capita. We obtain consumption, output, and physical investment data from the Bureau of Economic Analysis (BEA). For the purpose of cross-sectional analyses, we make use of several data sources at the micro-level, including (1) firm-level balance sheet data from the CRSP/Compustat Merged Fundamentals Annual Files, and (2) monthly stock returns from CRSP. In Appendix D, we provide more details on our data sources and constructions.

5.1 Calibration

We calibrate our model at the quarterly frequency and present the parameters in Table 4. We choose the relative risk aversion $\gamma = 10$ and the intertemporal elasticity of substitution $\psi = 2$, which is consistent with the long-run risks literature, as in Bansal and Yaron (2004). We set the discount factor $\beta = 0.997$ to match the level of risk-free rate. The capital share parameter $\alpha = 0.3$ and the quarterly depreciate rate of physical capital $\delta_k = 0.03$ are consistent with the standard real business cycles literature (e.g., Kydland and Prescott (1982)).

[Place Table 4 about here]

Our calibration of the parameters of the aggregate productivity shocks is standard in the long-run productivity risk literature. We calibrate μ and σ_a to match the mean and the volatility, respectively, of output growth in the U.S. economy in our sample period, 1930-2016. We set relative volatility at $\exp(\sigma_x - \sigma_a) = 0.12$ and the autocorrelation of long-run risk at $\rho_x = 0.946$, following Croce (2014).

In our model, there are three parameters intimately related to our key learning mechanism. The parameter ϕ is the rate of transition to the next capital age group, and the parameter λ^* governs the exposure of mature firms to contemporaneous aggregate shocks; in addition, the persistence of the cointegration residual ρ governs the speed of learning for young firms with imprecise information about their exposure to common productivity shocks. We choose the parameter ϕ to broadly match the steady state distribution of capital age and capital share. As we show in Appendix C, given the capital depreciation rate δ, both the capital share and average capital age across different capital age groups are functions of the transition rate parameter ϕ. As we show in Table 5, our calibration of $\phi = 0.08$ generates the capital age and capital share profile across age groups to be broadly consistent with the data. In Section 4, we provided supportive evidence and calibration details about the other two learning parameters λ^* and ρ.

[**Place Table 5 about here**]

In addition to our benchmark calibration, we also calibrate a RBC model with adjustment costs and report the result for the sake of comparison. When calibrating the RBC model, we retain the same parameter expect for three modifications. We keep 5 capital age groups but eliminate the learning channel ($\lambda_n = 1$ and $\rho_n = 0$). By doing so, we ensure that all firms are identical and have perfect information. We also adjust the volatility of short-run shocks to 1.69% ($(1-\alpha)\exp(\sigma_a) = 1.69\%$) to match the volatility of total output in the data. We increase the subjective discount factor to 0.998 to match the level of risk-free rate. Lastly, we add the adjustment cost of investment. Without the learning mechanism, adjustment cost is necessary to generate equity premium in standard RBC models, as shown in Croce (2014). We model the adjustment cost following Jermann (1998):

$$G(I, K) = K[\alpha_0 + \frac{\alpha_1}{1 - 1/\xi}(\frac{I}{K})^{1-1/\xi}].$$

$\{\alpha_0, \alpha_1\}$ are set such that in steady state $G = I$ and $G_I = 1$. We set the adjustment cost parameter $\xi = 2.3$ to obtain a same equity premium as in our benchmark model.

5.2 Aggregate moments

We now turn to the quantitative performance of the model at the aggregate level. We solve and simulate our model at the quarterly frequency and aggregate the model-generated data to compute annual moments.[11] We show that our model is broadly consistent with the key empirical features of macroeconomic quantities and asset prices.

[Place Table 6 about here]

In Table 6, we report the key moments of macroeconomic quantities (top panel) and those of asset returns (bottom panel) respectively, and compare them to their counterparts in the data when possible.

In terms of aggregate moments on macro quantities (top panel), our model has similar implications to those standard RBC models. In particular, our calibration features a low volatility of consumption growth (2.73%), and matches the mean, volatility, and autocorrelations of output and consumption growth with the data reasonably well. Due to the absence of adjustment costs, the volatility of investment in our benchmark calibration (6.41%) is reasonably high compared with the RBC model.

Turning our attention to asset pricing moments (bottom panel), our model produces a low risk-free rate (0.89%) and a high equity premium (3.76%) with a leverage ratio of 2, comparable to key empirical moments for aggregate stock markets. It is worth noting that our model can generate a high equity premium even without adjustment costs. Because new capital does not fully enjoy productivity growth, the consumption-smoothing effect of investment is mitigated. The RBC model with high adjustment cost can also generate a comparable equity premium, but with a lower investment volatility (5.04%). Our benchmark model also produces a sizable capital age premium as in the data. While in RBC model, because there is no information difference between old capital firms and young capital firms, it fails to generate any capital age spread. As in standard production based asset pricing models, without explicitly modeling labor market leverage, our model has difficulty in generating a significantly large stock market volatility. One potential resolution is to further incorporate the labor market frictions into the model framework, following Favilukis and Lin (2015).

[11] We solve the model by using a second-order local approximation around the steady state using the `Dynare` package.

5.3 Impulse response functions

To better understand the implications of the learning mechanism in our model, we plot the impulse response functions with respect to one-standard-deviation of short-run and long-run productivity shocks in Figure 1 and Figure 2, respectively. The impulse response of our benchmark model is plotted with a solid black line, and the impulse response of RBC model is plotted with a red dashed line.

Figure 1. Impulse Response Functions for Short-Run Shock

This figure shows percentage deviation from the steady state upon the realization of a positive short-run shock. The solid black line is the impulse response function of our benchmark capital model, and the dashed red line is the RBC model. Both the benchmark model and RBC model feature short- and long-run productivity shocks. The benchmark capital model features limited information and learning, but does not have adjustment cost. The RBC model has convex adjustment cost. Returns are not levered.

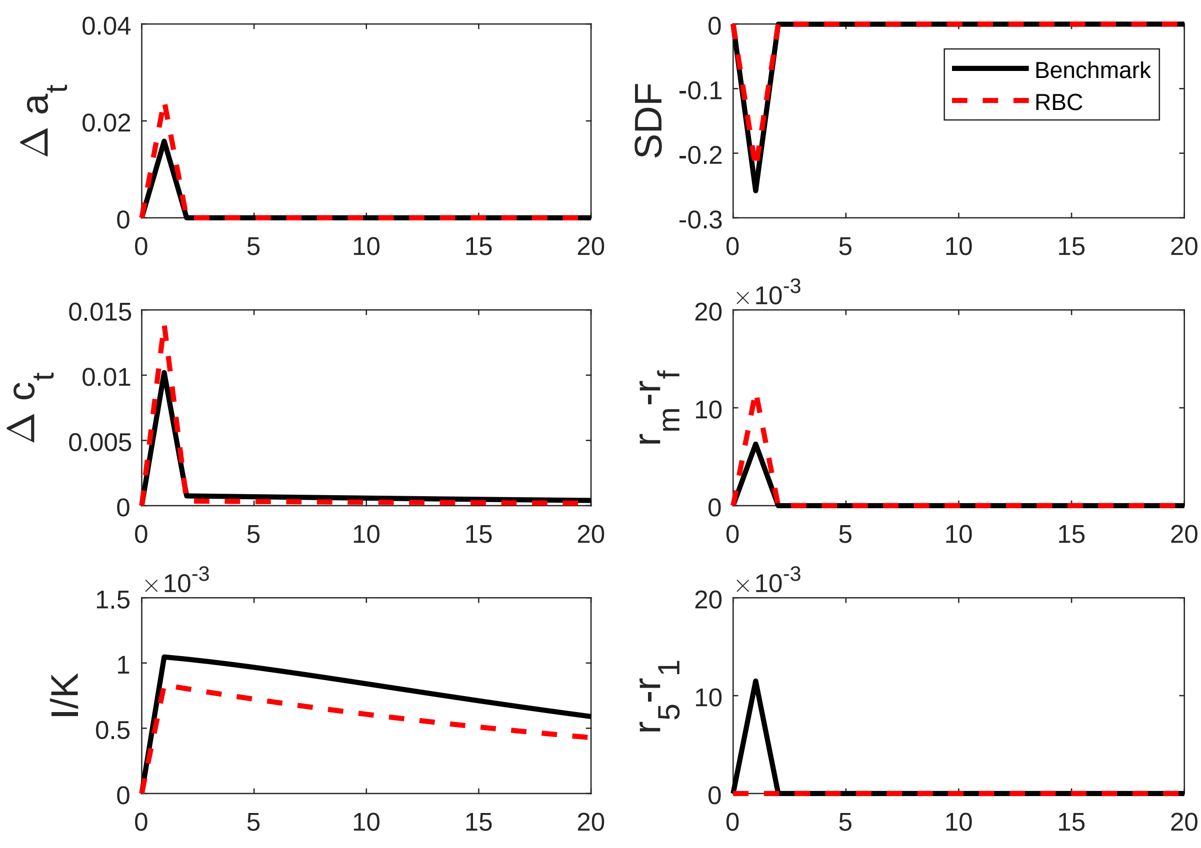
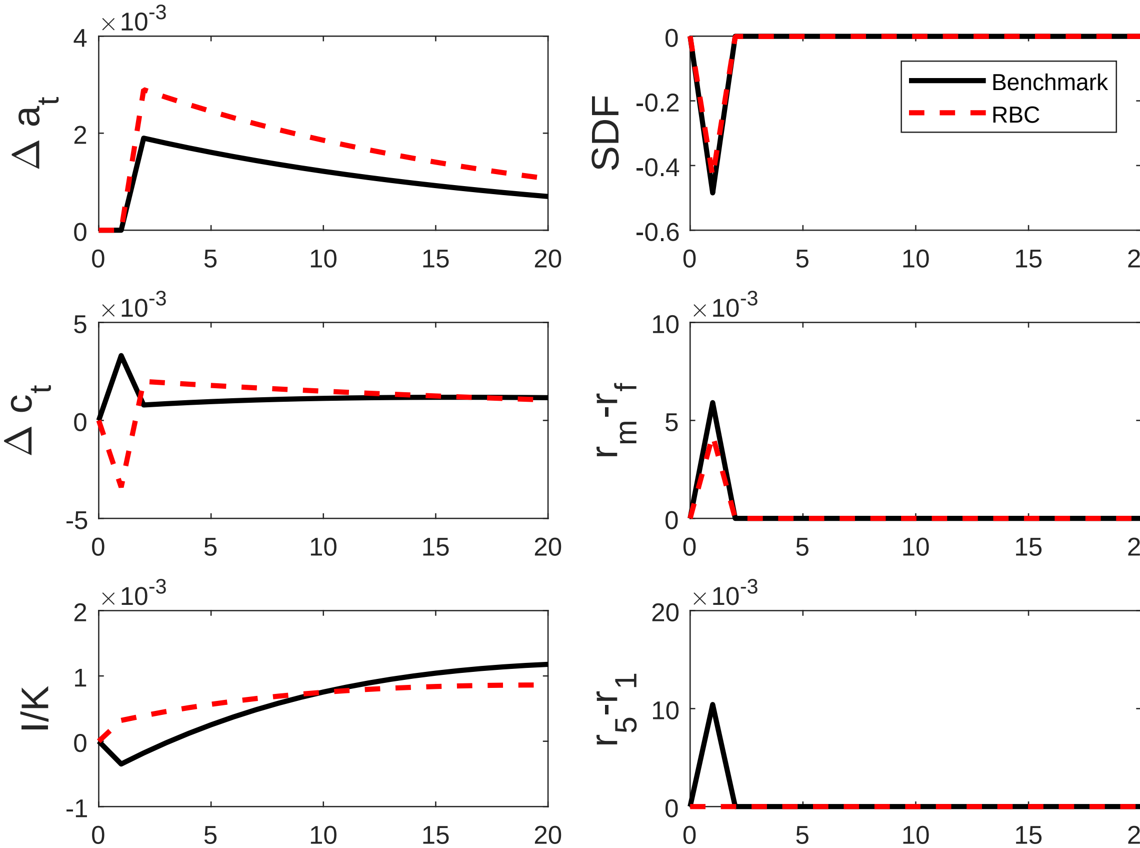


Figure 2. Impulse Response Functions for Long-Run Shock

This figure shows percentage deviation from the steady state upon the realization of a positive long-run shock. The solid black line is the impulse response function of our benchmark capital model, and the dashed red line is the RBC model. Both the benchmark model and RBC model feature short- and long-run productivity shocks. The benchmark capital model features limited information and learning, but does not have adjustment cost. The RBC model has convex adjustment cost. Returns are not levered.



The response of consumption growth and investments to shock-run shocks are similar in both the benchmark model and RBC model. Because macro quantities in the learning model are mostly driven by short-run shocks, our benchmark calibration inherits the success of standard RBC models. A notable difference is that, in the absence of adjustment cost, investment in our benchmark model responds more strongly to short-run shocks than that in the RBC model. As a result, our learning model can generate a higher investment volatility.

The risk premia associated with short-run shocks are small in both models. Therefore, to understand the impact of learning mechanism on asset prices, we need to focus on the long-run shocks. We make two observations from the impulse response functions with respect to long-run shocks.

First, we observe that the impulse responses of investment and consumption to long-run shocks are significantly different in two models. In the RBC model, investment responds positively to news shocks. With an IES of 2, upon the arrival of positive news about future productivity shocks, the substitution effect dominates. Investment rises and, since the productivity news has not materialized, consumption drops temporarily.

In stark contrast, in our learning model, investment responds negatively to positive news shocks. Over time, as news about future productivity materializes, investment gradually goes up. Intuitively, a positive news shock does not increase current period productivity, and its effect realizes slowly over time. On the one hand, the substitution effect is moderate. New investment only builds young capital firms, which cannot take full advantage of the rise in productivity. Because there is no adjustment cost, households will find it optimal to invest later when young capital firms gradually adopt the productivity growth. On the other hand, the income effect is strong because old capital firms immediately benefit from positive productivity shocks. As a result, investment temporarily drops and consumption increases.

Second, the excess return on mature firms and young firms response differently to productivity shocks. In our benchmark model, the response of return on mature firms to productivity shock is significantly larger than that of young firms, which helps to produce a capital age spread. When a positive productivity shock arrives, mature firms know how to allocate their resources efficiently and better utilize the rise in productivity. However, due to resource misallocation, young firms benefit little from the productivity growth. Mature firms' marginal product of capital is more sensitive to common shocks when compared with those of young firms. Therefore, mature firms' returns also responds more to shocks than those of young firms. In the RBC model, because all firms are identical, the return spread is always 0.

This pattern of impulse response has strong implications for the equity premium. In our benchmark model, since there is no adjustment cost, the marginal q of young firms always equals 1. On the other hand, because we can not directly create old capital firms, the marginal q of old capital firms responds strongly and positively to news, which allows our learning model to generate a sizable equity premium without investment frictions. In the RBC model, the positive response of marginal q is achieved by assuming a high adjustment cost, which generates a counterfactually low level of the investment volatility.

5.4 Cross-sectional implications

In this section, we study the capital age spread at the cross-sectional level. In Table 7, we report the average excess returns and cash flow durations across different age groups from the model, and then compare them with the data.

We make several observations. First, we use Panel A of Table 7 to report the average excess return of each capital age group in the data and model. We use a leverage ratio of 2 to compute the leverage equity return. Our model can generate a capital age spread as large as 6.65%, which is comparable to the data. The key mechanism to generate the return spread is as follows: mature firms have better information about their exposure to common shocks, so they demonstrate better allocation of resources. This has an amplification effect that makes their marginal product of capital more exposed to common productivity shocks.

[Place Table 7 about here]

Second, In Panel B of Table 7, we show the cash flow duration of each capital age group in the data and model. Appendix D.4 provides detailed construction of the cash flow duration.[12] In the data, young capital firms have higher cash flow duration than those of old capital firms. Our model generates the same monotonic pattern. Intuitively, firms in the young generation have lower average productivity, so they also have lower dividend payouts. As they grow into mature firms, they acquire better information and demonstrate better resource allocations; hence, their productivity and dividend payouts increase. Therefore, young firms have low cash flows at the short end and high cash flows at the far end. In contrast, the cash flow for mature firms is evenly distributed.

6 Empirical analysis

In this section, we first conduct test testable implications of our model in a two-step procedure. In the first step, we document that the exposure of firms' payout with respect to both the long-run and short-run productivity shocks increases in firms' capital age, providing direct empirical evidence to support our learning mechanism. In the second step, we examine the empirical evidence that differentiates our explanation from other alternative economic channels for the capital age premium, in particular, the technology frontier risk channel according to Lin, Palazzo, and Yang (2019).

[12] In our empirical calculation, the average cash flow duration over the whole sample is largely comparable with the mean cash flow duration in Weber (2018), although the latter paper did not calculate the duration measure across capital age groups.

Then, we provide additional empirical evidence for the positive relation between capital age and the cross-section of stock returns. We perform a battery of asset pricing factor tests to show that such a relation is largely unaffected by known return factors for other systematic risks. We then investigate the joint link between capital age and other firm-level characteristics on one hand and future stock returns in the cross-section on the other using Fama and MacBeth (1973) regressions as a valid cross-check for the positive relation between capital age and stock returns.

6.1 Testable model implications

6.1.1 Productivity shocks and payouts

The key implication of the learning mechanism in our model is that old capital firms with more precise information take advantage of technology growth; therefore, their payouts are more exposed to aggregate productivity shocks. In this subsection, we directly test this model prediction. We show the supporting evidence of positive relation between capital age and firms' payout exposure to aggregate short-run and long-run productivity shocks.

We proceed as follows. First, we measure firms' payout at the quarterly frequency by using their operating income before depreciation ($xintq$), net of interest expenses ($txtq$), income taxes ($oibdpq$), and common stock dividends ($dvy-dvpq$), following Croce, Marchuk, and Schlag (2018).

In the second step, we estimate exposures by regressing firm i's payout to lagged sales ratio with respect to short- and long-run productivity shocks and other control variables as follows:

$$Z_{i,t} = \beta_{0,i} + \beta_{srr}\varepsilon_{a,t} + \beta_{lrr}\varepsilon_{x,t} + \rho Z_{i,t-1} + \beta_x x_{t-1} + Controls_{i,t-1} + resid_t, \qquad (30)$$

in which we follow Ai, Croce, Diercks, and Li (2018) to construct short-run and long-run shocks (i.e., $\varepsilon_{a,t}$ and $\varepsilon_{x,t}$, respectively). Specifically, we project TFP growth on predictors proposed by Bansal and Shaliastovich (2013) plus the integrated volatility of stock market returns to identify the long-run growth component and disentangle it from short-run TFP shocks.

In the model, a linear approximation of the equilibrium dividend processes suggests the dependence of payout on both contemporaneously short- and long- productivity shocks and predetermined variables. For the sake of parsimony, we use the lagged values of the payout ratio to capture the role of the endogenous state variables (i.e., capital shocks), so we may avoid additional measurement errors. Under the null of the model, this is an innocuous assumption.

We report our main findings in Table 8. We observe an upward sloping pattern on coefficients for both short- and long-run productivity shocks from young to old capital age portfolios. That is, firms' payouts in the highest quintile portfolio face significantly higher exposure to short-run and long-run productivity shocks than do those in the lowest quintile.

In summary, payout exposures present an upward sloping pattern with respect to both short- and long-run productivity shocks, which is perfectly consistent with our model implication.

[**Place Table 8 about here**]

6.1.2 Alternative explanations

In this paper, we propose a learning mechanism to account for the capital age premium. The key driving force in our learning mechanism relies on the cross-sectional difference in information precision with respect to the firm specific exposure to aggregate TFP shocks. On the other hand, Lin, Palazzo, and Yang (2019) propose an alternative risk-based story, in which the driving source is the technology frontier shock (TFS). Old capital firms face higher exposure to TFS than young capital firms, and thus carry higher returns as risk compensation.

In order to differentiate our mechanism from that in Lin, Palazzo, and Yang (2019), we identify heterogenous exposure to TFS across Fama-French 30 industries and show that the capital age premium remains significant for firms in industries with low exposure to TFS.

We compute cross-correlations to TFS across industries in several steps. First, for each industry and each quarter, we compute the ± 4-quarter correlation between industry-level output (sales) growth and TFS using 20-quarter rolling windows. Second, we assign the maximum absolute ± 4-quarter cross-correlations to the corresponding industry to indicate TFS exposure. This procedure generates 30 industry-level indicators. Based on these indicators, we classify industries into high, medium, and low technology exposure groups, respectively. We report five portfolios sorted on capital age relative to their industry peers for industries with respect to high exposure in Panel A, medium exposure in Panel B, and low exposure in Panel C. To assure the capital age-return relation, we form a long-short portfolio that takes a long position in the highest quintile and a short position in the lowest quintile portfolio in each panel.

In Table 9, we report the average annualized excess returns in five quintile portfolios and long-short portfolio across three panels. The capital age spread in industries with low exposure to TFS attracts our attention. We show that the capital age premium (OMY) remains statistically significant for industries with low exposure to TFS (Panel C). On the other hand, the capital age spread is significant for industries with high exposure to TFS (Panel A) and larger than the spread in industries with low exposure, which is consistent with Lin, Palazzo, and Yang (2019). However, the statistical significance of the OMY portfolio in Panel C implies that the technology adoption mechanism alone may not fully account for the capital age premium. To fill this gap, we focus on a learning mechanism of firm-specific productivity exposure to explain for the capital age premium. This empirical evidence suggests that both channels co-exist and contribute to the capital age premium. In addition, it is worthwhile to note that our learning mechanism in this paper also coherently derives the negative relationships between capital age, capital misallocation, and cash flow duration simultaneously, as we show previously.

[Place Table 9 about here]

6.2 Asset pricing test

To examine the capital age-return relation, we form a long-short, old-minus-high (OMY), portfolio that takes a long position in the highest quintile and a short position in the lowest quintile portfolio sorted on capital age. We investigate the extent to which the variation in average returns of portfolio sorted on capital age can be explained by existing risk factors. Specifically, we explore whether the capital age-return relation reported in Table 10 reflects firms' exposure to existing systematic risk factors by preforming time-series regressions of portfolio excess returns sorted by capital age on the Fama and French (2015) five-factor model and on the Hou, Xue, and Zhang (2015) q-factor model, respectively.[13] Such time-series regressions enable us to estimate the betas (i.e., risk exposures) of each portfolio's excess return on various risk factors and to estimate each portfolio's risk-adjusted return (i.e., alphas in %).

[Place Table 10 about here]

We make several observations as follows. First, the variation in average returns of portfolio sorted on capital age cannot be captured by asset pricing factor models. As we show in Table 10, the risk-adjusted return (intercepts) of the old-minus-young (OMY) sorted on capital age remains large and significant, ranging from 4.63% for the Fama and French (2015) five-factor model to 3.31% from the Hou, Xue, and Zhang (2015) q-factor model, and these

[13]Data on the Fama-French five factors are from Kenneth French's website. We thank Kewei Hou, Chen Xue, and Lu Zhang for sharing the q-factor returns.

intercepts are 2.93 and 1.80 standard errors away from zero, which implies the 1% and 10% statistical significant level, respectively.

Second, the old-minus-young (OMY) portfolio presents a negative loading on the size factor in both panels, which implies that old capital firms are larger in their size. Such a positive size-age relationship is consistent with empirical findings documented in the literature.

Third, the long-short portfolio shows a positive coefficient with respect to the profitability factor. Such an empirical pattern is alighted with our model. Our interpretation is as follows. Given more precise information about their exposure to productivity shocks, old capital firms tend to take advantage of aggregate technology growth. Therefore, the profitability of old capital firms is higher than that of young capital firms.

Lastly but not least, we report a positive loading with respect to the investment factor on the long-short portfolio in both panels. In addition, loadings present an upward sloping pattern across portfolios sorted on capital age. This finding directly responds to the definition of capital age in Section D.2 , in which firms with lower capital age intensively investment in young capital vintage. However, the capital age spread is not subsumed by the investment factor, which suggests that capital age contains additional information not included in the investment factor.

In summary, results from asset pricing tests in Table 10 show that the cross-sectional return spread across portfolios sorted on capital age cannot be completely captured by either the Fama and French (2015) five-factor or the HXZ q-factor model (Hou, Xue, and Zhang (2015)). In the following subsection, we reassure the capital age-return relation by running Fama-Macbeth regressions to control a bundle of firm characteristics.

6.3 Firm-level return predictability regressions

For robustness, we extend our empirical analysis by providing further evidence for the positive relation between capital age and cross-section of stock returns. We first show that capital age positively predicts the cross-sectional expected stock returns, which is consistent with the result in the univariate portfolio sort on capital age. This analysis allows us to control for an extensive list of firm characteristics that predict stock returns and to verify whether the positive capital age-return relation is driven by other known predictors at the firm level. This approach has the advantage over the portfolio sort as the latter approach requires the specification of breaking points to sort the firms into portfolios and the selection of the number of portfolios. Also, it is difficult to include multiple sorting variables with unique information about future stock returns using a portfolio approach. Thus, Fama-MacBeth cross-sectional regressions provide a valid cross-check.

We run standard firm-level Fama-MacBeth cross-sectional regressions to predict stock returns using the lagged firm-level capital age after controlling for other characteristics. The specification of regression is as follows:

$$R_{i,t+1} - R_{f,t+1} \quad = \quad a + b \times Capital\ Age_{i,t} + c \times Controls_{i,t} + \varepsilon_{it}. \tag{31}$$

Following Fama and French (1992), we regress monthly excess returns of individual stock (annualized by multiplying 12) on capital age with different sets of variables that are known for 6 months prior portfolio formation, and industry fixed effects. Control variables include the natural logarithm of market capitalization (Size), the natural logarithm of book-to-market ratio (B/M), the investment rate (I/A), and the return on equity (ROE), and industry dummies based on Fama and French (1997) 30 industry classifications. All independent variables are normalized to zero mean and one standard deviation after winsorization at the 1th and 99th percentile to reduce the impact of outliers, and adjusted for standard errors by using Newey-West adjustment.

[Place Table 11 about here]

In Table 11, we the results from cross-sectional regressions performed at a monthly frequency. The reported coefficient is the average slope from monthly regressions and the corresponding t-statistics is the average slope divided by its time-series standard error. Findings in Table 11 are consistent with the results of portfolio sorts on capital age. In the first specification in Table 11, we regress stock returns on capital age and industry dummies without controlling for other variables. The estimated coefficient on capital age is statistically significant. Our interpretation is as follows. Capital age strongly predicts stock returns with a slope coefficient of 2.59, which is 5.57 standard deviations far away from zero. It suggests that a one-standard-deviation increases in (log) capital age leads to a significant increase of 2.59% in the average equity return per annum. The difference in average (log) capital age between firms in the highest and lowest quintile portfolio sorted by capital age, respectively, is equivalent to 2.23 standard deviations. The coefficient we obtain in the Specification (1) implies a difference in expected returns of 5.78%, which is equivalent roughly to 5.79%, the value of old-minus-young (OMY) portfolio reported in Table 2.

From specification 2 to 5, capital age significantly predicts for stock returns when we control for the size, book-to-market ratio, investment rate, and profitability one by one. The estimated coefficient on capital age remains statistically significant, ranging from 1.65 (Specification 3) to 3.83 (Specification 5), and is at least more than three standard deviation away from zero. When we put all control variables together in Specification 6, the slope of coefficient on capital age is still significantly positive. On top of that, the result from Specification 6 highlights that the predictability of capital age is not subsumed by known predictors for stock returns in the literature.

7 Conclusion

In this paper, we study the role of parameter learning on the cross-section of stock returns through the lens of a general equilibrium asset pricing models. There is a large class of models that link production and investment to the cross-section of expected returns, which assumes market participants directly observe firm-specific productivity and can also distinguish their systematic components from their idiosyncratic components. In this paper, we propose an alternative but more realistic imperfect information structure: that individual firms have imperfect information about their productivity and face a signal extraction problem. In particular, we assume that young capital vintage experiences less precise signals about their exposure when compared with old capital vintage.

We demonstrate that this parameter learning channel is highly useful for studying the joint link between capital age, reallocation efficiency, the timing of cash flow, and, most importantly, expected returns in the cross-section. Our model framework provides an unified explanation of a wide set of empirical facts in the cross-section: (1) a negative relation between capital age and capital misallocation, (2) a negative link between capital age and cash flow durations, and, most importantly, (3) a positive capital age and expected return relation, which we call the capital age premium. In contrast, a standard asset pricing model with full information generates flat relations between capital age and these aforementioned important economic variables. Our findings suggest that the imperfect information and parameter learning can have a significant impact on asset prices.

References

Ai, Hengjie, 2010, Information quality and long-run risk: Asset pricing implications, *The Journal of Finance* 65, 1333–1367.

Ai, Hengjie, Mariano Massimiliano Croce, and Kai Li, 2013, Toward a quantitative general equilibrium asset pricing model with intangible capital, *The Review of Financial Studies* 26, 491–530.

Ai, Hengjie, Mariano Max Croce, Anthony M. Diercks, and Kai Li, 2018, News shocks and the production-based term structure of equity returns, *The Review of Financial Studies* hhy015.

Bansal, Ravi, and Ivan Shaliastovich, 2013, A long-run risks explanation of predictability puzzles in bond and currency markets, *The Review of Financial Studies* 26, 1–33.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the Long Run: A Potential Resolution, *Journal of Finance* 59, 1481–1509.

Belo, Frederico, Jun Li, Xiaoji Lin, and Xiaofei Zhao, 2017, Labor-force heterogeneity and asset prices: The importance of skilled labor, *The Review of Financial Studies* 30, 3669–3709.

Belo, Frederico, and Xiaoji Lin, 2012, The inventory growth spread, *The Review of Financial Studies* 25, 278–313.

Belo, Frederico, Xiaoji Lin, and Santiago Bazdresch, 2014a, Labor hiring, investment, and stock return predictability in the cross section, *Journal of Political Economy* 122, 129–177.

Belo, Frederico, Xiaoji Lin, and Maria Ana Vitorino, 2014b, Brand capital and firm value, *Review of Economic Dynamics* 17, 150–169.

Belo, Frederico, Xiaoji Lin, and Fan Yang, 2018, External Equity Financing Shocks, Financial Flows, and Asset Prices, *The Review of Financial Studies* .

Blundell, Richard, and Stephen Bond, 2000, Gmm estimation with persistent panel data: an application to production functions, *Econometric reviews* 19, 321–340.

Brown, James R, Steven M Fazzari, and Bruce C Petersen, 2009, Financing innovation and growth: Cash flow, external equity, and the 1990s r&d boom, *The Journal of Finance* 64, 151–185.

Chan, Louis K C, Josef Lakonishok, and Theodore Sougiannis, 2001, The Stock Market Valuation of Research, *Journal of Finance* 56, 2431 – 2456.

Chen, Kaiji, and Zheng Song, 2013, Financial frictions on capital allocation: A transmission mechanism of tfp fluctuations, *Journal of Monetary Economics* 60, 683–703.

Cochrane, John H, 1991, Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations, *The Journal of Finance* 46, 209–237.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A Lochstoer, 2016, Parameter learning in general equilibrium: The asset pricing implications, *American Economic Review* 106, 664–98.

Croce, Mariano M, Martin Lettau, and Sydney C Ludvigson, 2014, Investor information, long-run risk, and the term structure of equity, *The Review of Financial Studies* 28, 706–742.

Croce, Mariano Massimiliano, 2014, Long-run productivity risk: A new hope for production-based asset pricing?, *Journal of Monetary Economics* 66, 13–31.

Croce, Mariano Max Massimiliano, Tatyana Marchuk, and Christian Schlag, 2018, The leading premium, *Working Paper* .

David, Alexander, 1997, Fluctuating confidence in stock markets: Implications for returns and volatility, *Journal of Financial and Quantitative Analysis* 32, 427–462.

David, Joel M, Hugo A Hopenhayn, and Venky Venkateswaran, 2016, Information, misallocation, and aggregate productivity, *The Quarterly Journal of Economics* 131, 943–1005.

David, Joel M, Lukas Schmid, and David Zeke, 2018, Risk-adjusted capital allocation and misallocation, *Working Paper* .

Dechow, Patricia M, Richard G Sloan, and Mark T Soliman, 2004, Implied equity duration: A new measure of equity risk, *Review of Accounting Studies* 9, 197–228.

Dunne, Timothy, 1994, Plant age and technology use in us manufacturing industries, *The RAND Journal of Economics* 488–499.

Eisfeldt, Andrea L., and Dimitris Papanikolaou, 2013, Organization capital and the cross-section of expected returns, *Journal of Finance* 68, 1365–1406.

Epstein, Larry G, and Stanley E Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption, *Econometrica* 57, 937–969.

Fama, Eugene F, and Kenneth R French, 1992, The cross-section of expected stock returns, *The Journal of Finance* 47, 427–465.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F, and Kenneth R French, 1997, Industry costs of equity, *Journal of financial economics* 43, 153–193.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Favilukis, Jack, and Xiaoji Lin, 2015, Wage rigidity: A quantitative solution to several asset pricing puzzles, *The Review of Financial Studies* 29, 148–192.

Feng, Ying, 2019, Firm life-cycle learning and misallocation, *Working Paper* .

Gondhi, Naveen, 2017, Rational inattention, misallocation, and asset prices, *Working Paper* .

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *The Review of Financial Studies* 28, 650–705.

Hsieh, Chang-Tai, and Peter J Klenow, 2009, Misallocation and manufacturing tfp in china and india, *The Quarterly journal of economics* 124, 1403–1448.

Hulten, Charles R, 1991, The measurement of capital 119–158.

İmrohoroğlu, Ayşe, and Şelale Tüzel, 2014, Firm-level productivity, risk, and return, *Management Science* 60, 2073–2090.

Jermann, Urban J, 1998, Asset pricing in production economies, *Journal of monetary Economics* 41, 257–275.

Kogan, Leonid, and Dimitris Papanikolaou, 2012, Economic activity of firms and asset prices, *Annual Review of Financial Economics* 4, 1–24.

Kreps, David M, and Evan L Porteus, 1978, Temporal resolution of uncertainty and dynamic choice theory, *Econometrica* 185–200.

Kydland, Finn E, and Edward C Prescott, 1982, Time to build and aggregate fluctuations, *Econometrica: Journal of the Econometric Society* 1345–1370.

Lettau, Martin, and Jessica A Wachter, 2007, Why is long-horizon equity less risky? a duration-based explanation of the value premium, *The Journal of Finance* 62, 55–92.

Li, Erica XN, Dmitry Livdan, and Lu Zhang, 2009, Anomalies, *The Review of Financial Studies* 22, 4301–4334.

Lin, Xiaoji, 2012, Endogenous technological progress and the cross-section of stock returns, *Journal of Financial Economics* 103, 411–427.

Lin, Xiaoji, Berardino Palazzo, and Fan Yang, 2019, The risks of old capital age: Asset pricing implications of technology adoption, *Journal of Monetary Economics* .

Mairesse, Jacques, 1978, New estimates of embodied and disembodied technical progress, in *Annales de l'INSEE*, 681–720, JSTOR.

Melitz, Marc J, 2003, The impact of trade on intra-industry reallocations and aggregate industry productivity, *Econometrica* 71, 1695–1725.

Pástor, L'uboš, and Pietro Veronesi, 2009a, Technological revolutions and stock prices, *American Economic Review* 99, 1451–83.

Pástor, L'uboš, and Pietro Veronesi, 2009b, Learning in financial markets, *Annu. Rev. Financ. Econ.* 1, 361–381.

Restuccia, Diego, and Richard Rogerson, 2008, Policy distortions and aggregate productivity with heterogeneous establishments, *Review of Economic Dynamics* 11, 707–720.

Salvanes, Kjell G, and Ragnar Tveteras, 2004, Plant exit, vintage capital and the business cycle, *The Journal of Industrial Economics* 52, 255–276.

Tuzel, Selale, 2010, Corporate real estate holdings and the cross-section of stock returns, *The Review of Financial Studies* 23, 2268–2302.

Veronesi, Pietro, 2000, How does information quality affect stock returns?, *The Journal of Finance* 55, 807–837.

Weber, Michael, 2018, Cash flow duration and the term structure of equity returns, *Journal of Financial Economics* 128, 486–503.

Zhang, Lu, 2005, The value premium, *Journal of Finance* 60, 67–103.

Table 1. Misallocations on Portfolios Sorted by Capital Age

This table reports time-series averages of capital misallocations within five capital age quintile portfolios. The capital misallocation is computed through a two-step procedure. First, we compute the cross-sectional dispersion of marginal product of capital (MPK) relative to industry peers within a narrowly defined industry (either Fama-French 30 industries or SIC 2-digit industry code). Second, we take the average of the dispersion measure across industries. Misallocation 1 measures MPK by the ratio of operating income before depreciation (*oidbp*) to a one-year-lag net plant, property, and equipment (*ppent*) as in Chen and Song (2013), while Misallocation 2 measures MPK as sales (*sale*) over a one-year-lag net plant, property, and equipment (*ppent*) as in David, Schmid, and Zeke (2018). The sample period is from December 1978 to July 2016 and excludes utility, financial, and R&D intensive industries from the analysis. The detailed definition of the variables can be found in the Appendix D.7 .

	Y	**2**	**3**	**4**	**O**
Capital Age	9.71	15.04	19.86	24.66	35.95
Misallocation 1					
SIC 2-digit	1.02	0.88	0.81	0.80	0.77
FF30	1.09	0.93	0.87	0.85	0.79
Misallocation 2					
SIC 2-digit	1.12	0.91	0.84	0.79	0.82
FF30	1.13	0.93	0.87	0.90	0.93
Number of Firms	480	469	471	467	458

Table 2. Univariate Portfolio Sorting on Capital Age

This table shows asset pricing tests for five portfolios sorted on capital age relative to a firm's industry peers, for which we use the Fama-French 30 industry classifications and rebalance portfolios at the beginning of January, April, July, and October. The results use monthly data, for which the sample period is July 1979 to December 2016 and excludes utility, financial, and R&D intensive industries from the analysis. We report average excess returns over the risk-free rate E[R]-R_f, standard deviations Std, and Sharpe ratios SR across portfolios. Standard errors are estimated by Newey-West correction with ***, **, and * indicating significance at the 1, 5, and 10% levels. We include t-statistics in parentheses and annualize the portfolio returns by multiplying 12. All portfolio returns correspond to value-weighted returns by firm market capitalization.

Variables	**Y**	**2**	**3**	**4**	**O**	**OMY**
E[R] - R_f (%)	3.32	6.50**	8.34***	8.19***	9.11***	5.79***
[t]	0.95	2.24	3.44	3.48	3.76	2.91
Std (%)	21.16	18.43	15.39	14.69	15.2	13.27
SR	0.16	0.35	0.54	0.56	0.60	0.44

Table 3. Exposure to Aggregate Productivity Shocks and Learning Rate

This table shows aggregate exposures and learning rates by age groups. Panel A reports the aggregate productivity exposures of five firm groups sorted on a firm's capital age. All estimates are based on the following regression:

$$\Delta \ln A_{i,j,t} = \xi_{0i} + \xi_n \Delta \ln A_t + X_{i,j,t} + \varepsilon_{i,j,t}, \text{ for firms in age group n}$$

in which $X_{i,j,t}$ are control variables for firms' fundamentals, including size, book-to-market ratio, investment rate, and profitability. The exposures are normalized so that the firm exposure of the whole sample regression is equal to 1. Regressions (1) and (2) differ in that they use two alternative estimation methods in the first stage to estimate $\Delta \ln A_{i,j,t}$. Regression (1) is based on the fixed effect procedure, whereas Regression (2) is based on the dynamic error component method of Blundell and Bond (2000). These estimation methods are described in Appendix D.5 , following Ai, Croce, and Li (2013). Standard errors are adjusted for heteroscedasticity and clustered at the firm level. In the last row ("Model"), we report the model-implied ξ_1 based on our calibrated parameters, λ and ϕ. Panel B reports learning rates from the persistence of co-integration residuals. The ratio between the productivity of young firms ($n < \bar{n}$) and that of the mature firms ($n = \bar{n}$) denotes:

$$\chi_{n,t+1} = \ln \mathbf{A_{\bar{n},t+1}} - \ln \mathbf{A_{n,t+1}}$$

in which $n = 1, 2, \cdots, 4$ refers to the capital age sorted group from Y to 4, respectively, and $\bar{n}$ refers to group O. For each n, we estimate the autocorrelation $\rho_{n,s}$ by running a AR(1) regression of $\chi_{n,t}$. Standard errors are estimated by Newey-West correction. We report t-statistics in parentheses, and use *, **, and *** to indicate significance at the 1, 5, and 10% levels.

	Y	**2**	**3**	**4**	**O**
$\boldsymbol{\xi_1}$	Panel A: Aggregate Exposures				
(1)	0.50	0.63***	0.97***	0.97***	1.47***
[t]	1.03	3.57	8.91	3.65	3.88
(2)	0.67*	0.58***	1.00***	1.18***	1.82***
[t]	1.84	5.58	4.92	5.40	5.44
Model	0.72	0.85	1.00	1.17	1.38
$\boldsymbol{\rho_{n,s}}$	Panel B: Learning Rates				
(1)	0.88***	0.81***	0.80***	0.58***	
[t]	12.91	8.16	8.40	3.84	
(2)	0.85***	0.86***	0.75***	0.53***	
[t]	9.22	11.75	7.41	4.04	
Model	0.85	0.72	0.61	0.52	

Table 4. Calibration Parameters

This table reports a summary of parameters for our quarterly calibrations. The benchmark capital model features limited information and learning.

Preference parameters		
Risk aversion	γ	10
Intertemporal elasticity of substitution	ψ	2
Discount factor	β	0.997
Technology parameters		
Capital share	α	0.3
Depreciation rate of capital	δ	0.03
Learning parameters		
Capital stage transition rate	ϕ	0.08
Productivity exposure of mature generation	λ^*	1.91
Cointegration speed	ρ_s	0.96
Common productivity parameters		
Average growth rate	μ	0.0011
Volatility of short-run risk	$\exp(\sigma_a)$	0.016
Relative volatility of long-run risk	$\exp(\sigma_x)/\exp(\sigma_a)$	0.12
Auto-correlation of expected growth	ρ_x	0.946

Table 5. Capital Age and Capital Share

This table reports the average capital age (measured in years) and capital share of each capital-age firm group in the data and the benchmark model. Capital share is defined as the time series average of group *ppent* share ($\frac{ppent_n}{\sum_{n=1}^{5} ppent_n}$). The benchmark capital model features limited information and learning, but does not have adjustment cost. A detailed calculation of model counter-part is described in Appendix C.

Panel A: Capital Age					
	Y	2	3	4	O
Data	3.44	5.39	6.59	8.48	14.2
Model	2.26	4.52	6.78	9.04	16.73
Panel B: Capital Share					
Data	0.12	0.19	0.25	0.26	0.19
Model	0.29	0.21	0.15	0.10	0.25

Table 6. Aggregate Moments

This table reports macro quantities and asset returns in the model and data. The benchmark capital model features limited information and learning and is calibrated as in Table 4. RBC is the real business cycle model with convex adjustment costs. Panel A reports the moments of output, consumption, and investment. Panel B reports the equity premium, the risk-free rate and the capital age spread. $E(r_m - r_f)$ is the levered equity premium. $E(r_5 - r_1)$ is the levered spread between capital age group 5 and group 1. We assume a constant financial leverage ratio of 2. All the moments are annualized.

		Data	Benchmark	RBC
Panel A: Aggregate Quantities				
Average output growth	$E(\Delta y)$	2.00	2.00	2.00
Volatility of output growth	$\sigma(\Delta y)$	3.49	3.49	3.49
Volatility of consumption growth	$\sigma(\Delta c)$	2.53	2.73	3.13
Volatility of investment	$\sigma(\Delta i)$	16.40	6.41	5.04
Autocorrelation of output	$AC_1(\Delta y)$	0.45	0.53	0.42
Autocorrelation of consumption	$AC_1(\Delta c)$	0.49	0.67	0.46
Corr of consumption and investment	$corr(\Delta c, \Delta i)$	0.39	0.80	0.83
Panel B: Asset Prices				
Equity premium	$E(r_m - r_f)$	5.70	3.76	3.75
Risk-free rate	r_f	0.89	0.89	0.89
Volatility of equity return	$\sigma(r_m)$	17.61	2.83	4.38
Volatility of risk-free rate	$\sigma(r_f)$	0.97	1.13	1.28
Capital age premium	$E(r_5 - r_1)$	5.79	6.65	0

Table 7. Capital Age Premium and Cash Flow Duration

This table reports excess return and cash flow duration in the benchmark capital model and data. Panel A reports the average excess return (in annualized percentage term) of each capital age group. Panel B reports the cash flow duration (measured in years). The benchmark capital model features limited information and learning, but does not have adjustment cost.

Panel A: Excess Return						
	Y	2	3	4	O	OMY
Data	3.32	6.50	8.34	8.19	9.11	5.79
Model	0.24	2.92	4.78	6.11	6.89	6.65
Panel B: Cash Flow Duration						
Data	21.13	20.09	19.64	19.66	19.49	1.64
Model	20.95	20.68	20.57	20.51	20.46	0.49

Table 8. Payout Exposures to Productivity Shocks

This table shows payout exposures to short-run and long-run productivity shocks by capital age sorted quintile portfolios. All estimates are based on the following panel regression:

$$Z_{i,t} = \beta_{0,i} + \beta_{srr}\varepsilon_{a,t} + \beta_{lrr}\varepsilon_{x,t} + \rho Z_{i,t-1} + \beta_x x_{t-1} + Controls_{i,t-1} + resid_t,$$

in which $Z_{i,t}$ denotes firm i's payout (income-to-sales) ratio, $\varepsilon_{a,t}$ and $\varepsilon_{x,t}$ denotes short- and long-run shocks, respectively. Controls variables for a firm's fundamentals include size and book-to-market ratio. We further control for the predetermined value of the long-run component, x_{t-1}, and lagged payout ratio $Z_{i,t-1}$. Standard errors are adjusted for heteroscedasticity and clustered at the firm level. We report t-statistics in parentheses, and use *, **, and *** to indicate significance at the 1, 5, and 10% levels.

	Payout Exposures				
Variables	**Y**	**2**	**3**	**4**	**O**
$\varepsilon_{a,t}$	-0.01	0.18**	0.09	0.15	0.58***
[t]	-0.08	2.20	0.77	0.87	2.97
$\varepsilon_{x,t}$	0.37***	0.76***	0.91***	0.81***	1.04***
[t]	4.62	9.80	7.86	4.97	5.67
Z_{t-1}	0.36***	0.27***	0.18***	0.15***	0.22***
[t]	65.01	50.48	32.89	29.89	42.00
x_{t-1}	0.19*	0.30***	0.23	0.34*	0.91***
[t]	1.88	3.11	1.59	1.73	4.07
lagged log ME	2.55***	2.36***	3.96***	4.44***	5.71***
[t]	11.30	9.64	9.80	7.22	8.30
lagged B/M	-0.32***	-0.91***	-0.98***	-2.08***	-0.89***
[t]	-2.96	-8.37	-6.19	-9.14	-3.65
Observations	33,719	40,568	43,080	44,598	44,267
Firm FE	Yes	Yes	Yes	Yes	Yes

Table 9. Portfolio Sorting Conditional on Exposure to Technology Frontier Shocks

This table shows asset pricing test for five portfolios sorted on capital age conditional on industry-level exposures to technology frontier shocks (TFS), which is defined as the log difference in the number of new technology standards. First, for each industry and each quarter we compute the ±4-quarter correlation between industry-level output (sales) growth and TFS using 20-quarter rolling windows. Second, we assign the maximum absolute ±4-quarter cross-correlations to the corresponding industry to indicate TFS exposure. This procedure generates 30 industry-level indicators. Based on these indicators, We classify industries into high, medium, and low technology exposure groups, respectively. We report five portfolios sorted on capital age relative to their industry peers for industries with high exposure in Panel A, medium exposure in Panel B, and low exposure in Panel C, where we use the Fama-French 30 industry classifications and rebalance portfolios at the beginning of January, April, July, and October. The results use monthly data, where the sample starts from July 1979 to December 2016 and excludes financial, utility, and and R&D intensive industries from the analysis. We report average excess returns over the risk-free rate E[R]-R_f, standard deviations, and Sharpe ratios SR across portfolios. Standard errors are estimated by using Newey-West correction with ***, **, and * indicating significance at the 1, 5, and 10% levels. We include t-statistics in parentheses and annualize portfolio returns by multiplying 12.

	Y	**2**	**3**	**4**	**O**	**OMY**
	Panel A: Industries with High TAS Exposure					
E[R] - R_f (%)	0.18	6.05	8.66	8.84	10.39	10.21
[t]	0.04	1.83	3.27	2.95	4.23	3.37
Std (%)	21.71	19.3	16.3	16.5	16.87	17.09
SR	0.01	0.31	0.53	0.54	0.62	0.34
	Panel B: Industries with Medium TAS Exposure					
E[R] - R_f (%)	4.31	7.11	8.85	8.07	9.40	5.09
[t]	1.06	2.17	2.89	2.65	3.06	2.43
Std (%)	22.82	19.81	16.89	16.27	16.62	15.94
SR	0.19	0.36	0.52	0.50	0.57	0.36
	Panel C: Industries with Low TAS Exposure					
E[R] - R_f (%)	4.69	6.49	9.67	10.63	10.79	6.10
[t]	1.13	1.96	3.49	4.80	3.67	3.46
Std (%)	23.72	20.52	16.92	15.42	17.11	15.88
SR	0.20	0.32	0.57	0.69	0.63	0.36

Table 10. Asset Pricing Tests

This table shows asset pricing tests for five portfolios sorted on capital age relative to firm's industry peers, where we use the Fama-French 30 industry classifications and rebalence portfolios at the beginning of January, April, July, and October. The results use monthly data, where the sample period is July 1979 to December 2016 and excludes utility, financial, and R&D intensive industries from the analysis. In Panel A we report the portfolio alphas and betas by the Fama-French five-factor model, including MKT, SMB, HML, RMW, and CMA factors. In panel B we report portfolio alphas and betas by the HXZ q-factor model, including MKT, SMB, I/A, and ROE factors. Data on the Fama-French five-factor model are from Kenneth French's website. Data on I/A and ROE factor are provided by Kewei Hou, Chen Xue, and Lu Zhang. Standard errors are estimated by using Newey-West correction with ***, **, and * indicate significance at the 1, 5, and 10% levels. We include t-statistics in parentheses and annualize the portfolio alphas by multiplying 12. All portfolios returns correspond to value-weighted returns by firm market capitalization.

Variables	**Y**	**2**	**3**	**4**	**O**	**OMY**
			Panel A: Fama-French Five-Factor Model			
α_{FF5}	-5.80***	-3.02***	-1.00	-2.65***	-1.16	4.63***
[t]	-4.00	-3.52	-1.53	-2.72	-1.31	2.93
MKT	1.13***	1.11***	0.99***	1.01***	1.01***	-0.12**
[t]	28.72	45.25	41.87	61.79	40.19	-2.21
SMB	0.57***	0.25***	0.04	0.01	-0.08**	-0.65***
[t]	8.73	7.07	1.04	0.65	-2.16	-9.32
HML	0.09	0.07	-0.04	-0.00	0.08	-0.01
[t]	1.61	1.34	-0.80	-0.07	1.57	-0.12
RMW	0.07	0.16***	0.31***	0.38***	0.31***	0.23***
[t]	1.08	3.59	10.19	10.76	8.27	2.95
CMA	-0.36***	-0.13*	0.11*	0.38***	0.25***	0.61***
[t]	-3.74	-1.76	1.84	7.50	3.66	5.39
			Panel B: HXZ q-Factor Model			
α_{HXZ}	-4.25**	-1.81*	-1.22*	-1.72	-0.95	3.31*
[t]	-2.53	-1.83	-1.92	-1.09	-0.94	1.80
MKT	1.12***	1.09***	0.98***	0.97***	1.00***	-0.12**
[t]	28.23	38.72	38.13	52.76	32.32	-2.19
SMB	0.45***	0.16***	-0.01	-0.06**	-0.15***	-0.60***
[t]	4.67	3.10	-0.31	-2.57	-3.95	-7.29
I/A	-0.28***	-0.06	0.12***	0.39***	0.37***	0.65***
[t]	-3.41	-1.05	2.89	3.52	4.84	6.81
ROE	-0.10*	-0.00	0.24***	0.15***	0.19***	0.29***
[t]	-1.75	-0.05	8.45	2.61	3.42	2.89

Table 11. Fama-Macbeth Regressions

This table reports Fama-Macbeth regressions of individual stock excess returns on their capital age and other firm characteristics. The sample period is July 1979 to December 2016 and excludes financial industries. We regress monthly excess returns of individual stock on capital age with different sets of variables that are known for 6 months prior portfolio formation; we also control for industry fixed effects based on Fama-French 30 industry classifications. We present the time-series average and heteroscedasticity-robust t-statistics of the slopes (i.e., coefficients) estimated from the monthly cross-sectional regressions for different model specifications. All independent variables are normalized to a zero mean and a one-standard-deviation after winsorization at the 1th and 99th percentile of their empirical distribution. We include t-statistics and annualize individual stock excess returns by multiplying 12. Standard errors are estimated using the Newey-West correction, and ***, **, * indicate statistical significance at the 1, 5, and 10% levels.

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Cap. Age	2.59***	2.50***	1.65***	2.41***	3.83***	3.24***
[t]	5.57	5.50	3.51	5.33	5.90	4.94
Log ME		-1.59				-1.51
[t]		-1.34				-1.49
Log B/M			4.39***			3.13***
[t]			5.77			3.94
ROE				3.49***		3.09***
[t]				4.30		4.52
I/K					-2.81***	-2.35***
[t]					-4.04	-4.02
Constant	8.05	18.09***	11.34**	5.63	10.48*	9.02**
[t]	1.34	2.91	2.32	1.09	1.79	2.48
Observations	1,023,442	1,011,740	943,910	985,637	927,857	874,609
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes

Internet Appendix for "Learning and the Capital Age Premium"

A. Aggregation with learning

A.1 Static learning

Proof of Lemma 1. To prove Lemma 1, first we need to derive the optimal resource allocation in (5).

Lemma A.1. *1. The optimal resource allocation of problem (5) satisfies:*

$$\begin{aligned} n_i &= \frac{E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}}}{\int E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di} N, \\ k_i &= \frac{E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}}}{\int E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di} K, \end{aligned} \tag{A1}$$

where $K = \int k_i di$ is the total capital input and $N = \int n_i di$ is the total labor input.

2. The aggregate production of the firm group can be written as $Y = K^\alpha (AN)^{1-\alpha}$. The group level productivity A can be expressed as

$$A = \left[\int E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di\right]^{\frac{1-\nu}{(1-\alpha)\nu}}. \tag{A2}$$

which corresponds to (6) in the paper.

Proof. The first order conditions (FOC henceforth) of the firm i's profit maximization problem (4) are

$$\begin{aligned} \alpha E_s\left(p_i A_i^{1-\alpha}\right) k_i^{\alpha-1} n_i^{1-\alpha} &= R, \\ (1-\alpha) E_s\left(p_i A_i^{1-\alpha}\right) k_i^{\alpha} n_i^{-\alpha} &= W. \end{aligned} \tag{A3}$$

The FOCs imply that the expected marginal product of capital and labor are equalized across firms. The market price of intermediate good y_i is

$$p_i = \frac{\partial Y}{\partial y_i} = \left(\frac{y_i}{Y}\right)^{\nu-1}. \tag{A4}$$

Combine (A3) and (A4), we have

$$\frac{k_i}{k_{i'}} = \frac{n_i}{n_{i'}} = \frac{E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}}}{E_s\left(A_{i'}^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}}}, \quad \text{for any firm } i \text{ and } i'. \tag{A5}$$

Denote K as total capital input and N as total labor input, we get the resource allocation equations (A1). Plug equation (A1) into the production function (2), the output of a typical firm can be expressed as

$$\begin{aligned} y_i = A_i^{1-\alpha} k_i^\alpha n_i^{1-\alpha} &= A_i^{1-\alpha}\left[\frac{E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}}}{\int E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di} K\right]^\alpha \left[\frac{E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}}}{\int E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di} N\right]^{1-\alpha} \\ &= A_i^{1-\alpha}\frac{E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}}}{\int E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di} K^\alpha N^{1-\alpha}. \end{aligned} \tag{A6}$$

Total output of the firm group is

$$\begin{aligned} Y = \left(\int y_i^\nu di\right)^{\frac{1}{\nu}} &= \left\{\int \frac{E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{\nu}{1-\nu}} A_i^{(1-\alpha)\nu}}{\left[\int E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di\right]^\nu} di\right\}^{\frac{1}{\nu}} K^\alpha N^{1-\alpha} \\ &= \left[\int E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di\right]^{\frac{1-\nu}{\nu}} K^\alpha N^{1-\alpha}. \end{aligned} \tag{A7}$$

Define total productivity in equation (A7) as $A = \left[\int E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di\right]^{\frac{1-\nu}{(1-\alpha)\nu}}$, we have $Y = K^\alpha (AN)^{1-\alpha}$, as needed. $\square$

For a specific firm i, after observing the signal s_i, the posterior distribution of its exposure β_i is Normal with

$$\begin{aligned} Var_s[\beta_i] &= \frac{1}{\Delta a}\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau^2}}, \\ E_s[\beta_i] &= \frac{1}{\sigma^2+\tau^2}[\tau^2\mu + \sigma^2 s_i]. \end{aligned} \tag{A8}$$

Given $A_i = e^{m+\beta_i \Delta a}$ and β_i's posterior mean and volatility (A8), we can compute the expectation term in equation (A2) as

$$
\begin{aligned}
E_s(A_i^{(1-\alpha)\nu})^{\frac{1}{1-\nu}} &= e^{\frac{(1-\alpha)\nu}{1-\nu}m} E_s\left[e^{(1-\alpha)\nu\beta_i\Delta a}\right]^{\frac{1}{1-\nu}} \\
&= e^{\frac{(1-\alpha)\nu}{1-\nu}m} e^{\frac{(1-\alpha)\nu\Delta a}{1-\nu}E_s[\beta_i]+\frac{1}{2}\frac{(1-\alpha)^2\nu^2\Delta a^2}{1-\nu}Var_s[\beta_i]} \\
&= e^{\frac{(1-\alpha)\nu}{1-\nu}m+\left[(1-\alpha)\nu\frac{1}{\sigma^2+\tau^2}[\tau^2\mu+\sigma^2 s_i]+\frac{1}{2}(1-\alpha)^2\nu^2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right]\frac{1}{1-\nu}\Delta a}.
\end{aligned} \tag{A9}
$$

Because signal s_i follows a normal distribution with mean μ and variance $\frac{1}{\Delta a}[\sigma^2+\tau^2]$ across firms, the aggregate productivity of the firm group A can be computed as:

$$
\begin{aligned}
A &= \left[\int E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di\right]^{\frac{1-\nu}{(1-\alpha)\nu}} \\
&= \left\{e^{\frac{(1-\alpha)\nu}{1-\nu}m+\left[\frac{(1-\alpha)\nu}{\sigma^2+\tau^2}\tau^2\mu+\frac{1}{2}(1-\alpha)^2\nu^2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right]\frac{1}{1-\nu}\Delta a}\int e^{\frac{(1-\alpha)\nu}{\sigma^2+\tau^2}\sigma^2 s_i\frac{1}{1-\nu}\Delta a}di\right\}^{\frac{1-\nu}{(1-\alpha)\nu}} \\
&= \exp\left\{m+\left[\mu+\frac{1}{2}(1-\alpha)\nu\left(\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}+\frac{1}{1-\nu}\frac{\sigma^4}{\sigma^2+\tau^2}\right)\right]\Delta a\right\}.
\end{aligned} \tag{A10}
$$

Under the normalization condition $\mu = 1-\frac{1}{2}(1-\alpha)\nu\sigma^2$, the log firm group productivity can be rewritten as

$$
\ln A = m + \left[1+\frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{\sigma^4}{\sigma^2+\tau^2}\right]\Delta a = m + \lambda(\tau^2)\Delta a. \tag{A11}
$$

Take two extreme cases for example, if the firm obtains no information from the signal, i.e. $\tau^2 \to \infty$, then

$$
\lim_{\tau^2\to\infty}\lambda(\tau^2) = 1; \tag{A12}
$$

if the firm obtains full information, i.e. $\tau^2 \to 0$, then

$$
\lambda^* = \lim_{\tau^2\to 0}\lambda(\tau^2) = 1+\frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\sigma^2. \tag{A13}
$$

This completes the proof of lemma 1. □

Proof of Lemma 2. Firm i's realized marginal product of capital is

$$
MPK_i = p_i\frac{\partial y_i}{\partial k_i} = \frac{A_i^{(1-\alpha)\nu}}{E_s\left(A_i^{(1-\alpha)\nu}\right)}A^{1-\alpha}\left(\frac{N}{K}\right)^{1-\alpha}.
$$

The variance of realized log MPK can be computed as

$$
\begin{aligned}
&Var\left[\log(MPK_i) - \log(MPK)\right] = Var\left[\log(A_i^{(1-\alpha)\nu}) - \log(E_s(A_i^{(1-\alpha)\nu}))\right] \\
=&Var\left[(1-\alpha)\nu\beta_i\Delta a - [(1-\alpha)\nu\frac{1}{\sigma^2+\tau^2}[\tau^2\mu+\sigma^2 s_i] + \frac{1}{2}(1-\alpha)^2\nu^2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}]\Delta a\right] \\
=&Var\left[\left[(1-\alpha)\nu\beta_i\frac{\tau^2}{\sigma^2+\tau^2} - (1-\alpha)\nu\frac{\sigma^2}{\sigma^2+\tau^2}\epsilon_i\right]\Delta a\right] \\
=&(1-\alpha)^2\nu^2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\Delta a.
\end{aligned}
$$

□

A.2 Dynamic learning

In the dynamic setup, firm i's productivity follows $\ln A_{i,t+1} = \sum_{u=0}^{t+1}(m + \beta_{i,u}\Delta a_u)$. For ease of notation, we omit the firm index i and generation index n for the following analysis. For each period t, firms receive noisy signals about current period productivity exposure, β_t, as in the static model. In addition, we allow for perpetual learning: firms also receive signals about all past productivity exposures, $\{\beta_u | u < t\}$. We use the first subscript u to denote the timing of productivity exposure and the second subscript t to denote the arrival timing of signals. The signal follows:

$$
s_{u,t} = \beta_u + \epsilon_{u,t}, \qquad \epsilon_{u,t} \sim N\left(0, \frac{1}{\Delta a_u}\tau_{u,t}^2\right)
$$

To give a concrete example, we describe the information updating process of firm i in generation n as follows:

- In period 0, $lnA_0 = m + \beta_0\Delta a_0$. After observing the signal:

$$
s_{0,0} = \beta_0 + \epsilon_{0,0}, \qquad \epsilon_{0,0} \sim N\left(0, \frac{1}{\Delta a_0}\tau_{0,0}^2\right).
$$

  The posterior distribution of β_0 is updated as:

$$
\beta_0 \sim N\left(\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{0,0}^2}}\left[\frac{\mu}{\sigma^2}+\frac{s_{0,0}}{\tau_{0,0}^2}\right], \frac{1}{\Delta a_0}\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{0,0}^2}}\right).
$$

- In period 1, $lnA_1 = 2m + \beta_0\Delta a_0 + \beta_1\Delta a_1$. Manager receives a signal, $s_{1,1}$, on β_1. Under perpetual learning, firm manager also receive a signal $s_{0,1}$ on β_0 as well. The signals follows:

$$s_{0,1} = \beta_0 + \epsilon_{0,1}, \qquad \epsilon_{0,1} \sim N\left(0, \frac{1}{\Delta a_0}\tau_{0,1}^2\right),$$
$$s_{1,1} = \beta_1 + \epsilon_{1,1}, \qquad \epsilon_{1,1} \sim N\left(0, \frac{1}{\Delta a_1}\tau_{1,1}^2\right).$$

After observing these signals, the posterior distributions of β_0 and β_1 are updated as:

$$\beta_0 \sim N\left(\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{0,0}^2}+\frac{1}{\tau_{0,1}^2}}\left[\frac{\mu}{\sigma^2}+\frac{s_{0,0}}{\tau_{0,0}^2}+\frac{s_{0,1}}{\tau_{0,1}^2}\right], \frac{1}{\Delta a_0}\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{0,0}^2}+\frac{1}{\tau_{0,1}^2}}\right),$$
$$\beta_1 \sim N\left(\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{1,1}^2}}\left[\frac{\mu}{\sigma^2}+\frac{s_{1,1}}{\tau_{1,1}^2}\right], \frac{1}{\Delta a_1}\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{1,1}^2}}\right).$$

- In period 2, $lnA_2 = 3m + \beta_0\Delta a_0 + \beta_1\Delta a_1 + \beta_2\Delta a_2$. Manager receives signals on β_0, β_1 and β_2:

$$s_{0,2} = \beta_0 + \epsilon_{0,2}, \qquad \epsilon_{0,2} \sim N\left(0, \frac{1}{\Delta a_0}\tau_{0,2}^2\right),$$
$$s_{1,2} = \beta_1 + \epsilon_{1,2}, \qquad \epsilon_{1,2} \sim N\left(0, \frac{1}{\Delta a_1}\tau_{1,2}^2\right),$$
$$s_{2,2} = \beta_2 + \epsilon_{2,2}, \qquad \epsilon_{2,2} \sim N\left(0, \frac{1}{\Delta a_2}\tau_{2,2}^2\right).$$

The posterior distributions of β_0, β_1 and β_2 are updated correspondingly:

$$\beta_0 \sim N\left(\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{0,0}^2}+\frac{1}{\tau_{0,1}^2}+\frac{1}{\tau_{0,2}^2}}\left[\frac{\mu}{\sigma^2}+\frac{s_{0,0}}{\tau_{0,0}^2}+\frac{s_{0,1}}{\tau_{0,1}^2}+\frac{s_{0,2}}{\tau_{0,2}^2}\right], \frac{1}{\Delta a_0}\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{0,0}^2}+\frac{1}{\tau_{0,1}^2}+\frac{1}{\tau_{0,2}^2}}\right),$$
$$\beta_1 \sim N\left(\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{1,1}^2}+\frac{1}{\tau_{1,2}^2}}\left[\frac{\mu}{\sigma^2}+\frac{s_{1,1}}{\tau_{1,1}^2}+\frac{s_{1,2}}{\tau_{1,2}^2}\right], \frac{1}{\Delta a_1}\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{1,1}^2}+\frac{1}{\tau_{1,2}^2}}\right),$$
$$\beta_2 \sim N\left(\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{2,2}^2}}\left[\frac{\mu}{\sigma^2}+\frac{s_{2,2}}{\tau_{2,2}^2}\right], \frac{1}{\Delta a_2}\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau_{2,2}^2}}\right).$$

- In general, in period t, $\ln A_t = \sum_{u=0}^{t}(m + \beta_u\Delta a_u)$. Manager receives signals on the current and all past exposures $\{\beta_u | u \leq t\}$:

$$s_{0,t} = \beta_0 + \epsilon_{0,t}, \qquad \epsilon_{0,t} \sim N\left(0, \frac{1}{\Delta a_0}\tau_{0,t}^2\right);$$
$$s_{1,t} = \beta_1 + \epsilon_{1,t}, \qquad \epsilon_{1,t} \sim N\left(0, \frac{1}{\Delta a_1}\tau_{1,t}^2\right);$$
$$\vdots$$
$$s_{u,t} = \beta_u + \epsilon_{u,t}, \qquad \epsilon_{u,t} \sim N\left(0, \frac{1}{\Delta a_u}\tau_{u,t}^2\right);$$
$$\vdots$$
$$s_{t,t} = \beta_t + \epsilon_{t,t}, \qquad \epsilon_{t,t} \sim N\left(0, \frac{1}{\Delta a_t}\tau_{t,t}^2\right).$$

The updated posterior distributions of $\{\beta_u | u \leq t\}$ are:

$$\beta_0 \sim N\left(\frac{1}{\frac{1}{\sigma^2} + \sum_{v=0}^{t}\frac{1}{\tau_{0,v}^2}}\left[\frac{\mu}{\sigma^2} + \sum_{v=0}^{t}\frac{s_{0,v}}{\tau_{0,v}^2}\right], \frac{1}{\Delta a_0}\frac{1}{\frac{1}{\sigma^2} + \sum_{v=0}^{t}\frac{1}{\tau_{0,v}^2}}\right),$$
$$\beta_1 \sim N\left(\frac{1}{\frac{1}{\sigma^2} + \sum_{v=1}^{t}\frac{1}{\tau_{1,v}^2}}\left[\frac{\mu}{\sigma^2} + \sum_{v=1}^{t}\frac{s_{1,v}}{\tau_{1,v}^2}\right], \frac{1}{\Delta a_1}\frac{1}{\frac{1}{\sigma^2} + \sum_{v=1}^{t}\frac{1}{\tau_{1,v}^2}}\right),$$
$$\vdots$$
$$\beta_u \sim N\left(\frac{1}{\frac{1}{\sigma^2} + \sum_{v=u}^{t}\frac{1}{\tau_{u,v}^2}}\left[\frac{\mu}{\sigma^2} + \sum_{v=u}^{t}\frac{s_{u,v}}{\tau_{u,v}^2}\right], \frac{1}{\Delta a_u}\frac{1}{\frac{1}{\sigma^2} + \sum_{v=u}^{t}\frac{1}{\tau_{u,v}^2}}\right),$$
$$\vdots$$
$$\beta_t \sim N\left(\frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau_{t,t}^2}}\left[\frac{\mu}{\sigma^2} + \frac{s_{t,t}}{\tau_{t,t}^2}\right], \frac{1}{\Delta a_t}\frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau_{t,t}^2}}\right).$$

To have a concise representation of the aggregate productivity of generation n, we introduce the contemporaneous exposure parameter λ and learning speed parameter ρ. We show that if the information precision of signals $\{s_{u,u}, s_{u,u+1}, \cdots, s_{u,t}, s_{u,t+1}\}$ on time u exposure β_u satisfy

$$\lambda = 1 + \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{\tau_{t+1,t+1}^{-2}/\sigma^{-2}}{\sigma^{-2} + \tau_{t+1,t+1}^{-2}}, \tag{A14}$$

$$\rho = \frac{\sigma^{-2} + \sum_{v=u}^{t}\tau_{u,v}^{-2}}{\sigma^{-2} + \sum_{v=u}^{t+1}\tau_{u,v}^{-2}} \quad \text{for } u < t+1, \tag{A15}$$

for firms in generation n, then we have the recursive representation of lemma 3.

Equation (A14) specifies the relationship between generation n's contemporaneous productivity exposure and manager's signal precision on current period firm specific beta. Equation (A15) specifies the learning rate on past productivity. More specifically, denote the posterior variance of β_u at time $t+1$ as $V_{\beta_u}^{t+1}$. We have

$$\begin{aligned} V_{\beta_u}^{t+1} &= \frac{1}{\Delta a_u} \frac{1}{\sigma^{-2} + \sum_{v=u}^{t+1} \tau_{u,v}^{-2}} \\ &= \frac{\sigma^{-2} + \sum_{v=u}^{t} \tau_{u,v}^{-2}}{\sigma^{-2} + \sum_{v=u}^{t+1} \tau_{u,v}^{-2}} \frac{1}{\Delta a_u} \frac{1}{\sigma^{-2} + \sum_{v=u}^{t} \tau_{u,v}^{-2}} \\ &= \rho V_{\beta_u}^{t}, \end{aligned}$$

which means that the posterior variance of β_u decreases by a factor of ρ each period. Lower posterior variance implies a lower misallocation across firms, which, in turn, lead to higher productivity in the group level. Over time, firms in younger generations accumulate information to reduce the productivity gap with respect to the perfect information benchmark, i.e. $V_{\beta_u} = 0$. We can also manipulate the equation to have

$$\frac{V_{\beta_u}^{t} - V_{\beta_u}^{t+1}}{V_{\beta_u}^{t}} = 1 - \rho.$$

That is to say, every period, firms learn $1-\rho$ fraction of information, and $1-\rho$ fraction of productivity gap is reduced.

Up to now, we have talked about the dynamic learning process in a specific generation n. The crucial difference among different firm generations is that older generation firms have more precise information. Specifically, we assume mature firms (generation $\bar{n}$) learn perfect information and new firms (generation 1) learn no information on contemporaneous exposure. That is to say, $\lambda_{\bar{n}} = \lambda^*$ and $\lambda_1 = 1$ as what we have in the static model. For the generations in between $1 < n < \bar{n}$, we have $1 < \lambda_n < \lambda_{n+1} < \lambda^*$. For the mature firm generation $\bar{n}$, because it already obtains full information, $\rho_{\bar{n}} = 0$. For $n < \bar{n}$, older firm generations will have faster learning rate, i.e., $1 > \rho_n > \rho_{n+1} > 0$.

Proof of Lemma 3. At period $t+1$, after receiving a sequence of signals, the posterior distribution of β_u is:

$$N\left(\frac{1}{\sigma^{-2} + \sum_{v=u}^{t+1} \tau_{u,v}^{-2}} \left[\sigma^{-2}\mu + \sum_{v=u}^{t+1} \tau_{u,v}^{-2} s_{u,v} \right], \frac{1}{\Delta a_u} \frac{1}{\sigma^{-2} + \sum_{v=u}^{t+1} \tau_{u,v}^{-2}} \right).$$

Similar to Equation (A10) in Lemma 1, the aggregate productivity of a firm group can be computed as

$$\begin{aligned}\ln A_{t+1} &= \ln\left[\int E_s\left(A_{i,t+1}^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}} di\right]^{\frac{1-\nu}{(1-\alpha)\nu}} \\ &= \sum_{u=0}^{t+1}\left\{m + \left[\mu + \frac{1}{2}(1-\alpha)\nu\left(\frac{1}{\sigma^{-2}+\sum_{v=u}^{t+1}\tau_{u,v}^{-2}} + \frac{1}{1-\nu}\frac{\sigma^2\sum_{v=u}^{t+1}\tau_{u,v}^{-2}}{\sigma^{-2}+\sum_{v=u}^{t+1}\tau_{u,v}^{-2}}\right)\right]\Delta a_u\right\}.\end{aligned}$$

With the normalization condition $\mu = 1 - \frac{1}{2}(1-\alpha)\nu\sigma^2$, the aggregate productivity of a firm group can be rewritten as

$$\begin{aligned}\ln A_{t+1} &= \sum_{u=0}^{t+1}(m + \lambda(\tau_u)\Delta a_u) \\ &= \sum_{u=0}^{t+1}\left\{m + \left[1 + \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{\sigma^2\sum_{v=u}^{t+1}\tau_{u,v}^{-2}}{\sigma^{-2}+\sum_{v=u}^{t+1}\tau_{u,v}^{-2}}\right]\Delta a_u\right\}.\end{aligned} \tag{A16}$$

Now we are ready to derive the recursive representation of the perpetual learning model. For mature generation $\bar{n}$, because they have perfect signal each period, $\tau_{t+1,t+1} = 0$, their productivity growth follows:

$$\ln A_{\bar{n},t+1} = \sum_{u=0}^{t+1}\left\{m + \left[1 + \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\sigma^2\right]\Delta a_u\right\}$$

and

$$\ln A_{\bar{n},t+1} - \ln A_{\bar{n},t} = m + \lambda^*\Delta a_{t+1}.$$

Denote $\chi_{n,t+1}$ as the productivity difference between mature generation and young generation

$(n < \bar{n})$. With Equations (A14) and (A15), we have

$$
\begin{aligned}
\chi_{n,t+1} &= \ln A_{\bar{n},t+1} - \ln A_{n,t+1} \\
&= \sum_{u=0}^{t+1} \left(\frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{1}{\sigma^{-2} + \sum_{v=u}^{t+1} \tau_{u,v}^{-2}}\Delta a_u \right) \\
&= \sum_{u=0}^{t} \left(\frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{\sigma^{-2} + \sum_{v=u}^{t} \tau_{u,v}^{-2}}{\sigma^{-2} + \sum_{v=u}^{t+1} \tau_{u,v}^{-2}}\frac{1}{\sigma^{-2} + \sum_{v=u}^{t} \tau_{u,v}^{-2}}\Delta a_u \right) + \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{1}{\sigma^{-2} + \tau_{t+1,t+1}^{-2}}\Delta a_{t+1} \\
&= \rho_n \sum_{u=0}^{t} \left(\frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{1}{\sigma^{-2} + \sum_{v=u}^{t} \tau_{u,v}^{-2}}\Delta a_u \right) + \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{1}{\sigma^{-2} + \tau_{t+1,t+1}^{-2}}\Delta a_{t+1} \\
&= \rho_n \left(\ln A_{\bar{n},t} - \ln A_{n,t} \right) + \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\sigma^2 \left(1 - \frac{\tau_{t+1,t+1}^{-2}}{\sigma^{-2} + \tau_{t+1,t+1}^{-2}} \right) \Delta a_{t+1} \\
&= \rho_n \chi_{n,t} + (\lambda^* - \lambda_n)\Delta a_{t+1}.
\end{aligned}
$$

With equation (A14), (A15) and the definition of $\chi_{n,t}$, the productivity growth of young generation takes the form:

$$
\begin{aligned}
\ln A_{n,t+1} - \ln A_{n,t} &= \sum_{u=0}^{t} \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{\tau_{u,t+1}^{-2}}{(\sigma^{-2} + \sum_{v=u}^{t+1} \tau_{u,v}^{-2})(\sigma^{-2} + \sum_{v=u}^{t} \tau_{u,v}^{-2})}\Delta a_u + m + \lambda_n \Delta a_{t+1} \\
&= \sum_{u=0}^{t} \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu} \left(1 - \frac{\sigma^{-2} + \sum_{v=u}^{t} \tau_{u,v}^{-2}}{\sigma^{-2} + \sum_{v=u}^{t+1} \tau_{u,v}^{-2}} \right) \frac{1}{\sigma^{-2} + \sum_{v=u}^{t} \tau_{u,v}^{-2}}\Delta a_u + m + \lambda_n \Delta a_{t+1} \\
&= (1-\rho_n) \sum_{u=0}^{t} \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{1}{\sigma^{-2} + \sum_{v=u}^{t} \tau_{u,v}^{-2}}\Delta a_u + m + \lambda_n \Delta a_{t+1} \\
&= m + (1-\rho_n)\chi_{n,t} + \lambda_n \Delta a_{t+1}.
\end{aligned}
$$

This completes the proof of Lemma 3. □

Proof of Lemma 4. The variance of realized log MPK can be computed as

$$
\begin{aligned}
Var\left[\log(MPK_i) - \log(MPK)\right] &= Var\left[\log(A_i^{(1-\alpha)\nu}) - \log(E_s(A_i^{(1-\alpha)\nu}))\right] \\
&= \sum_{u=0}^{t} (1-\alpha)^2 \nu^2 \frac{1}{\sigma^{-2} + \sum_{v=u}^{t} \tau_{u,v}^{-2}}\Delta a_u \\
&= 2(1-\alpha)(1-\nu)\chi_{n,t}.
\end{aligned}
$$

□

B. Solution to the dynamic model

This section provide the details of model solution. The social planner solves

$$V(\mathbf{A}_t, \mathbf{K}_t) = \max_{I_t} \left\{ (1-\beta) C_t^{1-\frac{1}{\psi}} + \beta (E_t[V(\mathbf{A}_{t+1}, \mathbf{K}_{t+1})^{1-\gamma}])^{\frac{1-1/\psi}{1-\gamma}} \right\}^{\frac{1}{1-1/\psi}}, \tag{B17}$$

subject to the capital dynamics (21), (22), (23) and resource constraint (24). Define $F(\mathbf{A}_t, \mathbf{K}_t)$ as the aggregate output function. The FOC and the envelope conditions are:

$$MC_t = E_t \left[MV_{t+1} V_{K_1,t+1} \right], \tag{B18}$$

$$\begin{aligned} V_{K_n,t} &= MC_t F_{K_n,t} + (1-\delta)(1-\phi) E_t \left[MV_{t+1} V_{K_n,t+1} \right] + (1-\delta)\phi E_t \left[MV_{t+1} V_{K_{n+1},t+1} \right], \\ V_{K_{\bar{n}},t} &= MC_t F_{K_{\bar{n}},t} + (1-\delta) E_t \left[MV_{t+1} V_{K_{\bar{n}},t+1} \right], \end{aligned} \tag{B19}$$

where $MC_t = (1-\beta) V_t^{\frac{1}{\psi}} C_t^{-\frac{1}{\psi}}$ and $MV_{t+1} = \beta V_t^{\frac{1}{\psi}} V_{t+1}^{-\gamma} E_t[V_{t+1}^{1-\gamma}]^{\frac{\gamma - \frac{1}{\psi}}{1-\gamma}}$.

Define the cum-dividend prices of capital to be:

$$p_{K_n,t} = \frac{1}{MC_t} V_{K_n,t}, \tag{B20}$$

and the ex-dividend prices of capital to be

$$q_{K_n,t} = E_t[\Lambda_{t,t+1} p_{K_n,t+1}], \tag{B21}$$

where $\Lambda_{t,t+1}$ is the pricing kernel:

$$\Lambda_{t,t+1} = \beta \frac{MC_{t+1} MV_{t+1}}{MC_t} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\psi}} \left(\frac{V_{t+1}}{E_t[V_{t+1}^{1-\gamma}]^{\frac{1}{1-\gamma}}} \right)^{\frac{1}{\psi} - \gamma}. \tag{B22}$$

With the definition (B20)-(B22), we can simplify the notation of FOC (B18) to have:

$$1 = E_t[\Lambda_{t,t+1} p_{K_1,t+1}],$$

as in Equation (28). We can also simplify the notation of envelope conditions (B19) to have a recursive representation of capital prices:

$$\begin{aligned} p_{K_n,t} &= MPK_{n,t} + (1-\delta)(1-\phi) q_{K_n,t} + (1-\delta)\phi q_{K_{n+1},t}, \\ p_{K_{\bar{n}},t} &= MPK_{\bar{n},t} + (1-\delta) q_{K_{\bar{n}},t}. \end{aligned}$$

as in Equations (26) and (27).

C. Firm distribution

The survival rate of firm is $1-\delta$ per quarter. Therefore, if the steady state investment is I_{ss} the total measure of firms with age a is $I_{ss}(1-\delta)^{a-1}$. Given the transition rate ϕ, the total measure of firms with age a in generation n (denoted as $M_{n,a}$) can be computed as

$$M_{n,a} = (1-\delta)^{a-1}\bar{C}_{a-1}^{n-1}(1-\phi)^{a-n}\phi^{n-1}I_{ss}, \quad \text{for} \quad n < \bar{n}, \quad \text{if} \quad a-n<0, \quad M_{n,a}=0,$$
$$M_{\bar{n},a} = (1-\delta)^{a-1} - \sum_{n=1}^{\bar{n}-1} M_{n,a},$$

where $\bar{C}$ denotes the mathematics notation for combinations. The total measure of firms in group n can be computed as

$$M_n = \sum_{a=1}^{\infty} M_{n,a} = \frac{((1-\delta)\phi)^{n-1}}{(1-(1-\delta)(1-\phi))^n} I_{ss},$$
$$M_{\bar{n}} = \sum_{a=1}^{\infty} M_{\bar{n},a} = \frac{((1-\delta)\phi)^{\bar{n}-1}}{(1-(1-\delta)(1-\phi))^{\bar{n}-1}\delta} I_{ss}.$$

The average capital age of firms in group i can be computed as

$$Kage_n = \sum_{a=1}^{\infty} \frac{aM_{n,a}}{M_n} = \frac{n}{1-(1-\delta)(1-\phi)},$$
$$Kage_{\bar{n}} = \sum_{a=1}^{\infty} \frac{aM_{\bar{n},a}}{M_{\bar{n}}} = \frac{1}{\delta} + \frac{\bar{n}-1}{1-(1-\delta)(1-\phi)}.$$

D. Data construction

Given that a firms's capital age is unobservable, we rely on methodologies developed in the empirical industrial organization literature to estimate a firm's capital age from the firm-level investment.[14] We follow this stream of literature to construct capital age for U.S. publicly listed companies.

[14] See Dunne (1994), Mairesse (1978), Hulten (1991), and Salvanes and Tveteras (2004).

D.1 Data

Our sample consists of firms in the intersection of quarterly Compustat and CRSP (Center for Research in Security Prices). We obtain accounting data from Compustat and stock returns data from CRSP. Our sample firms include those with domestic common shares (SHRCD = 10 and 11) trading on NYSE, AMEX, and NASDAQ, excluding utility firms with SIC 4-digit (Standard Industrial Classification) codes between 4900 and 4949 and finance firms with SIC codes between 6000 and 6999 (finance, insurance, trusts, and real estate sectors). We additionally exclude R&D intensive sectors (SIC codes 283, 357, 366, 367, 382, 384, and 737) from our sample, following Brown, Fazzari, and Petersen (2009). According Fama and French (1993), we further drop closed-end funds, trusts, American depository receipts, real estate investment trusts, and units of beneficial interest. To mitigate backfilling bias, firms in our sample must be listed on Compustat for two years before including them in our sample. Macroeconomic data refers to the Bureau of Economic Analysis (BEA) maintained by the United States Department of Commerce. To minimize the impact of outliers, we winsorize all variables at the 1% and 99% level.

D.2 Capital age

We measure a firm's capital age following Lin, Palazzo, and Yang (2019). We first denote an initial measure of firm-level capital stock ($K_{i,0}$) for firm i using net property plant and equipment (*ppentq*) as the initial measure of firm i's capital age. After obtaining the initial capital age by calculating the ratio of accumulated depreciation and amortization (*dpactq*) over current depreciation and amortization (*dpq*), we recursively construct a measure of firm-level capital stock as follows:

$$K_{i,t+1} = K_{i,t} + I_{i,t}^{N}, \tag{D23}$$

in which $I_{i,t}^{N}$ denotes firm i net investment between period t and $t+1$. Net investment is defined as the difference in net property plan and equipment (*ppentq*) between two consecutive quarters. We denote firm i's gross investment as

$$I_{i,t}^{N} = \delta_j K_{i,t} + I_{i,t}^{N}, \tag{D24}$$

in which δ_j refers to the depreciation rate of industry j calculated by using industry-level depreciation data from BEA. All the quantities are expressed in 2009 dollars using the seasonally adjusted implicit price deflator for non-residential fixed investment.

When we construct firm-level capital stock and gross investment observations, the capital age of firm i at time t, according to Salvanes and Tveteras (2004), is defined as:

$$AGE_{i,t} = \frac{(1-\delta_j)^t K_{i,0}(AGE_{i,0}+t) + \sum_{j=0}^{t-1}(1-\delta_j)^{t-j-1} I_{i,j}(t-j)}{K_{i,t}}. \tag{D25}$$

The above specification implies that firm i's capital age at time t is a weighted average of the age of each capital vintage. The weights stand for the relative importance of each capital vintage in determining total capital in place at time t. We assume that firm i always installs the newest capital for its investment. For example, firm i's capital age is equal to $AGE_{i,0}$ at time $t = 0$. At the time $t = 1$, firm i's capital age is a weighted average of the new installed capital vintage, which has an age 1, and the old vintage capital that after one period has age $AGE_{i,0} + 1$. The weights are $(1-\delta_j)K_{i,0}/K_{i,1}$ for the past vintage after depreciation and $I_{i,0}/K_{i,1}$ for the new vintage, respectively, in which $K_{i,1} = K_{i,0} + I_{i,0}$.

Following the assumption in Lin, Palazzo, and Yang (2019) for the analytical convenience, firm i disposes all capital vintages in proportion to their contribution to the total installed capital when disinvestment occurs. Such the assumption enables us to obtain the recursive expression for firm i's capital age as follows:

$$AGE_{i,t} = (1-\delta_j)\frac{K_{i,t-1}}{K_{i,t}}(AGE_{i,t-1}+1) + \frac{I_{i,t-1}}{K_{i,t}}. \tag{D26}$$

Moreover, $AGE_{i,t} = AGE_{i,t-1} + 1$ when firm i has no positive investment. Taken together, the above specification suggests that we allow firm i to reduce its capital age only via positive investment.

D.3 Capital misallocation

Following Chen and Song (2013), we measure the marginal product of capital by the ratio of operating income before depreciation (*oibdpq*) to one-year-lag net plant, property, and equipment (*ppentq*). For robustness, we follow David et al. (2018) to construct an alternative measure of marginal product of capital by replacing the operating income before depreciation (*oibdpq*) with sales (*saleq*).[15] All the quantities are expressed in 2009 constant dollars using the seasonally adjusted implicit price deflator for non-residential fixed investment. Following Hsieh and Klenow (2009), we extend from manufacturing to all sectors, except utility, financial, high R&D industries, and compute the cross-sectional standard deviation as the dispersion measure within narrowly defined industries, as classified by the 2-digit SIC (Standard Industry Classification) industries, or broadly defined industries, as

[15] Using sales (*saleq*) to proxy a firm's output alleviates the missing data concern, given that the coverage of sales (*saleq*) is higher than that of operating income before depreciation (*oibdpq*) in Compustat.

classified by the Fama-French 30 industries. Specifically, for firm i in industry j, we compute

$$\log\left(\frac{MPK_{i,j}}{MPK_j}\right),$$

where MPK_j is the cross-sectional average of MPK measured at industry-level. We build the misallocation measure as follows. First, we compute the standard deviation of $\log\left(\frac{MPK_{i,j}}{MPK_j}\right)$ at industry-level within each portfolio sorted on capital age, where the number of observations within each narrowly or broadly defined industry must be larger than 10 to avoid biased standard deviations driven by a few extreme values.[16] Second, we take the cross-sectional average of standard deviations across industries within each portfolio. Finally, we report time-series averages of the cross-sectional dispersions of MPK in five portfolios sorted on capital age.

D.4 Cash flow duration

We construct firm i's cash flow duration to reflect the timing of cash flows, according to the model proposed by Lettau and Wachter (2007). Duration (Dur) is the equity implied cash flow duration. Dechow, Sloan, and Soliman (2004) proposes the measure of cash flow duration and documents a negative relationship between cash flow durations and stock returns; in addition, Weber (2018) recently studies asset pricing implications, including exposure to existing risk factors, time variations in the slope, and the effect of short-sale constraints.

Duration (Dur) resembles the traditional Maculay duration for bonds, which reflects the weighted average time to maturity of cash flows. The ratio of discounted cash flows to price determines the weights:

$$Dur_{i,t} = \frac{\sum_{s=1}^{T} s \times CF_{i,t+s}/(1+r)^s}{P_{i,t}}, \tag{D27}$$

in which we denote $Dur_{i,t}$ as firm i's duration at the end of fiscal year t, $CF_{i,t+s}$ is the cash flow at time $t+s$, $P_{i,t}$ is the current stock price, and r is the expected return on equity. Following Dechow, Sloan, and Soliman (2004), we assume the expected return on equity to be constant across both stocks and time. Relaxing such an assumption for firm-specific

[16] Industry coverage attrition issue is more salient for narrowly defined industries. To be concrete, after we impose this restriction, the valid number of industry coverage according to SIC 2-digit industry classifications drops from 75 to 42 for both MPK measures; in contrast, the valid number of industry coverage according to Fama-French 30 industry classifications drops from 28 to 24 for the MPK measure following Chen and Song (2013) and to 26 for the MPK measure following David, Schmid, and Zeke (2018).

discount rates generates larger cross-sectional variations in the duration measure since firms with high cash flow duration tend to be growth firms with lower stock returns. For the analytical simplicity, we focus on the measure of cash flow duration by using the constant expected return.[17]

In contrast to fixed income securities, such as bonds, stocks cannot have a well-defined finite maturity, $t+T$, and cash flows are not known in advance. Therefore, we split the duration formula into a finite detailed forecasting period and an infinite terminal value. We also assume that the later component is paid out as the level perpetuity for simplicity. Such the assumption allows us to rewrite the Equation (D27) as follows:

$$Dur_{i,t} = \frac{\sum_{s=1}^{T} s \times CF_{i,t+s}/(1+r)^s}{P_{i,t}} + \left(T + \frac{1+r}{r}\right) \times \frac{P_{i,t} - \sum_{s=1}^{T} CF_{i,t+s}/(1+r)^s}{P_{i,t}}. \quad \text{(D28)}$$

Moreover, we impose a clean surplus accounting, based on an accounting identify, and forecast cash flows via forecasting return on equity (ROE), $E_{i,t+s}/BV_{i,t+s-1}$, and growth in book equity, $(B_{i,t+s} - B_{i,t+s-1})/BV_{i,t+s-1}$:

$$\begin{aligned} CF_{i,t+s} &= E_{i,t+s} - (B_{i,t+s} - B_{i,t+s-1}) \\ &= B_{i,t+s-1} \times \left[\frac{E_{i,t+s}}{B_{i,t+s-1}} - \frac{B_{i,t+s} - B_{i,t+s-1}}{B_{i,t+s-1}}\right]. \end{aligned} \quad \text{(D29)}$$

Following Dechow, Sloan, and Soliman (2004), we model returns on equity and growth in equity as autoregressive process based on recent findings in financial accounting literature. In Weber (2018), the author estimates autoregressive parameters by using the merged CRSP-Compustat universe and assumes the mean reversion of ROE to the average cost of equity. We also follow the estimated autoregressive parameters in Weber (2018) by assuming that the growth in book equity is mean reverting to the long-run average growth rate in the economy with a coefficient of mean reversion equal to the average historical mean reversion in sales growth. The persistence of AR(1) model is 0.41 for ROE and 0.24 for BV, respectively. We assign the discount rate r to a value 0.12, which is equal to a steady-state average cost of equity of 0.12. Finally, we assign the average long-run nominal growth rate to a value 0.06, and use a detailed forecasting period of 15 years.

Because our model is highly stylized, the model implied duration can not directly be mapped into the data. Instead of selling the new project and pay out cash immediately, as what is assumed in the model, real companies will keep the new project in their balance

[17] According to Weber (2018), the variation over time in the return on equity r does not alter the cross-sectional ranking, which alleviate the concern for the cross-sectional implications.

sheet and collect cash flow later. Though the assumption in our model will generate the same valuation and return as in the real world, the timing of cash flow will be very different, which will lead to a much shorter cash flow duration. Therefore, in order to compute the corresponding duration in the model, we define a "company" that reinvest and postpone the collection of cash flow to their later stage. Define φ as the rate of investment that attributed to the existing firms, the cash flow duration of mature firms can be computed as:

$$MD_{\bar{n},t}q_{K_{\bar{n}},t} = E_t\left\{\Lambda_{t,t+1}\left[D_{\bar{n},t+1} + \varphi\frac{I_t}{K_t}MD_{1,t+1} + (1-\delta)(MD_{\bar{n},t+1}+1)q_{K_{\bar{n}},t+1}\right]\right\}.$$

For $n < \bar{n}$, the Macaulay duration MD_n can be computed as

$$MD_{n,t}q_{K_n,t} = E_t\left\{\Lambda_{t,t+1}\left[\begin{array}{c} D_{n,t+1} + \varphi\frac{I_t}{K_t}MD_{1,t+1} + \\ (1-\delta)[(1-\phi)(MD_{n,t+1}+1)q_{K_n,t+1} + \phi(MD_{n+1,t+1}+1)q_{K_{n+1},t+1}] \end{array}\right]\right\}.$$

We set φ to be 0.88 to broadly match the level of duration in the data.

D.5 Firm-level productivity estimation details

Firm-level productivity estimation Data and firm-level productivity estimation are constructed as follows. We consider publicly traded companies on U.S stock exchanges listed in both the annual Compustat and the CRSP (Center for Research in Security Prices) databases for the period 1950-2016. In what follows, we report the annual Compustat items in parentheses and defined industry at the level of two-digit SIC codes. The output, or value added, of firm i in industry j at time t, $y_{i,j,t}$, is calculated as sales (*sale*) minus the cost of goods sold (*cogs*) and is deflated by the aggregate gross domestic product (GDP) deflator from the U.S. National Income and Product Accounts (NIPA). We measure the capital stock of the firm, $k_{i,j,t}$, as the total book value of assets (*at*) minus current assets (*act*). This allows us to exclude cash and other liquid assets that may not be appropriate components of physical capital. We use the number of employees in a firm (*emp*) to proxy for its labor inputs, $n_{i,j,t}$, because data for total hours worked are not available.

We assume that the production function at the firm level is Cobb-Douglas and allow the parameters of the production function to be industry-specific:

$$y_{i,j,t} = A_{i,j,t}k_{i,j,t}^{\alpha_{1,j}}n_{i,j,t}^{\alpha_{2,j}},$$

where $A_{i,j,t}$ is the firm-specific productivity level at time t. This is consistent with our original specification because the observed physical capital stock, $k_{i,j,t}$, corresponds to the mass of production units owned by the firm.

We estimate the industry-specific capital share, $\alpha_{1,j}$, and labor share, $\alpha_{2,j}$, using the dynamic error component model adopted in Blundell and Bond (2000) to correct for endogeneity. Details are provided in Appendix D.5 . Given the industry-level estimates for $\widehat{\alpha_{1,j}}$ and $\widehat{\alpha_{2,j}}$, the estimated log productivity of firm i is computed as follows:

$$\ln \widehat{A_{i,j,t}} = \ln y_{i,j,t} - \widehat{\alpha_{1,j}} \cdot \ln k_{i,j,t} - \widehat{\alpha_{2,j}} \cdot \ln n_{i,j,t}.$$

We allow for $\widehat{\alpha_{1,j}} + \widehat{\alpha_{2,j}} \neq 1$, but our results hold also when we impose constant returns to scale in the estimation, that is, $\widehat{\alpha_{1,j}} + \widehat{\alpha_{2,j}} = 1$.

We use the multi-factor productivity index for the private non-farm business sector from the BLS as the measure of aggregate productivity.

Endogeneity and the dynamic error component model. We follow Blundell and Bond (2000) and write the firm-level production function as follows:

$$\begin{aligned} \ln y_{i,t} &= z_i + w_t + \alpha_1 \ln k_{i,t} + \alpha_2 \ln n_{i,t} + v_{i,t} + u_{i,t} \\ v_{i,t} &= \rho v_{i,t-1} + e_{i,t}, \end{aligned} \tag{D30}$$

where z_i, w_t, and $v_{i,t}$ indicate a firm fixed effect, a time-specific intercept, and a possibly autoregressive productivity shock, respectively. The residuals from the regression are denoted by $u_{i,t}$ and $e_{i,t}$ and are assumed to be white noise processes. The model has the following dynamic representation:

$$\begin{aligned} \Delta \ln y_{i,j,t} &= \rho \Delta \ln y_{i,j,t-1} + \alpha_{1,j} \Delta \ln k_{i,j,t} - \rho \alpha_{1,j} \Delta \ln k_{i,j,t-1} + \alpha_2 \Delta \ln n_{i,j,t} - \rho \alpha_2 \Delta \ln n_{i,j,t-1} \\ &\quad + (\Delta w_t - \rho w_{t-1}) + \Delta \kappa_{i,t}, \end{aligned} \tag{D31}$$

where $\kappa_{i,t} = e_{i,t} + u_{i,t} - \rho u_{i,t-1}$. Let $x_{i,j,t} = \{\ln(k_{i,j,t}), \ln(n_{i,j,t}), \ln(y_{i,j,t})\}$. Assuming that $E[x_{i,j,t-l} e_{i,t}] = E[x_{i,j,t-l} u_{i,t}] = 0$ for $l > 0$ yields the following moment coditions:

$$\begin{aligned} E[x_{i,i,t-l} \Delta \kappa_{i,t}] &= 0 \text{ for } l \geq 3 \\ E[x_{i,j,t-l} \Delta \kappa_{i,t}] &= 0 \text{ for } l \geq 3. \end{aligned} \tag{D32}$$

that are used to conduct a consistent GMM estimation of Equation (D31). Given the estimates $\widehat{\alpha}_{1,j}$ and $\widehat{\alpha}_{2,j}$, log productivity of firm i is computed as

$$\ln \widehat{a}_{i,j,t} = \ln y_{i,j,t} - \widehat{\alpha}_{1,j} \ln k_{i,j,t} - \widehat{\alpha}_{2,j} \ln n_{i,j,t}, \tag{D33}$$

where $\widehat{a}_{i,j,t}$ is the productivity for firm i in industry j.

Endogeneity and fixed effects. An alternative way to estimate the production function avoiding endengeneity issues is to work with the following regression:

$$\ln y_{i,j,t} = v_j + z_{i,j} + w_{j,t} + \alpha_{1,j} \ln k_{i,j,t} + \alpha_{2,j} \ln n_{i,j,t} + u_{i,j,t}. \tag{D34}$$

The parameter v_j, $z_{i,j}$, and $w_{j,t}$ indicate an industry dummy, a firm fixed effect, and an industry-specific time dummy, respectively. The residual from the regression is denoted by $u_{i,j,t}$. Given our point estimate of $\widehat{\alpha}_{1,j}$ and $\widehat{\alpha}_{2,j}$, we can use Equation (D33) to estimate $\widehat{a}_{i,j,t}$. Given this estimation of firms' productivity, we obtain the alternative estimation of firms' productivity.

D.6 More detailed firm characteristics

In Table D1, we report average capital age and other characteristics across five quintile portfolios sorted on capital age.

Table D1. Firm Characteristics

This table reports time-series averages of the cross-sectional averages of firm characteristics in five portfolios sorted on capital age, relative to their industry peers, where we use the Fama-French 30 industry classifications and rebalence portfolios at the beginning of January, April, July, and October. The sample period is from December 1978 to July 2016 and excludes utility, financial, and R&D intensive industries from our analysis. The detailed definition of the variables refers to Appendix D.

Variables	Y	2	3	4	O
Capital Age	9.71	15.04	19.86	24.66	35.95
Log ME	9.02	9.54	9.83	9.85	10.05
B/M	0.52	0.52	0.51	0.56	0.61
I/K	0.10	0.07	0.05	0.04	0.03
ROE	0.03	0.04	0.05	0.04	0.04
TANG	0.31	0.35	0.38	0.42	0.44
Number of Firms	480	469	471	467	458

On average, our sample contains $2,345$ firms. Firms are evenly distributed across five portfolios sorted on capital age, where the average number of firms in each portfolio ranges from 469 to 480. The cross-sectional variations in capital age are large. The average capital age in the lowest portfolio is 9.71; in contrast, the average capital age in the highest portfolio is 35.95. Both the size and book-to-market ratio (B/M) increases in portfolios sorted on capital age, which implies old capital firms, on average, are large and value firms. Moreover, we observe a downward sloping pattern of investment rate, since old capital firms with higher book to market ratio have less investment opportunities and thus invest less intensively than young capital firms. In addition, old capital firms are more profitable than young capital firms, since the former has more precise information to take advantage of technology growth, as we mentioned in Section 6.2. Finally, as we know that the size of old capital firms is lager, old capital firms hold more capital and have higher tangibility.

In summary, firms with high capital age have larger market capitalizations, higher book-to-market ratios, higher investment rates, higher profits, and higher tangibility.

D.7 Definition of firm characteristics

Market capitalization is calculated by using data from CRSP and it is equal to the number of shares outstanding (*shrout*) multiplied by the share price (*prc*). When size is reported to levels, we express it in 2009 dollars using the seasonally adjusted implicit price deflator for non-residential fixed investment. Quarter book value of equity is constructed following Hou, Xue, and Zhang (2015), and it is equal to shareholder's equity (*seqq*) plus deferred taxes and investment tax credit (*txditcq*, if available) minus the book value of preferred stock (*pstkrq*). If shareholder's equity is not available, we use common equity (*ceqq*) plus carrying value of the preferred stock (*pstkq*). If common equity is not available, we measure shareholder's equity as the difference between total assets (*atq*) and total liabilities (*ltq*). The book-to-market ratio is the book value of equity divided by the market capitalization (*prccq* times *cshoq*) at the end of the fiscal quarter. We measure the investment rate as gross investment $\delta_j K_{i,t} + I_{i,t}^N$ divided by the beginning of the period of capital stock $K_{i,t}$. Profitability is defined as income before extraordinary items (*ibq*) divided by the previous quarter book value of equity. The tangibility is net property plant and equipment (*ppentq*) divided by total assets (*atq*).