Learning and the Capital Age Premium

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

November 18, 2018

Abstract

This paper studies the implications of parameter learning on the cross-section of stock returns. We propose a production-based general equilibrium model to study the link between capital age, timing of cash flows and expected returns in the cross-section of stocks. Our model features slow learning about firms' exposure to aggregate productivity shocks over time. Firms with old capital are assumed to have more information about their exposure than firms with young capital. Our framework provides a unified explanation of the following stylized empirical facts: 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 call it the capital age premium; (3) have shorter cash-flow duration, as 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), Chi-Yang Tsou (bentsou@ust.hk) and Chenjie Xu (cxuab@ust.hk) are affiliated with the Hong Kong University of Science and Technology. We are grateful for helpful comments from Hengjie Ai, Utpal Bhattacharya, Yan Ji, Xiaoji Lin and Chu Zhang. We also thank the participants of HKUST Finance Brown Bag. The usual disclaimer applies. Address correspondence to Kai Li via E-mail: kaili@ust.hk.

1 Introduction

Parameter uncertainty is ubiquitous in finance. In this paper, we study the implications of parameter learning on the cross-section of stock returns. We develop a production-based general equilibrium asset pricing model and its key mechanism is that individual firms have imperfect information about their productivity and have to learn about their exposure to aggregate productivity shock over time. We show this model framework provides a unified explanation of a wide set of empirical facts on the links between capital age, resource allocation efficiency, timing of cash flows and expected returns in the cross-section of stocks.

The key model assumption is that old capital age firms have more information about their exposure than young capital age firms though learning. It is directly motivated by the empirical evidence in Section 2 that old capital age firms have less capital misallocations than young firms, because more information allows firms to better allocate resources. This key assumption leads to several important implications in the cross-section. First, In the equilibrium, heterogeneity in information translates into heterogeneity in risk exposures across different capital vintages. In particular, more precise information makes old capital age firms to be more sensitive to the common productivity shocks, and therefore, it implies a high average return of old capital age firms.

Second, this assumption also generates a negative link between capital age and duration of cash flow. Since young capital age firms have less information, hence, they have lower resource reallocation efficiency, are less able to take advantage of the aggregate technology growth, and pay lower payouts than when they get older. Therefore, young capital age firms should feature longer duration, as their future cash flows get more weight. Old capital firms, in contrast, should have shorter duration.

These model implications are strongly supported by the 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 (2012), Ai, Croce, Diercks, and Li (2018) and Lin, Palazzo, and Yang (2017), we measure 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 the standard procedure and sort firms into quintile portfolios based on these firms' capital ages within Fama-French 30 industries. As reported 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 the young capital age 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 et al. (2017). Moreover, we also find 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).

We 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 the finding Ai et al. (2012). 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. Third, we also find that the normalized payouts of young capital firms have a lower exposure to both the long-run and short-run aggregate productivity shocks than those of old capital firms.

We also examine the empirical evidence that differentiates our explanation from other alternative economic channels for the capital age premium. Lin et al. (2017) argues the technology adoption shock (TAS) as an additional source of risk that drives the capital age premium. We document that, even within those industries whose output growth or technology are not affected much by technology adoption shocks, the capital age premium are still present and significant, though its magnitude of the return spread is lower than that among high TAS exposure industries. This empirical evidence implies that both channels are in play 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.

Our calibrated model quantitatively matches the conventional macroeconomic quantity dynamics and asset pricing moments, and, more importantly, it is able to quantitatively account for the empirical relationship between capital age, duration of cash flows, and expected returns in the cross-section.

Literature Review This paper belongs to the literature on asset pricing in production economy, for which Kogan and Papanikolaou (2012) provide an excellent survey. It departs from previous articles in two significant aspects. First, our paper addresses the equity premium puzzle, so does the rest of the literature, but more importantly we focus on the spread between returns on old versus 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 have to learn about it over time, while most of previous work assumes perfect information. In this regard, our paper is closely related to Ai et al. (2012) and Ai et al. (2018), but, differently, we study the implications on the cross-section.

Our paper is also connected to the literature that links investment to the cross section of expected returns. 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. Our paper is closely related to Lin et al. (2017). Both papers document the capital age premium. The key difference is that Lin et al. (2017) provides a partial equilibrium model and emphasizes an additional source of technology adoption risk, while our paper is a general equilibrium framework and the key learning mechanism endogenously generates the asymmetric exposures of old versus young capital firms to the conventional aggregate productivity shocks. Moreover, the learning mechanism in our paper also coherently drives the negative relationships between capital age, capital misallocation and cash flow duration.

The learning mechanism that we emphasize in this paper is related to the literature that studies the impact of learning on asset market valuations. Pastor and Veronesi (2009) provide an excellent survey on learning models in finance. David (1997) and Veronesi (2000) study how learning and information affect both asset valuations and the risk premium on the equity market. Pástor and Veronesi (2009) present a model in which learning impacts the life-cycle dynamics of firms and their exposure to aggregate risks. The implication of their model that young firms are less exposed to aggregate shock than older firms is consistent with ours.

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. Some additional testable implications are presented in Section **??**. Section 7 concludes.

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 leave details of the data construction to Appendix D.

First, we investigate the empirical link between capital age and capital misallocation.[1]

[1] "Misallocation" is somewhat of a misnomer in our environment, as firms are acting optimally given the

Following Salvanes and Tveteras (2004), Ai et al. (2012), Ai et al. (2018) and Lin et al. (2017), we measure 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 the standard procedure and sort firms into quintile portfolios based on these firms' capital ages within Fama-French 30 industries. Table 1 reports the cross-sectional dispersion of the marginal product of capital (MPK hereafter) within each quintile portfolios, as a measure of capital misallocation following Hsieh and Klenow (2009). Within each quintile portfolios 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.

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, ranging from 1.13 in the young capital age quintile to 0.77 in the old capital age quintile. Such downward sloping pattern of misallocation across capital age sorted portfolios 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.

The negative relationship between capital age and capital misallocation provides suggestive evidence to support our key model assumption that young capital age firms contain less information about their exposures to the common productivity shocks than old capital age firms. Consistent with David, Hopenhayn, and Venkateswaran (2016), less information leads to lower resource reallocation efficiency. In our model, we assume old capital age firms contain full information about their exposure, but the evidence shows they still display positive MPK dispersion. This may be attributable to other factors, for example, adjustment costs, financial constraints, taxes, and regulations, that are not inside our model.

Next, we present the evidence on the cross-section of stock returns based on capital age sorted portfolios. Table 2 reports average annualized excess returns, t-statistics, and Sharpe ratios of the five capital age sorted portfolios. The average equity excess return for firms with old capital age (Quintile O) is 5.79% higher on an annualized basis than that of the young capital age 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 et al. (2017), with a slight difference that we sort portfolios within industries to control for industry heterogeneity, while Lin et al. (2017) does

information at hand. We follow the literature and use the term to refer broadly to deviations from marginal product equalization.

Table 1. Misallocations on Capital Age Sorted Portfolios

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 their industry peers with narrowly defined industry (either Fama-French 30 industries or SIC 2-digit industry code). Second, we take average of the dispersion measure across industries. Misallocation 1 measures MPK by the ratio of operating income before depreciation (OIDBP) to one-year-lag net plant, property and equipment (PPENT) as in Chen and Song (2013), while Misallocation 2 measures MPK as sales over 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 refers to Appendix.

	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

not. In this paper, we propose a learning mechanism that emphasizes firms are uncertain about their firm-specific exposures, and hence, our theory guides us to compare firms with the same industry that presumably have the same industry common exposures.

Table 2. Univariate Portfolio Sorting on Capital Age

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. We report average excess returns over the risk-free rate E[R]-R_f, standard deviations σ, and Sharpe ratios SR across portfolios. Standard errors are estimated by 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**
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

In sum, the above two facts together suggest that the learning mechanism proposed in this paper provides a potential coherent explanation. On one hand, old firms contain more precise information about their exposure to common productivity shocks, and hence has lower capital misallocation; on the other hand, old firms with more information allows them to take better advantage of aggregate technological progress, and therefore, they feature a high exposure to aggregate shocks and hence expect to earn a higher average return. In the next section, we develop a production-based general equilibrium model with learning to formalize the above 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 shock over time. In this section, we first describe the learning mechanism in a static framework. Then we incorporate learning into a 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 indexed with i. They produce intermediate inputs y_i, which can be transformed into final output Y using a CES production function:

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

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

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

We assume that $A_i = e^{\beta_i \Delta a}$, where Δa is a common shock that affects the productivity of all firms, and β_i is the firm-specific exposure to the common shock Δa. We assume the firm managers do not know exactly their exposure, β_i, and have to make production decision based on their interference on β_i. To facilitate a close-form solution, we assume that conditioning on the common shock Δa, β_i has a prior of a normal distribution $N(\mu, \frac{1}{\Delta a}\sigma^2)$. Before making the production decision, each firm receives a noisy signal of its own exposure:

$$s_i = \beta_i + \epsilon_i, \tag{2}$$

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

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

$$\max_{k_i, n_i} E_s[k_i^\alpha (A_i n_i)^{1-\alpha} p_i] - Rk_i - Wn_i, \tag{3}$$

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

In this economy, we adopts Dixit-Stiglitz aggregate production function among intermediate inputs with imperfect substitution, but different from Melitz (2003) and 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, and promises the second welfare theorem still holds. Under this assumption, all the intermediate firms take price p_i as given when solving their maximization problems, and each firm's production quantity has no impact on the market price p_i.

Define the aggregate production function of the firm group as

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

where 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 can be written as $Y = K^\alpha (\mathbf{A}N)^{1-\alpha}$, where

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

For simplicity, we impose the normalization condition $\mu = 1 - \frac{1}{2}(1-\alpha)\nu\sigma^2$ throughout the paper. As we will 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 group level production function is given by the following lemma:

Lemma 1. *The aggregate production function is given by*

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

where the aggregate productivity is given by $\ln \mathbf{A} = \lambda(\tau^2)\Delta a$*, and* $\lambda(\tau^2)$ *is defined as*

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

In the no private information case,

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

and in the full information case,

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

Proof. See Appendix A □

The above lemma has two intuitive implications. First, as firms acquire better information about their productivity, they can better allocate capital and labor across each other, as a result, the level of the aggregate productivity shock $\mathbf{A}N$ increases with information precision because $\lambda(\tau^2)$ is decreasing in τ^2. If we compare $\lambda(\tau^2)$ under the no information case and full information case, it is quite clear that we have $\lambda^* > 1$. Second, better allocation induced by higher information precision also acts as a risk exposure amplification mechanism, because the exposure to common productivity shocks increases with information precision.

3.1.2 Dynamic perpetual learning

Now we extend the above setup to a dynamic setting. Firm i productivity follows the following stochastic growth process:

$$A_{i,t} = \exp[\sum_{u=0}^{t} \beta_{i,u}\Delta a_u], \tag{4}$$

where $\{\Delta a_u\}_{u=0}^t$ is a sequence of 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 $\{\beta_{i,u}\}_{u=0}^t$ to be 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, that is, we allow firms to receive new signals about the entire history of their exposure coefficients in every period t. We describe the signal arrival process as below:

For a typical firm i starting its operation at time 0, its productivity $A_{i,t}$ follows equation (4). 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 for inferring the exposure coefficient $\beta_{i,u}$. In another word, for a typical 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, and the third subscript t denotes the arrival time of this signal. At the next period $t+1$, for each coefficient in $\{\beta_{i,u}\}_{u=0}^{t+1}$, there will be a new signal sequence $\{s_{i,u,t+1}\}_{u=0}^{t+1}$.

The signal processes at time t are described as below:

$$\begin{aligned}
s_{i,0,t} &= \beta_{i,0} + \epsilon_{0,t} \text{ with } \epsilon_{0,t} \sim N(0, \frac{1}{\Delta a_0}\tau_{0,t}^2); \\
s_{i,1,t} &= \beta_{i,1} + \epsilon_{1,t} \text{ with } \epsilon_{1,t} \sim N(0, \frac{1}{\Delta a_1}\tau_{1,t}^2); \\
&\vdots \\
s_{i,u,t} &= \beta_{i,u} + \epsilon_{u,t} \text{ with } \epsilon_{u,t} \sim N(0, \frac{1}{\Delta a_u}\tau_{u,t}^2); \\
&\vdots \\
s_{i,t,t} &= \beta_{i,t} + \epsilon_{t,t} \text{ with } \epsilon_{t,t} \sim N(0, \frac{1}{\Delta a_t}\tau_{t,t}^2),
\end{aligned} \tag{5}$$

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 the learning model. The parameter σ is the dispersion of firms' exposure to the aggregate productivity shocks. Intuitively, higher dispersion implies more benefit of reallocating resources across firms.

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 the firms. We choose a setup which allows us to talk about the link between capital age and expected returns, meanwhile we do not need to keep 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 capital age sorted portfolios in the empirical section. We use $n = 1$ to denote the youngest generation, while use $n = \bar{n}$ to denote the 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 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 on a negative relation between capital age and capital misallocation, we make a key assumption that more senior generation firms have more information about their exposure coefficients to the aggregate productivity. In the model, higher information precision is reflected by lower signal noise (τ^2). In the extreme case, we assume that mature firms (generation $\bar{n}$) know the exact values of $\{\beta_{i,u}\}_{u=0}^{t}$, that is, mature firms have full information about their exposures. When a new project is launched, managers do not exactly know its idiosyncratic exposure to aggregate productivity. Over time, managers accumulate more precise information about the project and make better resource allocation decision. We do not emphasize the learning of the exposure that is common among similar technology or similar business model, which can potentially be learned from other firms or past experience. Instead, we emphasize the imperfect information about firm specific exposure that can only be learned through its operation.

The learning parameters at the micro-level $\{\{\tau_{u,t}\}_{u=0}^{t}\}_{t=0}^{\infty}$ is an infinite-dimensional object, and not directly observable in the data. In the Appendix A.2 , equations (A3) and (A4) show that we can specify the sequence of $\{\{\tau_{u,t}\}_{u=0}^{t}\}_{t=0}^{\infty}$ as functions of parameters λ_n and $\rho_{n,s}$ for $n = 1, 2, ..., \bar{n}$, and we further have a recursive representation of group level productivity growth as in Lemma 2:

Lemma 2. *The aggregate output of firm group n is*

$$Y_{n,t} = K_{n,t}^{\alpha}(\mathbf{A_{n,t}}N_{n,t})^{1-\alpha} \tag{6}$$

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

$$\mathbf{A_{\bar{n},t}} = \exp[\sum_{u=0}^{t} \lambda^{*}\Delta a_u] \tag{7}$$

If the sequence of noise parameters $\{\{\tau_{u,t}\}_{u=0}^{t}\}_{t=0}^{\infty}$ are specified as functions of λ_n and $\rho_{n,s}$ for $n = 1, 2, ..., \bar{n}$ as equations (A3) and (A4) 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 an AR(1) process:*

$$\chi_{n,t+1} = \ln \mathbf{A_{\bar{n},t+1}} - \ln \mathbf{A_{n,t+1}} = \rho_{n,s}\chi_{n,t} + (\lambda^* - \lambda_n)\Delta a_{t+1}. \tag{8}$$

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

$$\ln \mathbf{A_{\bar{n},t+1}} - \ln \mathbf{A_{\bar{n},t}} = \lambda^* \Delta a_{t+1}, \tag{9}$$

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

$$\ln \mathbf{A_{n,t+1}} - \ln \mathbf{A_{n,t}} = (1 - \rho_{n,s})\chi_{n,t} + \lambda_n \Delta a_{t+1}. \tag{10}$$

We make several comments about Lemma 2. First, it is clear that young firms with imperfect information about their exposures will be less productive than mature firms on average. In order to guarantee the balanced growth, we keep the specification of productivity in equation (4), 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 (5). Lemma 2 shows that, thanks to the perpetual learning, young firms can eventually obtain full information about their exposures, and it rules out permanent productivity gaps between young and mature firms and guarantees balanced growth. In particular, equations (8), (9) and (10) fully specify the aggregate productivity of young and mature firms, that features the balanced growth.

Second, the parameters λ_n characterizes young firms' contemporaneous exposure to common shocks. As shown in equation (A3), λ_n is decreasing in the noise of the signal, τ^2. This is consistent with Lemma that firms with less information precision are less sensitive to the aggregate productivity shocks. Based on equation (A4), $1 - \rho_{n,s}$ can be interpreted as the learning rate about productivity. $\rho_{n,s}$ is increasing in the sequence of variances of the signals. Intuitively, higher values of τ^2 imply that young firms' information is less precise and, as a result, the productivity gap between young firms and mature firms can persistent 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 the macro parameters, the contemporaneous exposure λ_n and the learning rate $\rho_{n,s}$, for each firm generation n. We estimate these parameters from the difference in the exposure of young and old firms with respect to aggregate productivity shocks. As will be detailed in Section 4, the empirical evidences shows that $\lambda_{n+1} > \lambda_n$ and $\rho_{n,s} > \rho_{n+1,s}$, that is, younger firms have lower exposure to contemporaneous exposure to common productivity shocks and features a lower learning rate. The evidence strongly support our assumptions that young firms have less information on their firm specific exposure to common productivity shocks than mature firms.

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, like in Epstein and Zin (1989):

$$V_t = \{(1-\beta)C_t^{1-\frac{1}{\psi}} + \beta(E_t[V_{t+1}^{1-\gamma}])^{\frac{1-1/\psi}{1-\gamma}}\}^{\frac{1}{1-1/\psi}} \tag{11}$$

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

3.2.2 Output Producers

Following the long-run risks literature, the stochastic process for the common productivity growth is specified as

$$\begin{gathered}\Delta a_{t+1} = \mu + x_t + e^{\sigma_a}\varepsilon_{a,t+1} \\ x_{t+1} = \rho x_t + e^{\sigma_x}\varepsilon_{x,t+1} \\ \begin{bmatrix}\varepsilon_{a,t+1} \\ \varepsilon_{x,t+1}\end{bmatrix} \sim i.i.d.N\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}\right)\end{gathered} \tag{12}$$

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 does 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.

By assumption, there are $\bar{n}$ generations of firms based on their information precision. 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:

$$\begin{gathered}F(\mathbf{A}_t, \mathbf{K}_t) = \max_{N_{1,t}, N_{2,t}, \cdots, N_{\bar{n},t}} \sum_{n=1}^{\bar{n}} K_{n,t}^{\alpha}(\mathbf{A_{n,t}}N_{n,t})^{1-\alpha} \\ \text{subject to } \sum_{n=1}^{\bar{n}} N_{n,t} = 1\end{gathered} \tag{13}$$

Despite featuring substantial heterogeneity across firms, the production side of our model can be summarized by the production of a representative firm with the production function $Y_t = F(\mathbf{A}_t, \mathbf{K}_t)$, where the law of motion for productivity are characterized by equations (8)-(10), and the dynamics capital stocks of firm generations are given by equations (14)-(16), to be discussed in the next subsection.

3.2.3 Firm dynamics

News firms are created by physical investment goods. Upon creation, they belong to the youngest generation ($n = 1$). For parsimony, we assume that all firms are subject to the same exit rate, δ. We use $K_{n,t}$ to denote the total measure of firms in generation n at time t. In each period, the surviving firms of generation $n(n < \bar{n})$, $(1-\delta)K_{n,t}$, become the next generation, $n+1$, with a constant probability ϕ. Under this assumption, the law of motion of the mass of mature firms, $K_{\bar{n}}$, is

$$K_{\bar{n},t+1} = (1-\delta)K_{\bar{n},t} + (1-\delta)\phi K_{\bar{n}-1,t}, \tag{14}$$

The capital dynamics for middle firm generation ($1 < n < \bar{n}$) follow:

$$K_{n,t+1} = (1-\delta)(1-\phi)K_{n,t} + (1-\delta)\phi K_{n-1,t}, \tag{15}$$

The capital dynamics of the youngest firm generation follow:

$$K_{1,t+1} = (1-\delta)(1-\phi)K_{1,t} + I_t, \tag{16}$$

where I_t is investment.

To complete the model, we also have consumption plus investment equals total output:

$$C_t + I_t = Y_t. \tag{17}$$

3.2.4 Equilibrium conditions

Standard welfare theorems applies in our economy, we can construct equilibrium prices and quantities from the solution to a planner's problem. The stochastic discount factor $\Lambda_{t,t+1}$ can be written as

$$\Lambda_{t,t+1} = \beta(\frac{C_{t+1}}{C_t})^{-\frac{1}{\psi}} \left(\frac{V_{t+1}}{E_t[V_{t+1}^{1-\gamma}]^{\frac{1}{1-\gamma}}} \right)^{\frac{1}{\psi}-\gamma} \tag{18}$$

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

$$p_{K_{\bar{n}},t} = \alpha A_{\bar{n},t}^{1-\alpha}(\frac{K_{\bar{n},t}}{N_{\bar{n},t}})^{\alpha-1} + (1-\delta)E_t[\Lambda_{t,t+1}p_{K_{\bar{n}},t+1}] \tag{19}$$

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

$$p_{K_n,t} = \alpha A_{n,t}^{1-\alpha}(\frac{K_{n,t}}{N_{n,t}})^{\alpha-1} + (1-\delta)\{(1-\phi)E_t[\Lambda_{t,t+1}p_{K_n,t+1}] + \phi E_t[\Lambda_{t,t+1}p_{K_{n+1},t+1}]\} \tag{20}$$

Equation (19) implies that the cum-dividends marginal value of mature firms equals the current period marginal product of capital, $A_{\bar{n},t}^{1-\alpha}(\frac{K_{\bar{n},t}}{N_{\bar{n},t}})^{\alpha-1}$, and the expected continuation value of future payoffs, $p_{K_{\bar{n}},t+1}$, adjusted for the survival probability $1-\delta$.

According to equation (20), the value of adolescent firms (generation $n < \bar{n}$) is determined by the marginal product of its capital in the current period, $\alpha A_{n,t}^{1-\alpha}(\frac{K_{n,t}}{N_{n,t}})^{\alpha-1}$, 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 probability $1-\phi$, 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:

$$1 = E_t[\Lambda_{t,t+1}p_{K_1,t+1}] \tag{21}$$

The left-hand side of equation (21) is the marginal cost of investment and the right-hand side is the marginal benefit of investment.

3.3 Asset returns

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

$$q_{K_n,t} = E_t[\Lambda_{t,t+1}p_{K_n,t+1}], for n = 1, 2, ..., \bar{n}. \tag{22}$$

The return of capital takes the form:

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

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

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

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

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

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 the macro parameters, the contemporaneous exposure λ_n and the learning rate $\rho_{n,s}$, for each firm generation n. In this section, we provide the empirical procedure 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 help us to pin down the key learning parameters for a quantitative study of the model.

4.1 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.6 . 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.

4.2 Firms' exposure to aggregate shocks

To pin down the contemporaneous exposure λ_i, we estimate the exposure of firms' productivity with respect to the aggregate productivity by different capital age groups $(n = 1, 2, ..., \bar{n})$ using the following regression:

$$\Delta \ln A_{i,j,t} = \xi_{0,i} + \xi_1 \Delta \ln \bar{A}_t + \varepsilon_{i,j,t}$$

where $\xi_{0,i}$ controls for the firm-specific fixed effect, and $\Delta \ln \bar{A}_t$ is the growth rate of aggregate productivity as measured by the U.S. Bureau of Labor Statistics (BLS).

Table 3 Panel A shows the regression result within different capital age groups. The productivity exposure increase with capital age, which supports our model assumption that younger age firms have less information about their exposure. To directly map data to our model parameters, we need to do some normalization. Specifically, we divide group specific exposures by whole sample exposure. The model counter-part is the generation specific exposures normalized by the steady state capital share weighted exposure $\frac{\lambda_n}{(\sum_{n=1}^{\bar{n}} K_{n,ss}\lambda_n)/K_{ss}}$.

To maintain 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

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 firm's capital age. All estimate are based on the following regression:

$$\Delta \ln A_{i,j,t} = \xi_{0i} + \xi_1 \Delta \ln \overline{A}_t + X_{i,j,t} + \widetilde{\varepsilon}_{i,j,t},$$

where $X_{i,j,t}$ 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 $\Delta \ln A_{i,j,t}$. Regressions (1) is based on the fixed effect procedure, whereas Regressions (2) is based on the dynamic error component method of Blundell and Bond (2000). These estimation methods are describe in Appendix D.6 , following Ai, Croce, and Li (2012). 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

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

where $n = 1, 2, \cdots, 4$ refers to the capital age sorted group from Y to 4, respectively, and $\bar{n}$ refers to the 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	

exposure of youngest group to be 1, thus $\lambda_1 = 1$. And we calibrate $\lambda = 1.18$ such that the model implied normalized exposure broadly match the pattern in data. Panel A of Table 3 shows the normalized exposure in model and data are closely in line with each other.

4.3 Estimation of the learning rate

In our model, the learning rate parameters $\rho_{n,s}$ is the persistence of 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 young age group $n < \bar{n}$. Specifically, we define

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

where n indicates the capital age sorted group and $\mathbf{A_{n,t}}$ is the value weighted average productivity of firms in group n. For each n, we estimate the autocorrelation $\rho_{n,s}$ by running a AR(1) regression of $\chi_{n,t}$.

Since $1 - \rho_{n,s}$ is the learning rate on exposure in the past and mature generations learn faster, we expect $\rho_{n,s}$ to decrease from young to mature generation. Panel B of Table 3 reports the estimated autocorrelation of capital age groups 1-4. Indeed, there is a decreasing pattern from young to old firms. This directly support the learning mechanism of our model. Again, to maintain parsimony, we assume $\rho_{n,s}$ follows a exponentially decay pattern from young to mature generation: $\rho_{n,s} = \rho_s^n$ (In the model, the oldest generation has perfect information on their exposure. thus, $\rho_{\bar{n},s} = 0$). We calibrate the quarterly learning parameter $\rho_s = 0.96$ such that the model implied annual autocorrelations match 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. Consumption, output and physical investment data are 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 in the CRSP/Compustat Merged Fundamentals Annual Files, and (2) monthly stock returns from CRSP. Appendix D provides more details on our data sources and constructions.

Table 4 Calibration Parameters

This table reports a summary of parameters for our quarterly calibrations

Preference parameters		
Risk aversion	γ	10
Intertemporal elasticity of substituion	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	$\lambda^*\mu$	0.0021
Volatility of short-run risk	$(1-\alpha)\lambda^* \exp(\sigma_a)$	0.021
Relative volatility of long-run risk	$\exp(\sigma_x)/\exp(\sigma_a)$	0.12
Autocorrelation of expected growth	ρ_x	0.946

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 $\psi = 2$, in line with the long-run risks literature, such as 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 $\delta_k = 0.03$, consistent with the standard real business cycles literature Kydland and Prescott (1982).

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, 1929-2015. We set relative volatility $\exp(\sigma_x - \sigma_a) = 0.12$ and autocorrelation of long-run risk $\rho_x = 0.946$, in the same spirit of 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; and the persistence of the cointegration residual ρ_s governs the speed of learning for young firms with imprecise information about their exposure to comment productivity shocks. We

Table 5 Capital Age and Capital Share

This table reports the average capital age and capital share of each firm group in the data and model. Capital share is defined as time series average of group PPENT share ($\frac{PPENT_i}{\sum_{i=1}^{5} PPENT_i}$). 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	0.1	0.25

choose the parameter ϕ to broadly match steady state distribution of capital age and capital share. As we show in Appendix C, given 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 shown 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. We have already provide supportive evidence and calibration details about the other two learning parameters λ^* and ρ_s in previous Section 4.

In addition to our benchmark calibration, we also calibrate a RBC model with adjustment cost and report the result for comparison. When calibrating the RBC model, we retain the same parameter expect for three modifications. We keep 5 capital age groups but shut down the learning channel ($\lambda_i = 1$ and $\rho_{i,s} = 0$). That is to say 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 adjustment cost of investment to generate equity premium. We model the adjustment cost as in 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 Quantitative results

5.2.1 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.[2] We show that our model is broadly consistent with the key empirical features of macroeconomic quantities and asset prices.

Table 6 reports the key moments of macroeconomic quantities (top panel) and those of asset returns (bottom panel) respectively, and compares them to their counterparts in the data where available.

In terms of aggregate moments on macro quantities (top panel), our model have similar implications as the standard RBC models. In particular, our calibration features a low volatility of consumption growth (2.73%), match the mean, volatility and autocorrelations of output and consumption growth with the data reasonably well. The volatility of investment (6.41%) is reasonably high, though still a bit lower than the data counterpart, and correlation between consumption and investment growth (0.80) is a little bit overshooting. These features reflect a common issue for RBC model with adjustment cost to produce a sizable equity premium. We can remedy this by introducing intangible capital as in Ai et al. (2018), but as we focus on the learning mechanism and its cross-sectional implication, we intentionally avoid introducing additional complication.

Turning the attention to the 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 market. It is worth noting that our model is able to generate a high equity premium even without adjustment cost. Because new capital does not fully enjoy the productivity growth, the consumption-smoothing effect of investment is mitigated. RBC model with adjustment cost can also generate a comparable equity premium, but with much lower investment volatility (5.04%). In addition, RBC fail to generate any capital age spread.

5.2.2 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 compare them with the data.

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

Table 6 Aggregate Moments

This table reports macro quantities and asset prices in the model and data. RBC is the real business cycle model with adjustment cost. Panel A reports the moments of output, consumption and investment. Panel B reports equity premium and risk-free rate. $E(r_m - r_f)$ is the levered equity premium. $E(r_5 - r_1)$ is the levered spread between capital age group 5 and group 1. We use a leverage ratio of 2

		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 model and data. Panel A reports the average excess return of each capital age group. Panel B reports the cash flow duration.

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

We make several observations. First, Panel A of Table 7 reports 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 is able to 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 on their exposure to common shock, so they have better allocation of resources. This has an amplification that makes their marginal product of capital more exposed to common productivity shocks.

Second, Panel B of Table 7 shows the cash flow duration of each capital age group in the data and model. Appendix ?? provides detailed construction of cash flow duration. In the data, young capital age firms have higher cash flow duration than old capital age firms. Our model generate the same monotonic pattern. Intuitively, firms in the young generation have on average lower productivity, so they also have lower dividend payout. As they grow into mature firms, they acquire better information and have better resource allocation, hence their productivity and dividend payout increase. Therefore, young firms have low cash flow at the short end and high cash flow at the far end. While for mature firms, the cash flow is evenly distributed.

5.2.3 Impulse response functions

To better understand the above results, we plot the impulse response functions with respect to a 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 solid black line and the impulse response of RBC model is plotted with red dashed line for comparison.

The response of consumption growth and investments to shock-run shocks are similar in benchmark model and RBC model. Upon the arrival of a positive shock-run shock, consumption growth increase temporarily. Investment to capital ratio also increase and slowly decay.

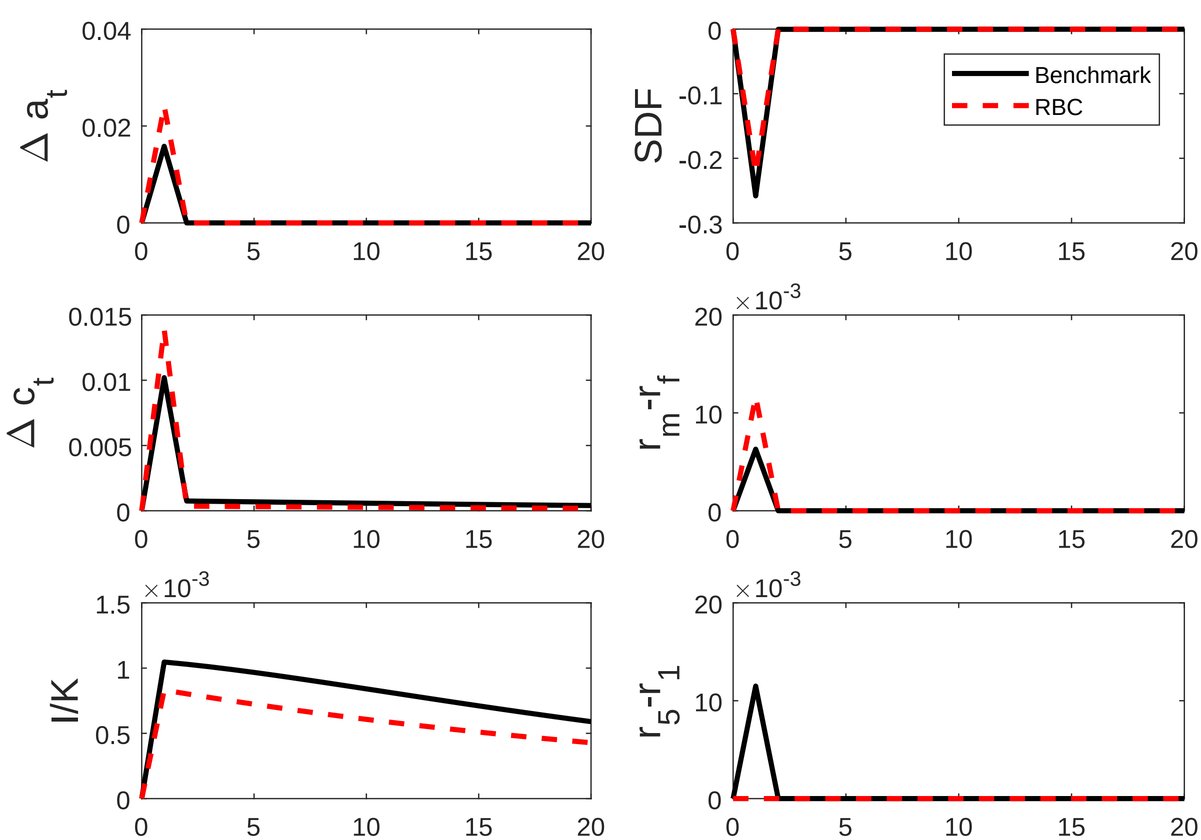
However, we observe that the impulse responses of investment and consumption to long-run shocks are very different in two models. Notice that, in the RBC model, investment responds positively to news shocks. With an IES of 2, upon the arrival of positive news about future productivity shocks, the substitution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

In 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 one hand, the substitution effect is moderate. New investment builds young capital firms, which cannot take full advantage of the rise in productivity. Since 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 the positive productivity shock. As a result, investment temporarily drops and consumption increases.

With $IES > 1$, news about future consumption growth requires a significant premium.

Figure 1. Impulse Response Functions for Short-Run Shock

This figure shows 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 model and the dashed red line is the RBC model with adjustment cost. Returns are not levered

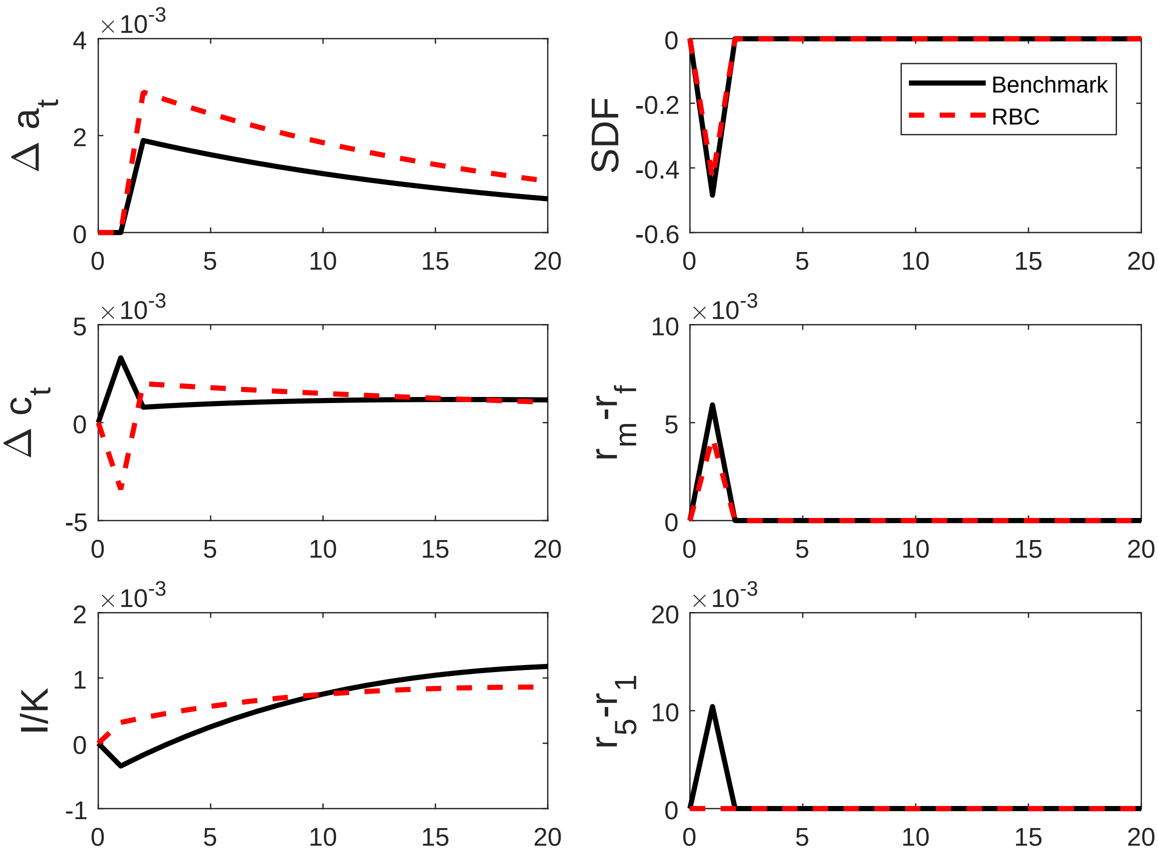


Long-run risks is the main source of risk premium. The market return response positively to productivity shocks in both models. Without adjustment cost, the marginal q of young firms always equals 1 in our benchmark model. However, since we can not directly create old capital firms. The marginal q of old capital firms will fluctuate with business cycle. This explains why in our benchmark model, though there are no investment frictions, market return also response to productivity shocks.

Our main focus is the different response of excess return on mature firms (group 5) and young firms (group 1). In our benchmark model, the response of return on mature firms to productivity shock is significantly larger than that of young firms. This contributes to generate 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 mis-allocation, young firms could only benefit a little from productivity

Figure 2. Impulse Response Functions for Long-Run Shock

This figure shows 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 model and the dashed red line is the RBC model with adjustment cost. Returns are not levered.



growth. Mature firms' marginal product of capital is more sensitive to common shocks than young firms. Therefore, mature firms' return also response more to shocks than young firms. In RBC model, because all firms are identical, the return spread is always 0.

6 Testable implications of the learning mechanism

In this section, we examine the empirical evidence that differentiates our explanation from other alternative economic channels for the capital age premium. That is, our learning mechanism, compared with the technology adoption mechanism in Lin, Palazzo, and Yang (2017), explains for different capital ages with different information precisions about exposures to aggregate TFP shocks. Moreover, we test testable implications of the learning mechanism in the following subsections.

6.1 Alternative explanations

Lin, Palazzo, and Yang (2017) provide empirical evidence of capital age premium and propose a risk-based story, where the driving source is the technology adoption shock (TAS), to interpret a positive relationship between the firm's capital age and future stock returns in the cross-section. As a result, old capital age firms face higher exposures to the TAS than young capital age firms, and thus carry higher returns as the risk compensation. On the other hand, we propose an alternative learning story to drive the capital age premium. The key driving force in our learning mechanism relies on the cross-sectional difference in information precisions about exposures to aggregate TFP shocks. We find empirical evidence that the capital age premium remains existing for firms in industries with low exposures to TAS. The details to construct exposures to TAS are as follows. First, we calculate cross-correlations to the TAS across industries. For each industry, we compute the ± 4-quarter cross-correlation between industry-level output (i.e., sales) growth and the TAS, which is defined as the log difference in the number of new technology standards, following Lin, Palazzo, and Yang (2017). Second, the summation of absolute ± 4-quarter cross-correlations is attained, and we assign the summation to the corresponding industry as indicator of TAS exposures. This procedure generates 30 industry-level indicators. Based on these indicators, we classify industries into high, middle, and low technology exposures, respectively. In Panel A of Table 8 we pool industries with high and middle exposures to TAS and implement univariate portfolio sort on capital age relative to their industry peers. In Panel B of Table 8 we sort on capital age within industry for firms in industries with low technology exposures. 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 both Panels.

Table 8 reports the average annualized excess returns in five quintile portfolios and long-short portfolio in Panel A and B. The portfolio returns are economically large, ranging from 2.25% to 10.32% in Panel A and from 4.66% to 9.38% in Panel B, and the long-short portfolios are statistically significant at least at 5% level. The capital age premium is robust to industries with low exposures to TAS, although the return spread is smaller in the low exposures industries (Panel B) than that in the high and middle group (Panel A). As the output of firms in industries with low TAS exposures are less affected by TAS risk, the statistically significance of OMY portfolio in Panel B implies that the technology adoption mechanism alone cannot fully account for the capital age premium. To fill the gap, we focus

Table 8. **Portfolio Sorting Conditional on Exposures to Technology Adoption Shock**

This table shows asset pricing test for five portfolios sorted on capital age conditional on industry-level exposures to TAS. First, for each industry, we compute the ±4-quarter correlation between industry-level output (sales) growth and the technology adoption shock, which is defined as the log difference in the number of new technology standards. Second, the summation of absolute ±4-quarter cross-correlations is attained, and we assign the summation to the corresponding industry as indicator of TAS exposures. This procedure generates 30 industry-level indicators. Based on these indicators, We classify industries into high, middle, and low technology exposures, respectively. We report five portfolios sorted on capital age relative to their industry peers for industries with high and middle exposures in Panel A and for industries with low exposures in Panel B, 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 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 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 portfolio returns by multiplying 12.

	Y	**2**	**3**	**4**	**O**	**OMY**
	Panel A: Industries with High & Medium TAS Exposures					
E[R] - R_f (%)	2.25	5.90*	7.67***	8.02***	10.32***	8.08***
[t]	0.61	1.91	2.82	3.19	4.27	3.55
	Panel B: Industries with Low TAS Exposures					
E[R] - R_f (%)	4.66	8.37***	10.12***	8.93***	9.38***	4.73**
[t]	1.18	2.75	3.82	3.70	3.68	2.17

on a learning mechanism of firm specific productivity exposure and reconcile the co-existence of capital age premium across different exposures to TAS, as our main contribution. In deed, both channels can be simultaneous in play, but our channel is qualitatively important to capture the essence of capital age premium.

6.2 Productivity shocks and payouts

The key implication of the learning mechanism in our model is the response of payouts to aggregate productivity shocks. In this subsection, we directly test this implication of our model using evidence on macroeconomic quantities. We show that firm's payout has a positive exposure to long-run shocks. The cross-sectional difference in productivity exposures increases in firms' capital ages.

We proceed as follows. First, we measure payout using firm's 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). Because our model abstracts away from leverage and capital structure decisions, payouts in our model cannot be directly compared to stock market dividends. We therefore use the model to guide our empirical measurement. Given the data limitation of payout disclosure in quarterly

Compustat at early periods, we start our sample here from 1984:Q2. To maximize sample length, our data include observations through 2016:Q1.

In the second step, we estimate the following panel regression:

$$Z_{i,t} = \beta_{0,i} + \beta_{srr}\varepsilon_{a,t} + \beta_{lrr}\varepsilon_{x,t} + \rho Z_{i,t-1} + \beta_x x_{t-1} + Controls_{i,t-1} + resid_t, \tag{23}$$

where $Z_{i,t}$ denotes firm i's payout (income-to-sales) ratio, $\varepsilon_{a,t}$ and $\varepsilon_{x,t}$ denote short-run and long-run shocks, respectively. Detailed in constructions for $\varepsilon_{a,t}$ and $\varepsilon_{x,t}$ refer to Ai, Croce, Diercks, and Li (2018). Controls variables for firm's fundamentals include size and book-to-market ratio. We divided our variable of interest by sales for three reasons as follows. First, since dividends could be negative, we cannot just focus on growth rates. Second, This is a common way to detrend our variable. Third, according to the model, it does not affect our ability to identify the sensitivity of our variable to long-run shocks, as sales is nearly invariant upon the arrival of pure long-run 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 payout ratio to capture the role of the endogenous state variables (i.e., capital shocks) to avoid additional measurement errors. Under the null of the model, this is an innocuous assumption. We also control for the predetermined value of the long-run component, x_{t-1}, size and book-to-market ratio for firms' fundamentals, and firm fixed effects. Our main findings are reported in Table 9. The responses of firm's payout to long-run risks are positive across portfolios. More importantly, we can observe an upward sloping pattern on coefficients for the long-run productivity shocks from young to old capital age portfolio. That is, firms' payouts in the highest quintile portfolio face higher exposure to the long-run productivity shocks than those in the lowest. In addition, we also observe an upward sloping pattern on coefficients for the short-run productivity shocks from young to old capital age portfolio. To sum up, payout features a positive response to both short- and long-run productivity shocks, which is perfectly consistent with our model implication.

7 Conclusion

In this paper we argue that parameter learning is a potential key determinants of the cross-section of stock returns. We develop a production-based general equilibrium asset pricing model and its key mechanism is that individual firms have imperfect information about their productivity and have to learn about their exposure to aggregate productivity

Table 9. Payout Exposures

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:

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

where $Z_{i,t}$ denotes firm i's payout (income-to-sales) ratio, $\varepsilon_{a,t}$ and $\varepsilon_{x,t}$ denote short- and long-run shocks, respectively. Controls variables for 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

shock over time. We show this model framework can provide a unified explanation of a wide set of empirical facts: 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 call it the capital age premium; (3) have shorter cash-flow duration, as compared with young capital firms.

A Aggregation with learning

A.1 Static learning

To prove Lemma 1, first we need to derive the optimal resource allocation.

Lemma 3. *1. The resource allocation satisfy:*

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

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

2. The capital rent and wage rate can be written as:

$$\begin{aligned} R &= \alpha A^{1-\alpha} K^{\alpha-1} N^{(1-\alpha)}, \\ W &= (1-\alpha) A^{1-\alpha} K^{\alpha} N^{-\alpha}, \end{aligned} \tag{A2}$$

where $A = \left[\int E_s(A_i^{(1-\alpha)\nu})^{\frac{1}{1-\nu}}di\right]^{\frac{1-\nu}{(1-\alpha)\nu}}$ is the total productivity.

Proof. The profit maximization problem of the firm is

$$\max_{k_i,n_i} E_s\left(p_i k_i^{\alpha}(A_i n_i)^{1-\alpha}\right) - W n_i.$$

The first order condition of the above problem is

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

The market price of intermediate good y_i is

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

Together, this imply that for any i and i'

$$\frac{k_i}{k_{i'}} = \frac{n_i}{n_{i'}} = \frac{E_s\left(A_i^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}}}{E_s\left(A_{i'}^{(1-\alpha)\nu}\right)^{\frac{1}{1-\nu}}}.$$

Denote K as total capital and N as total labor, we get equation (A1). The output of a typical firm is

$$y_i = A_i^{1-\alpha}\frac{E_s(A_i^{(1-\alpha)\nu})^{\frac{1}{1-\nu}}}{\int E_s(A_i^{(1-\alpha)\nu})^{\frac{1}{1-\nu}}di}K^{\alpha}N^{1-\alpha}.$$

Total output of the sector is

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

Define total productivity as $A = \left[\int E_s(A_i^{(1-\alpha)\nu})^{\frac{1}{1-\nu}}di\right]^{\frac{1-\nu}{(1-\alpha)\nu}}$, and plug into the profit function and FOC, we have equation (A2). □

Proof of Lemma 1. The posterior distribution of β_i can be derived as

$$Var_s[\beta] = \frac{1}{\Delta a}\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau^2}},$$
$$E_s[\beta] = \frac{1}{\sigma^2+\tau^2}[\tau^2\mu+\sigma^2 s].$$

Then we can compute the expectation term in aggregate productivity as

$$E_s(A_i^{(1-\alpha)\nu})^{\frac{1}{1-\nu}} = \exp\left\{\left[(1-\alpha)\nu\frac{1}{\sigma^2+\tau^2}[\tau^2\mu+\sigma^2 s]+\frac{1}{2}(1-\alpha)^2\nu^2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right]\frac{1}{1-\nu}\Delta a\right\}.$$

As signal s follows a normal distribution with mean μ and variance $\frac{1}{\Delta a}[\sigma^2+\tau^2]$ across firms,

$$\int E_s(A_i^{(1-\alpha)\nu})^{\frac{1}{1-\nu}}di = \exp\left\{\left[(1-\alpha)\nu\mu+\frac{1}{2}(1-\alpha)^2\nu^2(\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}+\frac{1}{1-\nu}\frac{\sigma^4}{\sigma^2+\tau^2})\right]\frac{1}{1-\nu}\Delta a\right\}.$$

The aggregate productivity is

$$A^{1-\alpha} = \left[\int E_s(A_i^{(1-\alpha)\nu})^{\frac{1}{1-\nu}} di\right]^{\frac{1-\nu}{\nu}} = \exp\left\{\left[(1-\alpha)\mu + \frac{1}{2}(1-\alpha)^2\nu(\frac{\sigma^2\tau^2}{\sigma^2+\tau^2} + \frac{1}{1-\nu}\frac{\sigma^4}{\sigma^2+\tau^2})\right]\Delta a\right\}.$$

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

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

□

Lemma 4. *In static setup, the realized log MPK dispersion (cross-sectional variance) follows:*

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

Proof. The realized marginal product of capital is

$$MPK_i = \frac{A_i^{(1-\alpha)\nu}}{E_s(A_i^{(1-\alpha)\nu})}A^{1-\alpha}(\frac{N}{K})^{1-\alpha}.$$

The variance of realized log MPK can be computed as

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

□

A.2 Dynamic learning

In the dynamic setup, firm i productivity $\ln A_{i,t+1} = \sum_{u=0}^{t+1}\beta_{i,u}\Delta a_u$. For ease of notation, we omit the firm index i for the following analysis. For each period t, firms receive noisy signals about current period productivity exposure β_t. Old generation firms have more precise information. Specifically, mature firms (generation $\bar{n}$) learn perfect information and new firms (generation 1) learn no information. Firms also receive signals about all past productivity exposures. We use the first subscript u to denote the timing of productivity exposure and the second subscript t to denote the timing of signals. The signal follows:

$$s_{u,t} = \beta_u + \epsilon_{u,t},$$

with $\epsilon_{u,t} \sim N(0, \frac{1}{\Delta a_u}\tau_{u,t}^2)$. To summarize, each period t, mature firms receive signal $s_{t,t}$ which perfectly reveals β_t, while young firms receive a set of noisy signals $\{s_{u,t}|u \leq t\}$ on current and past exposure $\{\beta_u|u < t\}$.

If the sequence of signals $\{s_{u,u}, s_{u,u+1}, \cdots, s_{u,t+1}\}$ on β_u for young generation n satisfy:

$$\lambda_n = 1 + \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{\tau_{t+1,t+1}^{-2}/\sigma^{-2}}{\sigma^{-2} + \tau_{t+1,t+1}^{-2}}, \tag{A3}$$

$$\frac{1}{\sigma^{-2} + \sum_{v=u}^{t+1}\tau_{u,v}^{-2}} = \rho_{n,s}\frac{1}{\sigma^{-2} + \sum_{v=u}^{t}\tau_{u,v}^{-2}} \quad \text{for } u < t+1, \tag{A4}$$

then we have the recursive representation of lemma 2.

Equation (A3) specifies the relationship between contemporaneous productivity exposure and signal precision about current period idiosyncratic beta. Equation (A4) specifies the learning rate on past beta. $\frac{1}{\sigma^{-2}+\sum_{v=u}^{t+1}\tau_{u,v}^{-2}}$ relates to the posterior variance of β_u. Equation (A4) states that each period, the decay rate of posterior variance is $\rho_{n,s}$. We can also manipulate the equation to have:

$$\frac{\tau_{u,t+1}^{-2}}{\sigma^{-2} + \sum_{v=u}^{t+1}\tau_{u,v}^{-2}} = 1 - \rho_{n,s}.$$

That is to say, precision wise, firms learn a fixed proportion $1-\rho_{n,s}$ of information on past productivity.

Proof of lemma 2. At period $t+1$, after receiving a sequence of signals, the posterior distribution of β_u is:

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

From the static case, the aggregate productivity can be computed as

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

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

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

For mature generation $\bar{n}$, since they have perfect signal each period, $\tau_{t+1,t+1} = 0$, their productivity growth follows:

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

Denote $\chi_{n,t+1}$ as the productivity difference between mature generation and young generation ($n < \bar{n}$). With equation A3, we have:

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

Use equation A4, we have:

$$\chi_{n,t+1} = \rho_{n,s}\chi_{n,t} + (\lambda^* - \lambda_n)\Delta a_{t+1}.$$

The productivity growth of young generation takes the form:

$$\ln A_{n,t+1} - \ln A_{n,t} = \sum_{u=0}^{t} \frac{1}{2}(1-\alpha)\frac{\nu^2}{1-\nu}\frac{\tau_{u,t+1}^{-2}}{(\sigma^{-2} + \sum_{v=u}^{t+1} \tau_{u,v}^{-2})(\sigma^{-2} + \sum_{v=u}^{t} \tau_{u,v}^{-2})}\Delta a_u + \lambda_n \Delta a_{t+1}.$$

With equation A4 and the definition of $\chi_{n,t}$, we have

$$\ln A_{n,t+1} - \ln A_{n,t} = (1-\rho_{n,s})\chi_{n,t} + \lambda_n \Delta a_{t+1}.$$

□

Lemma 5. *In the dynamic setup, the realized log MPK dispersion (cross-sectional variance) follows:*

$$Var\left[\log(MPK_i) - \log(MPK)\right] = \sum_{u=0}^{t} (1-\alpha)^2 \nu^2 \frac{1}{\sigma^{-2} + \tau_{u,u}^{-2}} \rho_{i,s}^{t-u} \Delta a_u.$$

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

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

With equation A4, the above equation can be rewritten as

$$var\left[\log(MPK_i) - \log(MPK)\right] = \sum_{u=0}^{t}(1-\alpha)^2\nu^2 \frac{1}{\sigma^{-2} + \tau_{u,u}^{-2}}\rho_{n,s}^{t-u}\Delta a_u.$$

□

B Model solution

This section provide the details of model solution.

B.1 Social Planner's Problem

The second welfare theorem applies in this economy. The social planner solves

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

subject to the capital dynamics (14), (15), (16) and resource constraint (17). Define $F(\mathbf{A}_t, \mathbf{K}_t)$ as the aggregate output function. The FOC and the envelope condition are:

$$MC_t = E_t MV_{t+1} V_{K_1,t+1},$$
$$V_{K_n,t} = MC_t F_{K_n,t} + (1-\delta)(1-\phi)E_t MV_{t+1} V_{K_n,t+1} + (1-\delta)\phi E_t MV_{t+1} V_{K_{n+1},t+1},$$
$$V_{K_{\bar{n}},t} = MC_t F_{K_{\bar{n}},t} + (1-\delta)E_t MV_{t+1} V_{K_{\bar{n}},t+1},$$

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:

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

and the ex-dividend prices of capital to be

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

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

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

We can simplify the notation to have F.O.C:

$$1 = q_{K_1,t}.$$

Prices of capital must satisfy the recursion:

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

C Firm distribution

C.1 Capital age and capital share

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

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

where C is the 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

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

C.2 Cash flow duration

The Macaulay duration $MD_{\bar{n}}$ of capital $K_{\bar{n}}$ 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 recursively as

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

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

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

Note that our capital level duration can not be mapped directly 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 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 firm duration in the model, we define a "firm" that reinvest and postpone the cash flow from investment opportunity 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:

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

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

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

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

D Data Construction

Given that the 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 firm-level investment.[3] We follow this stream of literature to construct capital age for U.S. publicly listed companies.

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, except utility firms that have SIC 4-digit (Standard Industrial Classification) codes between 4900 and 4949 and finance firms that have SIC codes between 6000 and 6999 (finance, insurance, trusts, and real estate sectors). We also 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 are from the Bureau of Economic Analysis (BEA) maintained by the United States Department of Commerce. Patent data is from the database provided by the authors of Kogan, Papanikolaou, Seru, and Stoffman (2017), using the National Bureau of Economic Research (NBER) patent database and citation data, both of which are originally extracted from the USPTO.[4] The combined data are linked to the public firm universe using the bridge file provided by NBER, allowing us to establish the full list of patents that a firm owns at each point of time between 1926 and 2010.[5] To minimize the impact of outliers, we winsorize all variables at the 1% and 99% levels.

D.2 Capital Age

We measure capital age following the methodology in Salvanes and Tveteras (2004). We start by defining an initial measure of firm-level capital stock ($K_{i,0}$) for firm i using net

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

[4]Data on patent-level information are primarily from the updated NBER database initially constructed by Hall, Jaffe, and Trajtenberg (2001). Data on the patent in our sample are from Noah Stoffman's personal website maintained by Indiana University (`https://iu.app.box.com/v/patents`).

[5]The original database from NBER covers patent related information from 1976 to 2006.

property plant and equipment (*ppentq*) and an initial measure of firm-level capital age. The initial capital age is calculated using the ratio of accumulated depreciation and amortization (*dpactq*) over current depreciation and amortization (*dpq*), and then we recursively construct a measure of firm-level capital stock

$$K_{i,t+1} = K_{i,t} + I_{i,t}^N, \tag{D5}$$

where $I_{i,t}^N$ is net investment of firm i 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 define gross investment as

$$I_{i,t}^N = \delta_j K_{i,t} + I_{i,t}^N, \tag{D6}$$

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

After we obtain a time-series of capital stock and gross investment observations at the firm level, we follow Salvanes and Tveteras (2004) and define the capital age of firm i at time t as:

$$AGE_{i,t} = \frac{(1-\delta_j)^t K_{i,0}(AGE_{i,0}+t) + \sum_{j=0}^{t-1}(1-\delta_j)^{t-j-1} I_{i,j}(t-j)}{K_{i,t}}. \tag{D7}$$

In the above formulation, capital age at each time t is a weighted average of the age of each capital vintage. The weights are the relative importance of each capital vintage in determining total capital in place at time t. We assume that a firm always installs the newest capital when it invests so that if a firm has capital age equal to $AGE_{i,0}$ at time $t = 0$, then the time $t = 1$ capital age is a weighted average of the new installed capital vintage, which has age 1, and the old vintage capital which 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 and $I_{i,0}/K_{i,1}$ for the new vintage, where $K_{i,1} = K_{i,0} + I_{i,0}$.

For analytical convenience, we assume that when a firm disinvests, it disposes all capital vintages in proportion to their contribution to the total installed capital. Under this assumption, the expression of capital age can be written recursively as

$$AGE_{i,t} = (1-\delta_j)\frac{K_{i,t-1}}{K_{i,t}}(AGE_{i,t-1}+1) + \frac{I_{i,t-1}}{K_{i,t}}. \tag{D8}$$

Therefore, $AGE_{i,t} = AGE_{i,t-1} + 1$ when the firm has no positive investment. In addition, the above formulation implies a firm can reduce the capital age only via positive investment, consistent with our model economy.

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*).[6] 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 and financial 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 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 has to be larger than 10 to avoid biased standard deviations driven by a few extreme values.[7] 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 Other Firm Characteristics

Market capitalization is calculated 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

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

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

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 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*).

D.5 Cash Flow Duration

We construct the firm-level cash flow duration to reflect the timing of cash flows, according to the model in Lettau and Wachter (2007). Duration (Dur) is the equity implied cash flow durations measure of Dechow et al. (2004). Dechow et al. (2004) propose this measure and report a negative relationship between cash flow durations and stock returns, while Weber (2018) further studies asset pricing implications, including exposures to existing risk factors, time variation in the slope, and the effect of short-sale constraints.

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

$$Dur_{i,t} = \frac{\sum_{s=1}^{T} s \times CF_{i,t+s}/(1+r)^s}{P_{i,t}}, \tag{D9}$$

where $Dur_{i,t}$ is the duration of firm i at the end of fiscal year t, $CF_{i,t+s}$ denotes the cash flow at time $t+s$, $P_{i,t}$ is the current price, and r is the expected return on equity. The expected return on equity is constant across both stocks and time. Allowing for firm-specific discount rates *cteris paribus* amplifies cross-sectional differences in the duration measure because high-duration firms tend to have lower returns on equity. Firm-specific discounts, however, would not change the ordering, and hence had no effect on my later results. The relative ranking is also insensitive to changes in the level of r.[8]

[8]Following Weber (2018), variation over time in return on equity r does not alter the cross-sectional

Unlike fixed income securities, such as bonds, stocks do not 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 and assume the later is paid out as level perpetuity to deal with the first complication. Such assumption to rewrite the equation (D9) as

$$Dur_{i,t} = \frac{\sum_{s=1}^{T} s \times CF_{i,t+s}/(1+r)^s}{P_{i,t}} + \left(T + \frac{1+r}{r}\right) \times \frac{P_{i,t} - \sum_{s=1}^{T} CF_{i,t+s}/(1+r)^s}{P_{i,t}} \tag{D10}$$

Furthermore, we impose a clean surplus accounting, start from 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, $(B_{i,t+s} - B_{i,t+s-1})/BV_{i,t+s-1}$:

$$\begin{aligned} CF_{i,t+s} &= E_{i,t+s} - (B_{i,t+s} - B_{i,t+s-1}) \\ &= B_{i,t+s-1} \times \left[\frac{E_{i,t+s}}{B_{i,t+s-1}} - \frac{B_{i,t+s} - B_{i,t+s-1}}{B_{i,t+s-1}}\right]. \end{aligned} \tag{D11}$$

Following Dechow et al. (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 using the pooled CRSP-Compustat universe and assume ROE mean reverts to the average cost of equity. We also follow Weber (2018) to assume that growth in book equity mean reverts to the average growth rate of the economy with a coefficient of mean reversion equal to average historical mean reversion in sales growth. ROE has an AR(1) coefficient of 0.41 and BV of 0.24. We assume a discount rate r of 0.12, a steady-state average cost of equity of 0.12, an average long-run nominal growth rate of 0.06, and a detailed forecasting period of 15 years.

D.6 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)

ranking, which alleviate the concern for the cross-sectional implications.

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.6 . 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:

$$\begin{aligned} \ln y_{i,t} &= z_i + w_t + \alpha_1 \ln k_{i,t} + \alpha_2 \ln n_{i,t} + v_{i,t} + u_{i,t} \\ v_{i,t} &= \rho v_{i,t-1} + e_{i,t}, \end{aligned} \tag{D12}$$

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:

$$
\begin{aligned}
\Delta \ln y_{i,j,t} &= \rho \Delta \ln y_{i,j,t-1} + \alpha_{1,j} \Delta \ln k_{i,j,t} - \rho \alpha_{1,j} \Delta \ln k_{i,j,t-1} + \alpha_2 \Delta \ln n_{i,j,t} - \rho \alpha_2 \Delta \ln n_{i,j,t-1} \\
&\quad + (\Delta w_t - \rho w_{t-1}) + \Delta \kappa_{i,t}, \qquad \text{(D13)}
\end{aligned}
$$

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:

$$
\begin{aligned}
E[x_{i,i,t-l} \Delta \kappa_{i,t}] &= 0 \text{ for } l \geq 3 \\
E[x_{i,j,t-l} \Delta \kappa_{i,t}] &= 0 \text{ for } l \geq 3. \qquad \text{(D14)}
\end{aligned}
$$

that are used to conduct a consistent GMM estimation of equation (D13). Given the estimates $\widehat{\alpha}_{1,j}$ and $\widehat{\alpha}_{2,j}$, log productivity of firm i is computed as

$$
\ln \widehat{a}_{i,j,t} = \ln y_{i,j,t} - \widehat{\alpha}_{1,j} \ln k_{i,j,t} - \widehat{\alpha}_{2,j} \ln n_{i,j,t}, \qquad \text{(D15)}
$$

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:

$$
\ln y_{i,j,t} = v_j + z_{i,j} + w_{j,t} + \alpha_{1,j} \ln k_{i,j,t} + \alpha_{2,j} \ln n_{i,j,t} + u_{i,j,t}. \qquad \text{(D16)}
$$

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 (D15) to estimate $\widehat{a}_{i,j,t}$. Given this estimation of firms' productivity, we obtain the alternative estimation of firms' productivity.

E Additional Empirical Evidence

In this section, we provide additional empirical evidence, univariate portfolios sorted on the capital age, asset pricing tests, and more detailed firm characteristics.

E.1 Asset Pricing Test

We implement the standard procedure and sort firms into quintile portfolios based on these firms' capital ages within Fama-French 30 industries. At the beginning of January, April, July, and October, we rank firms' capital ages by using 30 industry-specific breaking points based on Fama and French (1997) classifications and construct portfolios as follows. We sort firms with a positive capital age in previous six months into five groups from low to high. 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. After six portfolios (from low to high and long-short portfolios) are determined, we calculate the value-weighted monthly returns annualized by multiplying 12 and hold these portfolios over the next three months.

Panel A of Table E1 reports the average annualized excess returns and Sharpe ratios in five quintile portfolios and long-short portfolio. The spread of long-short (OMY) portfolio is economically large (5.79% per annum) and statistically significant at 1% level with t-statistics close to 3. The annualized Sharpe ratio is economically sizable, amounting to 0.44, which is almost comparable to that of the aggregate stock market index (around 0.5). We call the return spread of OMY as the capital age premium.

We investigate the extent to which the variation in average returns of the capital age sorted portfolios can be explained by existing risk factors. We then examine whether the capital age-return relation reported in Panel A reflects firms' exposures to the existing systematic risk factors by preforming time-series regressions of capital age sorted portfolios' excess returns on the Fama and French (2015) five-factor model (the market factor-MKT, the size factor-SMB, the value factor-HML, the profitability factor-RMW, and the investment factor-CMA) in Panel B and on the Hou, Xue, and Zhang (2015) q-factor model (the market factor-MKT, the size factor-SMB, the investment factor-I/A, and the profitability factor-ROE) in Panel C, respectively.[9] 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 %).

[Insert Table E1 Here]

We make several observations. First, the risk-adjusted returns (intercepts) of the capital age sorted old-minus-young (OMY) portfolio remain large and significant, ranging from 4.63% for Fama and French (2015) five-factor model to 3.31% from the Hou, Xue, and Zhang (2015) q-factor model, and these intercepts are 2.93 and 1.80 standard errors away

[9]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.

from zero, as reported in the t-statistics above 1% and 10% statistical significant level, respectively. Second, the alpha implied by the Fama-French five-factor model is slightly lower than the the capital age sorted portfolio spread in the univariate sorting (Panel A), while the alpha implied by HXZ q-factor model remains half of the long-short portfolio sorted on capital age. Third, our old-minus-young portfolio have insignificantly negative betas with respect to the value factor in the Fama and French (2015) five-factor model. The old-minus-young (OMY) portfolio presents negative loadings on market and size factors, but positive loadings on profitability and investment factors for Fama-French five-factor model (Panel B). Similarity, the OMY portfolio presents negative loadings on market and size factors, but positive loadings on profitability and investment factors for HXZ q-factor model (Panel C). Loads on market, size, profitability, and investment are statistically significant, which suggest firms with old capital bear less exposure to market risk, are larger in size, earn more profits, and incur less investment than firms with young capital. In summary, results from asset pricing tests in Table E1 suggest that the cross-sectional return spread across capital age sorted portfolios cannot be completely explained by either the Fama and French (2015) five-factor or the HXZ q-factor model (Hou et al. (2015)).

E.2 More Detailed Firm Characteristics

Table E2 documents how differences in firms' capital age are related to other characteristics. We report average capital age and other characteristics across five quintiles sorted on capital age for financially constrained firms.

[Insert Table E2 Here]

On average, our sample contains 2345 firms. Firms are evenly distributed across five capital age sorted portfolios, where the average number of firms in each portfolio ranges from 469 to 480. The cross-sectional variations in capital age are large, ranging from 9.71 to 35.95 across five portfolios. Both size and book-to-market ratio (B/M) increases with capital age sorted portfolios. Moreover, firms with a lower capital age are prone to have a higher investment rate (I/K) to reflect more investment opportunities, which is consistent with the pattern of book-to-market ratio. That is, high capital age firms are likely to be value firms with less investment opportunities. In addition, we observe a downward sloping pattern of profitability (ROE). Firms with higher capital age are less profitable than firms with lower capital age. On the other hand, there is a positive relationship between capital age and tangibility.

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

Table E1. 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 E2. 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 July 1988 to December 2015 and excludes utility, financial, and R&D intensive industries from the analysis. The detailed definition of the variables refers to Appendix.

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	0.11	0.09	0.08	0.07	0.06
TANG	0.31	0.35	0.38	0.42	0.44
Number of Firms	480	469	471	467	458

References

Ai, Hengjie, Mariano Massimiliano Croce, and Kai Li, 2012, 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 Amir Yaron, 2004, Risks for the Long Run: A Potential Resolution, *Journal of Finance* 59, 1481–1509.

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.

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 .

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 .

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, 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.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg, 2001, The nber patent citation data file: Lessons, insights and methodological tools, Technical report, National Bureau of Economic Research.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting Anomalies: An Investment Approach, *Review of Financial Studies2* 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, in *Fifty years of economic measurement: The jubilee of the Conference on Research in Income and Wealth*, 119–158, University of Chicago Press.

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.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, 2017, Technological innovation, resource allocation, and growth, *The Quarterly Journal of Economics* 132, 665–712.

Kreps, David M, and Evan L Porteus, 1978, Temporal resolution of uncertainty and dynamic choice theory, *Econometrica: journal of the Econometric Society* 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.

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, 2017, The risks of old capital age: Asset pricing implications of technology adoption .

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.

Pastor, Lubos, and Pietro Veronesi, 2009, Learning in financial markets, *Annu. Rev. Financ. Econ.* 1, 361–381.

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

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

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* .

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