

# DISCUSSION PAPER SERIES

DP12661

## **NEWS SHOCKS AND THE PRODUCTION- BASED TERM STRUCTURE OF EQUITY RETURNS**

Hengjie Ai, Mariano Massimiliano Croce, Anthony  
Diercks and Kai Li

**FINANCIAL ECONOMICS**



# NEWS SHOCKS AND THE PRODUCTION-BASED TERM STRUCTURE OF EQUITY RETURNS

*Hengjie Ai, Mariano Massimiliano Croce, Anthony Diercks and Kai Li*

Discussion Paper DP12661

Published 29 January 2018

Submitted 29 January 2018

Centre for Economic Policy Research  
33 Great Sutton Street, London EC1V 0DX, UK  
Tel: +44 (0)20 7183 8801  
[www.cepr.org](http://www.cepr.org)

This Discussion Paper is issued under the auspices of the Centre's research programme in **FINANCIAL ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Hengjie Ai, Mariano Massimiliano Croce, Anthony Diercks and Kai Li

# NEWS SHOCKS AND THE PRODUCTION-BASED TERM STRUCTURE OF EQUITY RETURNS

## Abstract

We propose a production-based general equilibrium model to study the link between timing of cash flows and expected returns both in the cross section of stocks and along the aggregate equity term structure. Our model incorporates long-run growth news with time-varying volatility and slow learning about the exposure that firms have with respect to these shocks. Our framework provides a unified explanation of the stylized features of the slope of the term structure of equity returns, its variations over the business cycle, and the negative relationship between cash-flow duration and expected returns in the cross section of book- to-market-sorted portfolios.

JEL Classification: E2, E3, G1

Keywords: N/A

Hengjie Ai - [hengjie.ai@gmail.com](mailto:hengjie.ai@gmail.com)  
*Carlson School of Management*

Mariano Massimiliano Croce - [mmc287@gmail.com](mailto:mmc287@gmail.com)  
*The University of North Carolina and CEPR*

Anthony Diercks - [antoined11@gmail.com](mailto:antoined11@gmail.com)  
*Board of Governors of the Federal Reserve System*

Kai Li - [kaili825@gmail.com](mailto:kaili825@gmail.com)  
*Hong Kong University of Science and Technology*

# News Shocks and the Production-Based Term Structure of Equity Returns

Hengjie Ai      M. Max Croce      Anthony M. Diercks      Kai Li\*

## Abstract

We propose a production-based general equilibrium model to study the link between timing of cash flows and expected returns both in the cross section of stocks and along the aggregate equity term structure. Our model incorporates long-run growth news with time-varying volatility and slow learning about the exposure that firms have with respect to these shocks. Our framework provides a unified explanation of the stylized features of the slope of the term structure of equity returns, its variations over the business cycle, and the negative relationship between cash-flow duration and expected returns in the cross section of book-to-market-sorted portfolios.

*Keywords:* Production Economy, Term Structure, Learning

*First Draft:* October 20, 2012. *This Draft:* September 4, 2017

*JEL Classification:* E2, E3, G1

---

\*Hengjie Ai is an associate professor at the Carlson School of Management, University of Minnesota (hengjie.ai@gmail.com). Mariano Massimiliano Croce is an associate professor at the Kenan-Flagler Business School, UNC-Chapel Hill (mmc287@gmail.com). Anthony M. Diercks is an economist at the Board of Governors of the Federal Reserve System (Anthony.M.Diercks@frb.gov). Kai Li is an assistant professor at the HKUST School of Business and Management (kaili825@gmail.com). We thank the editor, Itay Goldstein, and two anonymous referees. We also thank Andrew Abel, Ravi Bansal, Joao Gomes, Monika Piazzesi, Nick Roussanov, Lukas Schmid, and Amir Yaron for their helpful comments on our article. We are grateful to our discussants Frederico Belo, Adlai Fisher, John Heaton, Jun Li, Francisco Palomino, Hakon Tretvoll, Jules van Binsbergen, and Fang Yang. We also thank seminar participants at the AFA meetings, the AEA meetings, the WFA meetings, the EFA conference, the Carlson School of Management (UMN) Macro-Finance Conference, the Finance Cavalcade Conference, the China International Conference in Finance, the Wharton School (University of Pennsylvania), the STERN School of Business (NYU), the Kenan-Flagler Business School (UNC), London School of Economics, the BI Norwegian Business School, the Hong Kong Joint Finance Research Conference, the Ross School of Business (University of Michigan), and the Fuqua School of Business (Duke University).

# 1 Introduction

The link between the timing of equity cash flows and equity expected returns has been studied both in the cross section of stocks and in the aggregate. Working with aggregate cash-flow strips, Binsbergen et al. (2012), Binsbergen et al. (2013), and Binsbergen and Koijen (2016, 2017) document several stylized facts on the term structure of equity returns, that is, the relationship between the return on claims to aggregate dividend strips and their maturity. First, the slope of the term structure varies substantially over time and is significantly negative in the Great Recession. Second, the returns on short-term dividend claims have higher volatility but lower market beta than an index on aggregate dividends. Third, the CAPM  $\beta$ s of claims to aggregate dividends are countercyclical and this time variation of  $\beta$ s decreases with maturity. In addition, the literature on the cross section of expected returns documents a negative relationship between cash-flow duration and expected returns in book-to-market-sorted portfolios (Da (2009), Dechow et al. (2004)).

In this article, we propose a novel production-based model to provide a unified explanation of the relationship between the timing of cash flows and their expected returns both for the aggregate stock market dividends and for the cross section of book-to-market-sorted portfolios. We construct a vintage capital model in which individual firms have imperfect information about their productivity and have to learn about it over time. In this setting, the endogenous response of firms' investment and payout to news about future productivity can explain the relationship between cash-flow duration and risk premia documented in the above literature.

While the term structure of real interest rates is determined by the properties of the stochastic discount factor alone, the term structure of equity returns depends on the dynamics of both the stochastic discount factor and that of the cash-flow process. In endowment economy models (see, among others, Lettau and Wachter (2007, 2011) and Santos and Veronesi (2010)), dividends are exogenously specified. On the other hand, in investment-based partial equilibrium models (for example, those of Zhang (2005); Liu et al. (2009); Lin and Zhang (2007)) the stochastic discount

factor is often taken as given. Therefore, the empirical evidence on the term structure of equity returns does not provide a direct discipline on the aforementioned models. However, the negative relationship between dividend maturity and expected returns during recessions does provide a litmus test of general equilibrium production models in which both payouts and the pricing kernel are simultaneously and endogenously determined.

We start by showing that the equity term structure evidence constitutes a challenge in a large class of neoclassical growth models (henceforth RBC). In a setting with production, the total payout to investors is given by

$$\text{Payout} = (1 - \text{Labor Share}) \times \text{Output} - \text{Cost of Investment}.$$

In the data, the volatility of labor share is fairly small, about 2% per year (Choi and Ruis-Rull (2009)), and is therefore calibrated to be constant in most RBC models. However, RBC models produce an investment process that is highly volatile and procyclical with respect to contemporaneous productivity shocks. As a result, over the short horizon, investment acts like a hedge and the total payout is countercyclical. This endogenous correlation structure implies that short-maturity dividend strips command a negative risk premium and that the term structure of equity returns is unambiguously upward sloping along the cycle. Both implications are inconsistent with the empirical evidence on equity term structure.<sup>1</sup>

Our model resolves the above puzzle by building on the long-run risk framework of Bansal and Yaron (2004). In our model, investment responds strongly and positively to contemporaneous productivity shocks, as in the data. However, its reaction to news about future productivity shocks is negative upon impact. As a result, the total payout to the household increases upon the arrival of good long-run news. Therefore, the impulse response to contemporaneous productivity

---

<sup>1</sup>Boguth et al. (2012) point out some limitations of the empirical evidence in Binsbergen et al. (2012, 2013). The implications for the term structure of equity obtained from the standard RBC model are strongly rejected even under the most conservative interpretation of the empirical evidence.

shocks leads to an upward-sloping term structure of dividends, while the response of investment to news shocks provides a mechanism for a high risk premium for short-term dividend strips and a downward-sloping term structure over short maturities that is absent in the RBC model.

Guided by the above theoretical insights, we provide novel evidence for the time-varying relative volatility of news shocks and contemporaneous productivity shocks that can account for the time variation in the slope of term structure. We show that the volatility of the persistent component of productivity shocks exhibits substantial variation over time, and it peaks during recessions. When incorporated into our model, our productivity-based volatility factors produce a procyclical term structure slope which turns negative during severe recessions, consistent with the data. In addition, the presence of two risk factors, long-run and short-run productivity risks, allows our model to capture the failure of CAPM in the data.

The key feature of our model, that is, the fact that investment responds negatively to news shocks upon impact, is due to a novel learning mechanism. In our economy, firms have heterogeneous exposure to aggregate productivity shocks. Adolescent firms have limited information about their exposure to aggregate shocks but receive noisy signals from which they learn over time. Adolescent firms therefore are less capable of taking advantage of advances in aggregate productivity, and the correlation between their output and aggregate productivity is lower than that of firms with full information.

In this setup, upon a positive news shock about aggregate productivity, investment does not immediately increase because the learning mechanism dampens the substitution effect: new investment creates adolescent firms that are less capable of taking advantage of technological progress and hence do not represent appealing investment opportunities. At the same time, most of the existing mature firms have full information, and their productivity is expected to rise in the future because they can take full advantage of the new productivity frontier. The income effect therefore is positive and dominates. As a result, consumption increases but investment falls upon positive news shocks. Over time, as news materializes and productivity eventually goes up, so

does investment.

As in Ai et al. (2012), we model assets in place as physical capital, and the stock of new business ideas and investment opportunities as intangible capital. This setup allows us to micro-found value stocks (stocks with a high book-to-market ratio) as the claim to physical asset-intensive firms and growth stocks (stocks with a low book-to-market ratio) as the equity of intangible capital-intensive firms.

Because physical capital and intangible capital are complements, the negative response of physical investment with respect to positive news shocks is associated with drops in the payoff to claims to intangible capital. Equivalently, intangible capital provides insurance against news shocks, and hence it commands a lower risk premium, consistent with the value premium empirical evidence. As a result, our model also accounts for the negative relationship between cash-flow duration and expected returns in the cross section.

**Related literature.** Several other papers have proposed alternative economic channels for the downward-sloping term structure of equity returns. In endowment economies, Andries et al. (2017) focus on the preference side and propose horizon-dependent risk aversion as an explanation for the term structure of equity risk compensation. Croce et al. (2015) obtain a downward-sloping term structure in a long-run risk model with limited information and bounded rationality. Hasler and Marfe (2017) present a rare-disaster model with recursive preferences and study their implications on the term structure of interest rates and the term structure of dividends. Belo et al. (2014) study the implications of capital structure and corporate payout decisions on the term structure of equity returns.

In production economies, Kogan et al. (2013) show that their model with investment-specific shocks is also consistent with the negative slope of the term structure of equity returns.<sup>2</sup> Favilukis and Lin (2013) and Marfe (2017) produce a downward-sloping term structure of equity returns by

---

<sup>2</sup>See also Papanikolaou (2011) and Kogan and Papanikolaou (2014, 2010, 2012).

means of wage rigidity and a time-varying labor share.

Our paper is also related to the literature on the cross section of equity returns, specifically the value premium. Berk et al. (1999), Gomes et al. (2003), Carlson et al. (2004), and Cooper (2006) propose equilibrium models of the value premium by explicitly modeling the heterogeneous risk exposure of assets in place and growth options. Zhang (2005) and Gala (2005) focus on models of adjustment cost. Dechow et al. (2004), and Da (2009) provide empirical evidence on the difference in cash-flow duration for value versus growth stocks.

None of the above-mentioned papers focuses on the variations of the slope of the term structure of dividends over the business cycle or links it to the empirical evidence on the time-varying volatility of productivity news shocks, nor do they study the link between cash-flow duration and expected returns along the aggregate equity term structure and in the cross section of stocks jointly.

The learning mechanism that we emphasize in this paper is related to the literature that studies the impact of learning on asset market valuations. David (1997) and Veronesi (2000) study how learning and information affect both asset valuations and the risk premium on the equity market. Pastor 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 remainder of the paper is organized as follows. In the next section, we present our model. In section 3, we study the term structure of equity returns in our setting with learning, and we compare it to a neoclassical RBC model with convex adjustment costs. In section 4, we introduce our empirical evidence on time-varying productivity-based volatility factors and explore the implications of our model for the dynamics of the term structure slope. We test novel predictions of our model in section 5 and conclude in section 6.

## 2 Model Setup

The key element of our model is that firms learn about their exposure to aggregate productivity over time. In equilibrium, heterogeneity in information translates into heterogeneity in risk exposures. In this section, we first describe a tractable analytical framework that models learning with heterogeneous productivity. We then incorporate learning into a general equilibrium model with production and derive the equilibrium conditions.

### 2.1 Aggregation with learning

We provide aggregation results supporting the key learning mechanism of our model, that is, that when firms are uncertain about their exposure to aggregate productivity shocks, more information allows them to take better advantage of aggregate technological progress, and therefore they feature a high exposure to aggregate shocks.

**The static problem.** We start with a static setup similar to that of Melitz (2003) and Hsieh and Klenow (2009). Consider a group of firms that produce intermediate inputs,  $y_j$ , that can be transformed into output  $Y$  using a CES production function:

$$Y = \left[ \int (y_j)^{\frac{\eta-1}{\eta}} dj \right]^{\frac{\eta}{\eta-1}}, \quad (1)$$

Firm  $j$  combines capital and labor to produce output using a Cobb-Douglas production technology,

$$y_j = A_j k_j^\alpha n_j^{1-\alpha}. \quad (2)$$

We assume that  $A_j = e^{\beta_j \Delta a}$ , where  $\Delta a$  is a common shock that affects the productivity of all firms and  $\beta_j$  is the firm-specific exposure to the common shock  $\Delta a$ . To facilitate a closed-form solution, we assume that conditioning on the common shock  $\Delta a$ ,  $\beta_j \sim i.i.d.N(\mu, \frac{1}{\Delta a} \sigma^2)$ . Before making its

production decision, firm  $j$  observes only a noisy signal of its own exposure,  $sig_j$ :

$$sig_j = \beta_j + \varepsilon_j, \quad (3)$$

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

We define the aggregate production function as

$$F(K, N) \equiv \max_{\{k_j, n_j\}} \left[ \int (A_j k_j^\alpha n_j^{1-\alpha})^{\frac{\eta-1}{\eta}} dj \right]^{\frac{\eta}{\eta-1}} \quad (4)$$

$$\text{subject to : } \int k_j = K, \quad (5)$$

$$\int n_j = N, \quad (6)$$

where for each  $j$ , the choices of  $\{k_j, n_j\}$  must be measurable with respect to firm  $j$ 's information. That is,  $k_j$  and  $n_j$  can only be functions of the signal  $sig_j$ . In Lemma 3 in the appendix, we prove that the optimality of resource allocation implies that the aggregate production can be written as  $Y = \mathbf{A} K^\alpha N^{1-\alpha}$ , where

$$\mathbf{A} = \left[ \int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds \right]^{\frac{1}{\eta-1}}. \quad (7)$$

For simplicity, we assume that  $\mu = 1 - \frac{1}{2} \frac{(\eta-1)}{\eta} \sigma^2$  throughout the paper. As we show below, this is a normalization assumption that implies that the exposure of aggregate productivity to  $\Delta a$  is 1 in the case of no information ( $\tau = \infty$ ). The following lemma provides the functional form of the aggregate productivity  $\mathbf{A}$  and analyze its elasticity with respect to the common shock  $\Delta a$ .

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

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

where  $\ln \mathbf{A} = \lambda(\tau^2) \Delta a$ , and  $\lambda(\tau^2)$  is defined as

$$\lambda(\tau^2) \equiv \left( 1 + \frac{1}{2} \frac{(\eta - 1)^2}{\eta} \frac{\sigma^4}{\sigma^2 + \tau^2} \right). \quad (8)$$

The exposure of aggregate productivity to common shocks,  $\lambda(\tau^2)$ , is decreasing in the amount of noise in the signals,  $\tau^2$ . In addition,

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

and

$$\lambda^* \equiv \lim_{\tau^2 \rightarrow 0} \lambda(\tau^2) = 1 + \frac{1}{2} \frac{(\eta - 1)^2}{\eta} \sigma^2. \quad (9)$$

**Proof.** See appendix A. ■

The result of the above lemma is intuitive. Better information allows firms to allocate capital and labor more efficiently across each other and, as a result, the level of the aggregate productivity shock  $\mathbf{A}$  increases with information precision because  $\lambda(\tau^2)$  is decreasing in  $\tau^2$ . The upper bound on the exposure is attained under full information and is denoted by  $\lambda^*$ . Below we extend the above setup to a multiperiod setting.

**The infinite-horizon setting.** To build up our fully dynamic model, we first present an aggregation result where firm  $j$  productivity,  $A_{j,t}$ , is determined by the following stochastic growth process:

$$A_{j,t} = \exp \left[ \sum_{s=0}^t \beta_{j,s} \Delta a_s \right], \quad (10)$$

where  $\{\Delta a_s\}_{s=0}^t$  is a sequence of shocks common across all firms. For  $s = 0, 1, \dots, t$ ,  $\beta_{j,s}$  is the exposure of firm  $j$ 's productivity with respect to the common shock  $\Delta a_s$ . Here we assume that each firm observes only one signal about its own exposure to the current-period common productivity shock,  $\beta_{j,t}$ , in every period  $t$ . For simplicity, we also assume that firms learn only from the observed signals but not from the history of realized output.

To avoid keeping track of the distribution of firms with heterogenous information, we assume that there are two groups of firms: (i) mature firms, which know the exact value of  $\{\beta_{j,s}\}_{s=0}^t$ , and (ii) adolescent firms, which in each period  $t$ , observe a noisy signal about  $\beta_{j,t}$ . Let  $\hat{K}$  denote the total measure of mature firms and  $\bar{K}$  be the total measure of adolescent firms. Also, let  $\hat{N}$  and  $\bar{N}$  denote the total labor input of mature firms and adolescent firms, respectively.

Extending the setup of the static model to the dynamic environment, we assume that the distribution of  $\beta_{j,t}$  is *i.i.d.* across  $j$  and  $t$ , and follows a normal distribution with mean  $\mu = 1 - \frac{1}{2} \frac{(\eta-1)}{\eta} \sigma^2$  and variance  $\frac{1}{\Delta a_t} \sigma^2$ . In each period  $t$ , adolescent firms observe a signal for  $\beta_{j,t}$  of the following form:

$$sig_{j,t} = \beta_{j,t} + \varepsilon_{j,t}, \quad (11)$$

where  $\varepsilon_{j,t} \sim N(0, \frac{1}{\Delta a_t} \tau_t^2)$  for all  $t$ . Given these assumptions, we can derive the posterior distribution of  $A_{j,t}$  and apply equation (7) to recover the aggregate production function, which is given by the following lemma.

**Lemma 2** *The total output of all mature firms,  $\hat{Y}_t$ , is*

$$\hat{Y}_t = \hat{\mathbf{A}}_t \hat{K}_t^\alpha \hat{N}_t^{1-\alpha}, \quad \text{with} \quad \ln \hat{\mathbf{A}}_t = \exp \left[ \sum_{s=0}^t \lambda^* \Delta a_s \right], \quad (12)$$

where  $\lambda^*$  is defined as in Lemma 1. The total output of all adolescent firms,  $\bar{Y}_t$ , is

$$\bar{Y}_t = \bar{\mathbf{A}}_t \bar{K}_t^\alpha \bar{N}_t^{1-\alpha}, \quad \text{with} \quad \ln \bar{\mathbf{A}}_t = \exp \left[ \sum_{s=0}^t \lambda (\tau_s^2) \Delta a_s \right], \quad (13)$$

where  $\lambda(\cdot)$  is defined as in equation (8).

**Proof.** See appendix A. ■

The above lemma highlights the basic learning mechanism in our model. Because adolescent firms have less information about their  $\{\beta_{j,s}\}_{s=0}^t$ , they are less capable of taking advantage of technological progress, and hence their aggregate productivity has a lower elasticity with respect to common shocks than that of mature firms ( $\lambda(\tau_s^2) \leq \lambda^*, \forall s$ ).

Under our specification, equation (12) implies that the law of motion of the productivity of mature firms satisfies

$$\ln \hat{\mathbf{A}}_{t+1} - \ln \hat{\mathbf{A}}_t = \lambda^* \Delta a_{t+1}, \quad (14)$$

and the law of motion of adolescent firms is

$$\ln \bar{\mathbf{A}}_{t+1} - \ln \bar{\mathbf{A}}_t = \lambda(\tau_t^2) \Delta a_{t+1}. \quad (15)$$

**Aggregation with perpetual learning.** It is clear from the above discussion that adolescent firms are less sensitive to aggregate productivity shocks because they do not fully observe their exposures,  $\{\beta_{j,s}\}_{s=0}^t$ . If we allow for long-term growth, that is,  $E[\Delta a_{t+1}] > 0$ , with  $\lambda(\tau_s^2) < \lambda^* \forall s$ , the lower exposure to aggregate productivity implies that adolescent firms will be less productive than mature firms on average. With  $E[\Delta a_{t+1}] > 0$ , equations (14) and (15) together imply that this difference will accumulate over time, and the economy cannot have a balanced growth path.

To guarantee balanced growth, we keep the specification of productivity in equation (10) 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  $t$ . We describe this process below:

- In period 0,  $\ln A_{j,0} = \beta_{j,0} \Delta a_0$ , where  $\beta_{j,0} \sim N(\mu, \frac{1}{\Delta a_0} \sigma^2)$  and adolescent firms have no additional information about their  $\beta_{j,0}$ .
- In period 1,  $\ln A_{j,1} = \beta_{j,0} \Delta a_0 + \beta_{j,1} \Delta a_1$ , where  $\beta_{j,s} \sim N(\mu, \frac{1}{\Delta a_s} \sigma^2)$  and  $s = 0, 1$ . Each

adolescent firm observes a signal,  $s_0^1 = \beta_{j,0} + \varepsilon_0^1$ , where  $\varepsilon_0^1 \sim N(0, \frac{1}{\Delta a_0} \tau_0^2)$  to update its belief about  $\beta_{j,0}$  and lower its posterior variance to  $\frac{1}{\Delta a_0} \frac{1}{\sigma^{-2} + \tau_0^{-2}}$ .

- In period 2,  $\ln A_{j,1} = \beta_{j,0} \Delta a_0 + \beta_{j,1} \Delta a_1 + \beta_{j,2} \Delta a_2$ , where  $\beta_{j,s} \sim N\left(\mu, \frac{1}{\Delta a_s} \sigma^2\right)$  and  $s = 0, 1, 2$ . Each adolescent firm observes a signal on  $\beta_{j,1}$ ,  $s_1^1 = \beta_{j,1} + \varepsilon_1^1$ , where  $\varepsilon_1^1 \sim N(0, \frac{1}{\Delta a_1} \tau_0^2)$  to lower its posterior variance to  $\frac{1}{\Delta a_0} \frac{1}{\sigma^{-2} + \tau_0^{-2}}$ . In addition, under perpetual learning, this firm also receives a signal about its previous exposure,  $\beta_{j,0}$ ,  $s_0^1 = \beta_{j,0} + \varepsilon_0^1$ , where  $\varepsilon_0^1 \sim N(0, \frac{1}{\Delta a_0} \tau_1^2)$ . As a result, it lowers further the posterior variance of  $\beta_{j,0}$  to  $\frac{1}{\Delta a_0} \frac{1}{\sigma^{-2} + \tau_0^{-2} + \tau_1^{-2}}$ .

Similarly, for  $t = 3, 4, \dots$ , each adolescent firm observes a sequence of signals,  $s_s^t = \beta_{j,s} + \varepsilon_s^t$ , where  $\varepsilon_s^t \sim N(0, \frac{1}{\Delta a_s} \tau_{t-s-1}^2)$  for  $s = 0, 1, \dots, t-1$  and updates its beliefs on all previous exposure coefficients,  $\{\beta_{j,s}\}_{s=0}^{t-1}$ .

In this setup, over time, firms will be constantly learning their exposures and improving their productivity, which allows us to modify equation (15) and ensure balanced growth. In appendix A, we show that the sequence  $\{\tau_t\}_{t=0}^{\infty}$  can be specified as a function of a parameter  $\rho_s \in (0, 1)$  so that the ratio between the productivity of adolescent firms and that of mature firms,  $\chi_{t+1} \equiv \ln\left(\frac{\hat{\mathbf{A}}_{t+1}}{\bar{\mathbf{A}}_{t+1}}\right)$ , is stationary and follows an AR(1) process:

$$\chi_{t+1} = \ln\left(\frac{\hat{\mathbf{A}}_{t+1}}{\bar{\mathbf{A}}_{t+1}}\right) = \rho_s \chi_t + (\lambda^* - 1) \Delta a_{t+1}. \quad (16)$$

In addition, the law of motion of  $\bar{\mathbf{A}}_t$  can be specified recursively with  $\chi_t$  being the only state variable:

$$\ln \bar{\mathbf{A}}_{t+1} - \ln \bar{\mathbf{A}}_t = (1 - \rho_s) \chi_t + \Delta a_{t+1}. \quad (17)$$

Together with (14), the above two equations fully specify the aggregate productivity of adolescent firms and mature firms.

**Summary of the micro-foundation of learning.** At the micro-level,  $\sigma$  and the sequence  $\{\tau_t^2\}_{t=0}^\infty$  are the primitive parameters of the model. The parameter  $\sigma$  is the dispersion of firms' exposure to the economy-wide common productivity. Intuitively, higher dispersion implies more benefit of reallocating resources across firms. As shown in equation (9), this implies that mature firms who have complete information about  $\{\beta_{j,s}\}_{s=0}^t$  are more exposed to aggregate shocks.

Thanks to perpetual learning, adolescent firms can eventually obtain full information about their exposures. This condition rules out permanent gaps between the productivity of adolescent and mature firms and guarantees balanced growth. As shown in equation (A8)–(A9), the sequence of variances of the signals,  $\{\tau_t^2\}_{t=0}^\infty$ , is increasing in  $\rho_s$ . Intuitively, higher values of  $\tau_t^2$  imply that adolescent firms' information is less precise and, as a result, the productivity gap between adolescent firms and mature firms can persist for many periods.

In our quantitative exercise, we do not directly specify the micro parameters  $\sigma$  and  $\{\tau_t^2\}_{t=1}^\infty$ . Rather, we calibrate the macro parameters  $\lambda^*$  and  $\rho_s$  from empirical evidence on the difference in the exposure of young and old firms with respect to aggregate productivity shocks.

Finally, given the dynamics of the productivity of adolescent and mature firms,  $\bar{\mathbf{A}}_t$  and  $\hat{\mathbf{A}}_t$ , we specify aggregation production as the solution to the following optimal resource allocation problem:

$$F(\hat{\mathbf{A}}_t, \bar{\mathbf{A}}_t, \hat{K}_t, \bar{K}_t, N_t) = \max_{\hat{N}_t, \bar{N}_t} \left\{ \hat{\mathbf{A}}_t \hat{K}_t^\alpha \hat{N}_t^{1-\alpha} + \bar{\mathbf{A}}_t \bar{K}_t^\alpha \bar{N}_t^{1-\alpha} \right\} \quad (18)$$

*subject to* :  $\hat{N}_t + \bar{N}_t = N_t$ .

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  $Y_t = F(\hat{\mathbf{A}}_t, \bar{\mathbf{A}}_t, \hat{K}_t, \bar{K}_t, N_t)$ , where the law of motion of productivity is given by (14), (17), and (16).

## 2.2 The full model

**Preferences.** Time is discrete and infinite, with  $t = 1, 2, 3, \dots$ . The representative agent has Kreps and Porteus (1978) preferences, as in Epstein and Zin (1989):

$$V_t = \left\{ (1 - \beta) u(C_t, N_t)^{1 - \frac{1}{\psi}} + \beta (E_t [V_{t+1}^{1-\gamma}])^{\frac{1-\frac{1}{\psi}}{1-\gamma}} \right\}^{\frac{1}{1-\frac{1}{\psi}}}, \quad (19)$$

where  $C_t$  and  $N_t$  denote, respectively, the total consumption and total hours worked at time  $t$ . For simplicity, we consider a Cobb-Douglas aggregator for consumption and leisure:

$$u(C_t, N_t) = C_t^o (1 - N_t)^{1-o}.$$

We normalize  $N_t = 1$  in the case of inelastic labor supply, i.e., when  $o = 1$ .

**Output producers.** The specification of aggregate output and individual firm output are as summarized in equations (14), (17), (16), and (18). Following the long-run risks literature, we specify the stochastic process for the common productivity  $\Delta a_t$  as follows:

$$\begin{aligned} \Delta a_{t+1} &= \mu + x_t + e^{\sigma_a} \varepsilon_{a,t+1}, \\ x_{t+1} &= \rho x_t + e^{\sigma_x} \varepsilon_{x,t+1}, \\ \begin{bmatrix} \varepsilon_{a,t+1} \\ \varepsilon_{x,t+1} \end{bmatrix} &\sim i.i.d.N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right), \quad t = 0, 1, 2, \dots \end{aligned} \quad (20)$$

According to the above specification, short-run productivity shocks,  $\varepsilon_{a,t+1}$ , affect contemporaneous output directly but have no effect on future productivity growth. Shocks to long-run productivity, represented by  $\varepsilon_{x,t+1}$ , carry news about future productivity growth rates but do not affect current output. The log standard deviations of these shocks,  $\sigma_a$  and  $\sigma_x$ , are constant over time.

**Firm dynamics.** For parsimony, we assume that all firms are subject to the same exit rate,  $\delta$ . For tractability, we assume that in each period the surviving adolescent firms,  $(1 - \delta)\bar{K}_t$ , become mature with a constant probability  $\phi$ . Under this assumption, the law of motion of the mass of mature firms,  $\hat{K}$ , can be written as

$$\hat{K}_{t+1} = (1 - \delta) \hat{K}_t + (1 - \delta) \phi \bar{K}_t.$$

Note that in our setup, maturity and age are positively but not perfectly correlated. The parameter  $\phi$  determines the speed of transition probability from adolescence to maturity in each period.

New firms are created by combining ideas and physical investment goods. We use  $S_t$  to denote the total measure of ideas or, equivalently, the total stock of intangible capital at time  $t$ . As in Ai et al. (2012), the total measure of new firms that can be created with total investment  $I_t$  is determined by a concave and constant return-to-scale production function  $G(I_t, S_t)$ .

Under these conditions, the law of motion of the total measure of young firms,  $\bar{K}_t$ , can be written as

$$\bar{K}_{t+1} = (1 - \delta) (1 - \phi) \bar{K}_t + G(I_t, S_t). \quad (21)$$

**Intangible capital.** We now specify the law of motion of intangible capital. Let  $S_t$  denote the total stock of intangibles available at time  $t$ . We follow Ai et al. (2012) in modeling intangible capital as a stock of growth options:

$$S_{t+1} = [S_t - G(I_t, S_t)] \times (1 - \delta_S) + J_t, \quad (22)$$

where  $J_t$  represents intangible investments at time  $t$ .

Each growth option can be used to build one unit of new firms. Under this normalization,  $G(I_t, S_t)$  is also the total amount of growth options exercised at time  $t$ . Ai et al. (2012) provide a micro-foundation for the aggregator  $G(I_t, S_t)$  by modeling explicitly the competition among ideas

with heterogenous quality. We adopt an aggregator  $G$  with constant elasticity of substitution between physical investment and intangible capital,

$$G(I, S) = \left( \nu I^{1-\frac{1}{\eta}} + (1-\nu) S^{1-\frac{1}{\eta}} \right)^{\frac{1}{1-\frac{1}{\eta}}}, \quad (23)$$

which conforms well to the data on the cross section of book-to-market ratios, as in Ai et al. (2012).

Equation (22) can therefore be interpreted as follows. At time  $t$ , the agent has a mass  $S_t$  of available growth options. If options are exercised optimally and the total amount of investment goods used to exercise options is  $I_t$ , then  $[S_t - G(I_t, S_t)] \times (1 - \delta_S)$  is the total amount of growth options left at the end of the period after depreciation.  $J_t$  is the amount of growth options newly produced in period  $t$ .

To complete the model, we note that consumption, investment in physical capital, and investment in intangible capital must sum up to total output:

$$C_t + I_t + J_t = F(\hat{\mathbf{A}}_t, \bar{\mathbf{A}}_t, \hat{K}_t, \bar{K}_t, N_t).$$

### 2.3 Equilibrium conditions

In our economy, standard welfare theorems apply, and we can construct equilibrium prices and quantities from the solution to the planner's problem. Let  $\Lambda_{t,t+1}$  be the one-step-ahead stochastic discount factor:

$$\Lambda_{t,t+1} = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-1} \left( \frac{u_{t+1}}{u_t} \right)^{1-\frac{1}{\psi}} \left( \frac{V_{t+1}}{E_t \left[ V_{t+1}^{1-\gamma} \right]^{\frac{1}{1-\gamma}}} \right)^{\frac{1}{\psi}-\gamma}. \quad (24)$$

Given the equilibrium quantities, we can show that the cum-dividend price of young firms,  $p_{\bar{K},t}$ , that of the mature firms,  $p_{\hat{K},t}$ , and that of the ideas,  $p_{S,t}$ , must jointly satisfy the following recursions:

$$p_{\bar{K},t} = \alpha \bar{\mathbf{A}}_t \left( \frac{\bar{K}_t}{\bar{N}_t} \right)^{\alpha-1} + (1 - \delta) \left\{ (1 - \phi) E [\Lambda_{t,t+1} p_{\bar{K},t+1}] + \phi E [\Lambda_{t,t+1} p_{\hat{K},t+1}] \right\}, \quad (25)$$

$$p_{\hat{K},t} = \alpha \hat{\mathbf{A}}_t \left( \frac{\hat{K}_t}{\hat{N}_t} \right)^{\alpha-1} + (1 - \delta) E [\Lambda_{t,t+1} p_{\hat{K},t+1}], \quad (26)$$

$$p_{S,t} = \frac{1 - \nu}{\nu} \left( \frac{I_t}{S_t} \right)^{\frac{1}{\eta}} + (1 - \delta_S) E [\Lambda_{t+1} p_{S,t+1}]. \quad (27)$$

According to equation (25), the value of adolescent firms is determined by the marginal product of its capital in the current period,  $\alpha \bar{\mathbf{A}}_t \left( \frac{\bar{K}_t}{\bar{N}_t} \right)^{\alpha-1}$ , plus the continuation value of their future payoffs. Conditional on surviving to the next period with probability  $1 - \delta$ , adolescent firms become mature with probability  $\phi$  and pay  $p_{\bar{K},t+1}$  going forward. With probability  $1 - \phi$ , they stay in adolescence, i.e., they continue to have limited information on their  $\beta$ s and pay the continuation value  $p_{\hat{K},t+1}$ .

Equation (26) implies that the cum-dividends marginal value of mature firms equals the expected present value of the marginal product of old capital adjusted for the survival probability  $1 - \delta$ . Similarly, equation (27) states that the cum-dividend value of a unit of intangible capital is equal to the present value of its marginal product,  $\frac{1-\nu}{\nu} \left( \frac{I_t}{S_t} \right)^{\frac{1}{\eta}}$ , accounting for the survival probability of  $1 - \delta_S$ .

Using the above notation, the optimality for investment in physical capital and intangible capital can be written as

$$E [\Lambda_{t+1} p_{\bar{K},t+1}] - \frac{1}{G_I(I_t, S_t)} = (1 - \delta_S) E [\Lambda_{t+1} p_{S,t+1}], \quad (28)$$

$$1 = E [\Lambda_{t+1} p_{S,t+1}]. \quad (29)$$

The left-hand side of equation (28) measures the net marginal benefit of exercising an additional option, that is, the present value of one additional unit of young physical capital net of the  $\frac{1}{G_I(I_t, S_t)}$

exercise cost.

The right-hand side of equation (28) is, in contrast, the opportunity cost of exercising an additional option, i.e., the market value of an unexercised option adjusted for the probability of death. Finally, equation (29) prescribes that intangible investment must be set so that the ex-dividend value of growth options equals their marginal production cost.

## 2.4 Term structures

Given a sequence of cash flows,  $\{CF_t\}_{t=0}^{\infty}$ , the time  $t$  present value of the time  $t+n$  component of the cash-flow sequence is denoted by  $P_{t,t+n}$  and can be computed as follows:

$$P_{t,t+n} = E_t[\Lambda_{t,t+n}CF_{t+n}] \quad n = 1, 2, \dots,$$

where  $\Lambda_{t,t+n} = \Lambda_{t,t+1} \times \Lambda_{t+1,t+2} \times \dots \times \Lambda_{t+n-1,t+n}$  is the  $n$ -step-ahead discount factor that can be computed from the one-step-ahead stochastic discount factors. The one-period return of the claim to  $CF_{t+n}$  from period  $t$  to  $t+1$  is simply  $\frac{P_{t+1,t+n}}{P_{t,t+n}}$ . We are interested in studying the risk premium,  $RP_t(n)$ , on this return for different maturities  $n$ :

$$RP_t(n) = E_t \left[ \frac{P_{t+1,t+n}}{P_{t,t+n}} - r_t^f \right], \quad n = 1, 2, \dots,$$

where  $r_t^f = \frac{1}{E[\Lambda_{t,t+1}]}$  is the one-period risk-free interest rate. The term structure of a cash-flow sequence  $\{CF_t\}_{t=0}^{\infty}$  refers to the link between  $RP_t(n)$  and  $n$ . Borrowing the terminology from the literature on the term structure of interest rates, we will call  $RP_t(n)$  the risk premium on the zero-coupon equity with maturity  $n$ .

While the term structure of real interest rates is determined by the properties of the stochastic discount factor alone, the term structure of equity returns depends on the dynamics of both the stochastic discount factor and that of the cash-flow process. Our goal is to study the slope of the

term structure of equities in the general equilibrium model we developed above, where both the stochastic discount factor and the cash flows are endogenously determined in equilibrium.

Binsbergen et al. (2012) and Binsbergen et al. (2013) present evidence for substantial variations in the slope of the term structure over time. In particular, Binsbergen et al. (2012) document a significant negative slope of the term structure for the aggregate stock market during recessions. Standard RBC models predict an unambiguously positive slope in the term structure of equity returns, and are therefore inconsistent with the data. In the rest of the paper, we study the term structure of equity return in our learning model in two steps. In section 3, we analyze our learning model with homoscedastic productivity shocks and contrast it with the standard RBC model. Although without time-varying volatility the slope of the term structure is constant over time, this analysis allows us to demonstrate that our learning mechanism creates a downward-sloping term structure, especially when long-run productivity shocks are important.

Guided by this intuition, in section 4 we provide empirical evidence on the time-varying volatility of productivity shocks and incorporate this feature into our learning model. We show that countercyclical variations in the relative volatility of productivity shocks allow our model to account for the variation in both the slope of the term structure and the market equity premium.

### 3 The Unconditional Term Structure

In this section we calibrate our benchmark learning model and compare it to one without learning and without intangible capital, which we call RBC. This is essentially the production economy studied in Croce (2014), and it can be obtained as a special case of our setting under two conditions: (i)  $\tau_{j,t}^2 = 0, \forall j, \forall t$ , that is, all firms have full information about their productivity; and (ii) the  $G(I, S)$  function in equation (21) is replaced by the following capital adjustment cost function:

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

**Table 1: Calibrated Parameter Values**

<b>Preference parameters</b>		
Effective risk aversion	$\gamma \cdot o$	10
Intertemporal elasticity of substitution	$\psi$	2
Discount factor	$\beta$	0.98
Leisure weight	$o$	0.33
<b>Technology parameters</b>		
Capital share	$\alpha$	0.3
Depreciation rate of physical capital	$\delta$	9.9%
Depreciation rate of intangible capital	$\delta_S$	9.9%
Weight on physical investment in $G$	$\nu$	92.5%
Elasticity of substitution in $G$	$\eta$	12
<b>Learning parameters</b>		
Percentage share of firms with limited information transitioning to full information	$\phi$	0.70
Productivity exposure of firms with full information	$\lambda^*$	6
Diffusion of information: cointegration speed	$\rho_s$	0.96
<b>Common productivity parameters</b>		
Average growth rate	$\lambda^* \mu$	00.02
Volatility of short-run risk	$\lambda^* \exp(\sigma_a)$	5.65%
Relative volatility of long-run risk	$\exp(\sigma_x) / \exp(\sigma_a)$	00.12
Autocorrelation of expected growth	$\rho_x$	0.965

This table reports the parameter values used for our annual calibrations. The benchmark model features both tangible and intangible capital, as well as full- and limited-information firms.

We first describe our calibration and then present the quantitative results.

### 3.1 Calibration

We list our calibrated parameter values in table 1. The discount rate ( $\beta$ ), risk aversion ( $\gamma$ ), and intertemporal elasticity of substitution ( $\psi$ ) are set as in standard long-run risk models. The weight of leisure in the utility function ( $o$ ) is chosen to match the average share of hours worked, that is,  $N = 1/3$  in steady state.

Both the capital share ( $\alpha$ ) and capital depreciation rate ( $\delta$ ) are standard in the RBC literature. The parameters governing the accumulation of growth options are chosen in the spirit of Ai et al.

(2012). For parsimony, the depreciation of intangible capital ( $\delta_S$ ) is set equal to that of physical capital. The shape of the aggregator  $G(I, S)$  is determined by two parameters, the weight on physical investment,  $\nu$ , and the elasticity  $\eta$ . As in Ai et al. (2012), we choose them to jointly match the steady-state consumption–tangible investment ratio and the consumption–intangible investment ratio.<sup>3</sup>

Our calibration of the parameters of the aggregate productivity shocks is standard in the long-run productivity risk literature. We calibrate  $\mu$  and  $\sigma_a$  to match the mean and the volatility, respectively, of output growth in the US economy in our sample period, 1929–2007. We set  $e^{\sigma_x - \sigma_a} = 0.12$  and  $\rho = 0.965$ , in the spirit of Croce (2014).

We calibrate the parameters for idiosyncratic productivity shocks to match moments of the joint distribution of firm age and exposure to aggregate productivity shocks. Using COMPUSTAT data, Ai et al. (2012) document a strong positive correlation between firm age and the exposure of firm-level productivity to measured aggregate productivity shocks. In our model, the parameter  $\phi$  is the rate of transition to maturity, and the parameter  $\lambda^*$  governs the difference between the exposure of adolescent and mature firms to aggregate shocks.

We simultaneously calibrate  $\phi$  and  $\lambda^*$  to target the moments of the conditional distribution of firm exposure to aggregate productivity shocks as a function of firm age. Note that  $\lambda^*$  is the difference between the exposure of mature and adolescent firms, and  $\phi$  determines the fraction of mature firms as a function of firm age. Jointly, the two parameters pin down the average exposure of productivity shocks for firms of all ages. We therefore choose  $\phi$  and  $\lambda^*$  jointly to target regression coefficients of the exposure-age relationship in the data. We describe in appendix B the details of this calculation, which yields  $\phi = 0.70$  and  $\lambda^* = 6$ . The persistence of the cointegration residual  $\zeta_t = 0.96$  is obtained by estimating the autocorrelation of the log productivity difference of the top-20 and bottom-20 percentiles of the firm age distribution.

When calibrating the RBC model, we retain the same calibration except for two modifications.

---

<sup>3</sup>Ai et al. (2012) show that the distribution of the productivity of growth options implied by this calibration is similar to its empirical counterpart measured from the distribution of the book-to-market ratio of IPO firms.

We set the subjective discount factor to 0.99 to match the average risk-free interest rate.<sup>4</sup> We also lower the volatility of short-run shocks to 4% ( $\lambda^* e^{\sigma_a} = 4\%$ ) to match the volatility of total output in the data. A lower level of  $\lambda^* e^{\sigma_a}$  matches the same level of volatility of output as the learning model because all firms are fully exposed to shocks to  $\Delta a_{t+1}$ . Additionally, we set the adjustment cost parameter  $\xi = 1.27$  to obtain an annual market risk premium the same as our benchmark learning model, 4% per year.

### 3.2 Quantitative results

We report the quantitative implications of our benchmark model and the RBC model in table 2. We make several observations. First, our benchmark model and the RBC model have similar implications for macroeconomic quantities, except that the RBC model produces a significantly lower volatility of investment. Due to the absence of adjustment costs, the ratio of the volatility of investment relative to that of consumption in our benchmark model is 4.03, much closer to its empirical counterpart, 5.29.

Second, both models produce a significant equity premium, but they yield very different term structures. In our benchmark model, the term structure of equity is downward sloping over short maturities, whereas that in the RBC model is upward sloping. The risk premium on the claim to the zero-coupon equity with two-year maturity ( $RP(2)$ ) in our benchmark model is 6.77% per year, close to the evidence reported by Binsbergen et al. (2012), who show that a strategy long in a dividend strip with a maturity of 1.9 years and short in a dividend strip with a maturity of 0.9 years pays an average annual excess return of 10.10%. In contrast,  $RP(2)$  in the RBC model is  $-26.76\%$ , and the high market risk premium is obtained by a very high right tail of the term structure.<sup>5</sup>

---

<sup>4</sup>The difficulty the RBC models have in simultaneously matching the interest rate and the investment-to-output ratio is well known (the risk-free rate puzzle). We chose the parameters to match the level of the risk-free rate, but not the investment-to-output ratio.

<sup>5</sup>By no arbitrage, the value-weighted return on all zero coupon equities must sum up to the market equity return. In the RBC model, the aggregate premium is substantial because the right tail of the term structure is

**Table 2: Vintage and Intangible Capital Models**

	Data	Benchmark	RBC
Panel A: Aggregate Quantities			
$E[I/Y]$	00.15 (00.05)	0.17	0.29
$\sigma(\Delta c)$	02.53 (00.56)	2.98	2.76
$\sigma(\Delta i)/\sigma(\Delta c)$	05.29 (00.50)	4.03	1.48
$\sigma(\Delta j)/\sigma(\Delta i)$	00.50 (00.07)	1.04	–
$\sigma(\Delta n)$	02.07 (00.21)	1.51	0.10
$AC_1(\Delta c)$	00.49 (00.15)	0.41	0.21
$\rho_{\Delta c, \Delta n}$	00.28 (00.07)	0.55	0.47
$\rho_{\Delta c, \Delta i}$	00.39 (00.15)	0.69	0.91
Panel B: Asset Prices			
$E[r^f]$	00.89 (00.97)	0.44	0.96
$E[r_K^{L,ex}]$	05.70 (02.25)	4.05	4.02
$E[r_K^L - r_S^L]$	04.32 (01.39)	3.83	–
$RP(2)$	10.08 (05.04)	6.77	-26.76

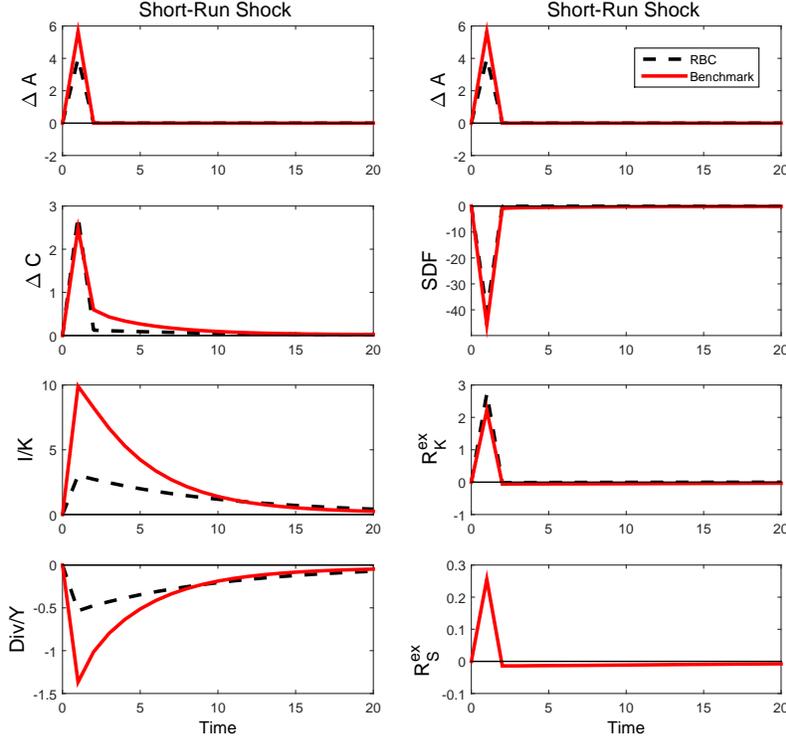
All entries for the models are obtained from repetitions of small samples. Data refer to the US and span the sample 1930–2007, unless otherwise stated. Numbers in parentheses are GMM Newey-West adjusted standard errors.  $E[r_K^L - r_S^L]$  and  $E[r_K^{L,ex}]$  measure the levered spread between tangible and intangible capital returns, and the levered excess returns of tangible capital, respectively. We assume a constant leverage of three, consistent with Garcia-Feijo and Jorgensen (2010). In the data, we use the Fama-French HML factor and the market excess return factor as the counterparts of  $E[r_K^L - r_S^L]$  and  $E[r_K^{L,ex}]$ , respectively. The annualized empirical counterpart of the risk premium on the zero-coupon equity with maturity of two years,  $RP(2)$ , is from Binsbergen et al. (2012). Volatility and correlations are denoted as  $\sigma(\cdot)$  and  $\rho_{\cdot, \cdot}$ , respectively. Our annual calibrations are reported in table 1.

Third, our benchmark model also produces a significant spread between the return on physical capital and the return on intangible capital. This implication of our model is consistent with the value premium evidence, that is, the fact that physical capital-intensive firms have a higher return than intangible capital-intensive firms.

To understand the above implications, we compare the impulse response functions of quantities and prices of our benchmark model with learning to those in the Croce (2014) model. In figure 1 we depict the response of quantities (left panel) and asset prices (right panel) to short-run shocks. The responses to long-run productivity shocks are shown in figure 2.

---

high and positive.



**Fig. 1: Impulse Response Functions for Short-Run Shocks**

This figure shows percentage annual log-deviations from the steady state upon the realization of a positive short-run shock. Returns are not levered. Both the RBC and benchmark models feature short- and long-run productivity risk. The RBC model has convex adjustment costs. The benchmark capital model features limited information and learning and is calibrated as detailed in table 1.

**Contemporaneous productivity shocks.** Focusing on contemporaneous shocks to productivity, or short-run shocks, in figure 1, we note that the benchmark capital model produces responses qualitatively similar to those in the RBC model, but with a few differences. First, in the absence of adjustment costs, physical investment responds strongly to short-run productivity shocks. As a result, our learning capital model generates a high volatility of investment, as in the data. The significant adjustment cost helps the RBC model produce a high level of equity premium, but at the cost of a counterfactually low level of volatility of investment.

Second, in our benchmark model, the response of the return on intangible capital ( $R_S^{ex}$ ) to short-run productivity shocks is significantly smaller than that of the physical capital ( $R_K^{ex}$ ). This result

contributes to the generation of a value premium in our model and can be explained as follows. Short-run productivity shocks directly affect the payoff of physical capital owned by existing firms. In addition, most of the existing firms are mature firms, and their marginal product of capital is more sensitive to these shocks than is that of adolescent firms. In contrast, productivity shocks affect the payoff of intangibles only indirectly: a growth option benefits from productivity shocks because option exercise leads to creation of an adolescent firm. Because adolescent firms are less able to take advantage of technological progress,  $R_S^{ex}$  is less sensitive to contemporaneous productivity shocks than  $R_K^{ex}$ .

Third, in both models investment responds positively to contemporaneous productivity shocks, whereas short-term dividends respond negatively. This creates a force in both models that pushes up the slope of the term structure of equity returns. The negative response of dividends to productivity shocks is the implication of the general-equilibrium resource constraint:

$$D_t = Y_t - W_t N_t - I_t - J_t = \alpha Y_t - I_t - J_t,$$

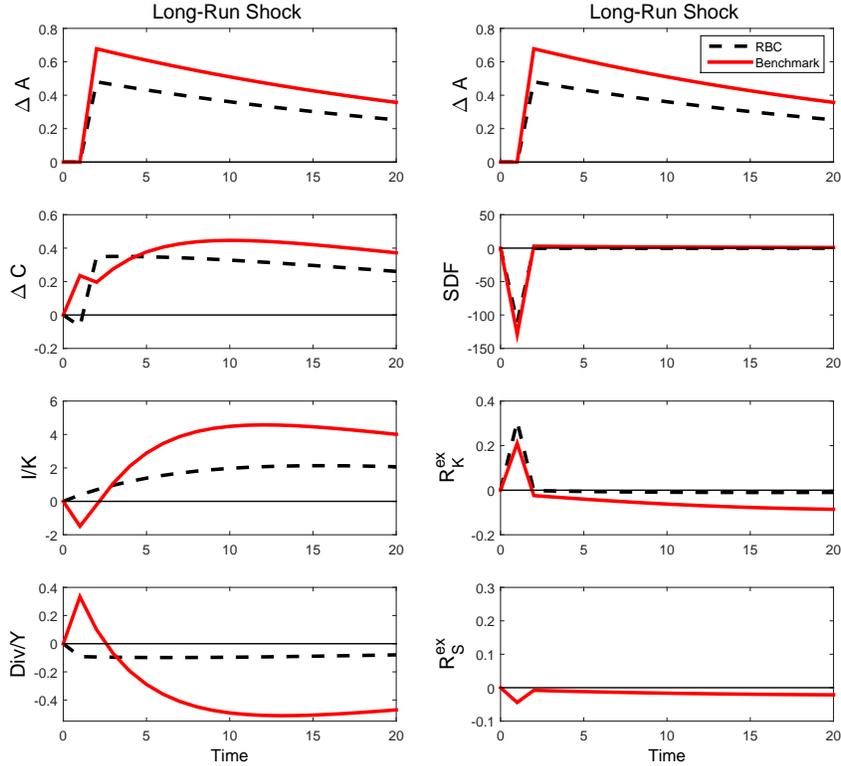
where  $W_t$  is the equilibrium wage, and the second equality results from our constant factor shares. Because investment responds strongly and positively to contemporaneous productivity shocks, dividends must respond negatively.

To understand the negative slope of the term structure of equity returns in our learning model, we need to turn to the impulse responses with respect to news shocks.

**News about future productivity shocks.** As in Bansal and Yaron (2004), Croce (2014), and Gourio (2012), news about future consumption growth requires a significant compensation under recursive preferences.<sup>6</sup> In figure 2, we plot the impulse response functions produced by our learning model (solid line) and by the RBC model (dashed line) with respect to shocks to news about future productivity,  $\varepsilon_{x,t+1}$ .

---

<sup>6</sup>In rare events model, news about the probability of disasters are also growth news shocks.



**Fig. 2: Impulse Response Functions for Long-Run Shocks**

This figure shows percentage annual log-deviations from the steady state upon the realization of a positive long-run shock. Returns are not levered. Both the RBC and benchmark models feature short- and long-run productivity risk. The RBC model has convex adjustment costs. The benchmark capital model features limited information and learning and is calibrated as detailed in table 1.

The impulse responses of investment are the key to understanding the difference in the asset pricing implications of the two models. First, note that in the RBC model investment responds positively to news shocks. With an IES of  $\psi = 2$ , upon the arrival of positive news about future productivity shocks, the substitution effect dominates, and investment rises. As for the case of positive short-run shocks, increases in investment are associated with decreases in dividends. This pattern of the impulse response makes short-term dividends less risky than long-term dividends, and hence it reinforces the upward-sloping term structure of equity produced by short-run productivity shocks.

In contrast, in our learning model, on impact, 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 because new investment builds adolescent firms, which cannot take full advantage of the rise in productivity. On the other hand, the income effect is strong because all existing mature firms immediately benefit from the positive productivity shock. As a result, investment immediately drops and a higher dividend payout is used to support more consumption. Over time, as the effect of the news materializes, investment eventually picks up, and dividends fall correspondingly.

This pattern of impulse response has strong implications for asset prices. First, even though investment drops upon the arrival of positive news, the return to physical capital responds strongly and positively to news, allowing our learning model to produce a high equity premium without resorting to adjustment costs. Note that equity is the claim to existing firms, most of which are mature and respond strongly to productivity shocks. As a result, the return on physical capital ( $R_K^{ex}$ ) rises immediately. In contrast, in the RBC model, the strong reaction of  $R_K^{ex}$  to news shocks is achieved by assuming a high adjustment cost, which produces a counterfactually low level of the volatility of investment.

Second, because short-term dividends respond positively to news, they are more risky and require a greater compensation. The impulse response of dividends in our learning model reflects that of the term structure of equity returns. As a result, our learning model generates a downward sloping term structure over short maturities, in contrast to the RBC model.

Third, the negative response of physical investment with respect to positive news about future productivity shocks also implies that the return to intangible capital declines, therefore providing a hedge against these shocks. Because intangible capital and physical investment are complements—less tangible investment implies that a smaller fraction of growth options can be exercised—the decline of physical investment is associated with a lower market value of growth options and therefore a lower return  $R_S^{ex}$ . At the equilibrium, news shocks enhance the value premium

generated in our benchmark model.

**Sensitivity Analysis.** In Appendix D, we show that our results are robust to a range of plausible values for the speed of information diffusion ( $\rho_s$ ) and the speed of learning ( $\phi$ ). Furthermore, we show that our results continue to hold even without intangible capital ( $\nu = 1$ ).

## 4 Dynamics of the Term Structure Slope

The analysis in the previous section implies that news shocks and contemporaneous productivity shocks have opposite effects on the slope of the term structure of equity. Because investment responds positively to contemporaneous productivity shocks, the aggregate payout has to react negatively due to the resource constraint. As a result, short-term dividends must be less risky than long-term dividends. In contrast, under our learning friction the income effect dominates upon positive news shocks, investment declines, and short-term dividends increase. As a result, news shocks make short-term dividends riskier and generate a downward-sloping term structure.

Naturally, if the relative importance of news shocks and contemporaneous productivity shocks is time varying, then our model has the potential to account for the time-varying slope of the term structure. Guided by this theoretical insight, in the rest of this section we present and estimate a model of time-varying productivity volatility. We then recalibrate our learning model to incorporate our novel empirical evidence. By doing so, we connect the time-varying slope of the term structure to fundamental macroeconomic volatility factors.

## 4.1 A model with time-varying volatility

**Model specification.** We replace the homoscedastic model in equation (20) with the following specification of the productivity process:

$$\log \frac{A_{t+1}}{A_t} \equiv \Delta a_{t+1} = \mu + x_t + e^{\bar{\sigma}_a + \sigma_{a,t}} \varepsilon_{a,t+1}, \quad (30)$$

$$x_{t+1} = \rho_x x_t + e^{\bar{\sigma}_x + \zeta_t + \sigma_{a,t}} \varepsilon_{x,t+1},$$

$$\sigma_{a,t+1} = \rho_\sigma \sigma_{a,t} + \sigma_\sigma \varepsilon_{\sigma,t+1}, \quad (31)$$

where  $[\varepsilon_{a,t+1}, \varepsilon_{x,t+1}, \varepsilon_{\sigma,t+1}]$  is a vector of standard normal shocks i.i.d. over time. The process  $\sigma_{a,t}$  is the time-varying stochastic log-volatility for contemporaneous productivity shocks. The key element in our estimation is the term  $e^{\zeta_t}$ , which captures the variations in the relative volatility between long-run and short-run productivity shocks. For parsimony, in the model we assume that the relative volatility is a negative log-linear function of the state variable  $x_t$ :

$$\zeta_t = -\beta_{\zeta|x} x_t. \quad (32)$$

In what follows, we present an empirical procedure to estimate the time-varying conditional volatilities of short- and long-run productivity growth shocks, and we investigate their properties to motivate our specification in (32). The empirical estimation also provides us guidance on the key parameters that are new to the stochastic volatility model, namely,  $\rho_\sigma$ ,  $\sigma_\sigma$ , and  $\beta_{\zeta|x}$ .

**Estimation procedure.** We use a quarterly sample ranging from 1947:Q1 to 2013:Q4. Our main goal is to obtain measures of the conditional volatilities of the long-run and short-run productivity shocks and further construct a measure of the ratio of the two volatilities,  $\zeta_t$ . Because none of the above quantities is directly observable, we adopt the following procedure to construct estimates of them. First, as in Croce (2014), we project the one-period-ahead productivity growth,

$\Delta a_{t+1}$ , on a set of predictive factors,  $F_t$ ,

$$\Delta a_{t+1} = \mu + \beta^x F_t + u_{a,t+1}, \quad (33)$$

and use the fitted value to construct our estimates of the latent predictive component of productivity,  $x_t = \widehat{\beta}^x F_t$ . The short-run shocks are identified using the residual of the regression above,  $u_{a,t+1}$ , and innovations in news shocks are constructed as the residual of the following estimated AR(1) process:

$$x_{t+1} = \rho_x x_{t+1} + u_{x,t+1}. \quad (34)$$

Second, we take the residuals,  $u_{a,t+1}$  and  $u_{x,t+1}$ , and construct measures of their conditional variances. For robustness, we adopt two alternative approaches. We call the first approach the predictive approach. That is, we regress the realized variances of  $u_{a,t+1}$  and  $u_{x,t+1}$  on the same vector of predictive variables,  $F_t$ :

$$\log \left( \frac{1}{h} \sum_{j=1}^h u_{a,t+j}^2 \right) = \beta_{a,0}^\sigma + \beta_a^\sigma F_t + error, \quad (35)$$

$$\log \left( \frac{1}{h} \sum_{j=1}^h u_{x,t+j}^2 \right) = \beta_{x,0}^\sigma + \beta_x^\sigma F_t + error. \quad (36)$$

We set  $h = 4$ , so that realized variance is measured as the sum of the squared innovations in the next four quarters. This procedure allows us to construct the demeaned conditional log standard deviation as the square root of the predictable component of the realized variances:  $\widehat{\sigma}_{a,t} = \frac{1}{2} \left( \widehat{\beta}_a^\sigma F_t \right)$  and  $\widehat{\sigma}_{x,t} = \frac{1}{2} \left( \widehat{\beta}_x^\sigma F_t \right)$ . Our second approach is the GARCH approach, in which we replace equations (35)–(36) with two GARCH(1,1) models.

Third, we construct the key object of interest, the relative volatility of long-run versus short-run shocks, as  $\widehat{\zeta}_t = \widehat{\sigma}_{x,t} - \widehat{\sigma}_{a,t}$  and investigate its empirical properties.

In our benchmark estimation, we use the four factors proposed by Bansal and Shaliastovich

**Table 3: Time-Varying Volatility in Productivity**

	$\rho_\sigma$	$\sigma_\sigma$	$\rho_x$	$\rho_\zeta$	$\beta_{\zeta_t x_t}$	$\frac{StD(\widehat{\zeta}_t)}{StD(\zeta_t)}$	$\beta_{\Delta y_{t+1} t \zeta_t}$	$\beta_{\Delta y_{t+2} t \zeta_t}$
Est.	0.91***	0.09***	0.77***	0.92***	-30.69***	0.71	-0.02*	-0.02**
St.Err.	0.05	0.02	0.08	0.09	12.06	–	0.01	0.01

Data refer to the U.S. and span the sample 1947:Q1–2016:Q4. All  $t$ -stats are based on GMM Newey-West adjusted standard errors. We jointly estimate the set of equations (30)–(36) as well as the following equations:

$$\begin{aligned}\zeta_t &= const + \rho_\zeta \zeta_{t-1} + \epsilon_{\zeta,t} \\ \zeta_t &= const + \beta_{\zeta_t|x_t} x_t + resid_t \\ \Delta y_{t+j|t} &= const + \beta_{y_{t+j|t}|\zeta_t} \zeta_t + resid_t \quad j = 1, 2,\end{aligned}$$

where  $\Delta y_{t+j|t}$  denotes real output growth over  $j$  periods. In this table,  $\widehat{\zeta}_t = \widehat{\beta}_{\zeta_t|x_t} x_t$ . Our five factors are the price-dividend ratio, the three-months Treasury bill yield, the three-year and five-year Treasury bond yields, and the integrated volatility of stock market returns.

(2013) plus the integrated volatility of stock market returns.<sup>7</sup> To ensure robustness, we also consider an alternative specification with the thirteen factors proposed by Jurado et al. (2015), which are principal components extracted from a very wide cross section of both macroeconomic and financial indicators and have significant predictive power for aggregate volatility. These factors are available only over the shorter sample 1960:Q3–2011:Q4.

In table 3 we summarize the results based on our benchmark specification. We report our robustness results in the appendix, table C.1. Across all procedures, we estimate all coefficients jointly by continuous GMM and use as many orthogonality conditions as parameters. Hence, our inference accounts for all possible layers of estimation uncertainty, and our point estimates are equivalent to those of a multistep OLS procedure. Standard errors are computed using the GMM efficient weighting matrix and are Newey-West adjusted. All data are quarterly to better capture variability in the conditional volatility of productivity.<sup>8</sup>

<sup>7</sup>The four factors are the price-dividend ratio, the three-month Treasury bill yield, the three-year and five-year Treasury bond yields. Integrated volatility is the sum of squared daily returns calculated at a quarterly frequency. This measure is based on stock market indices (NYSE/AMEX/NASDAQ) from CRSP and is expressed in annualized percentage units.

<sup>8</sup>Productivity is measured by the total factor productivity index for the US business sector published by the San Francisco FED. The price-dividend ratio is from R. Shiller’s webpage. Bond yields are from the Global Financial Database. Our sample starts in 1947:Q1 and ends in 2013:Q4. The Jurado et al. (2015) factors are available from

**Estimation results.** We summarize our main estimation results as follows. First, the estimated relative volatility,  $\zeta_t$ , exhibits significant variation over time. In table C.1, we report a Wald test for the hypothesis that both volatilities have the same loadings on our predictive factors:  $\beta_x^\sigma = \beta_a^\sigma$ . Under this null hypothesis, relative volatility would be constant. We strongly reject this null across all of our specifications.

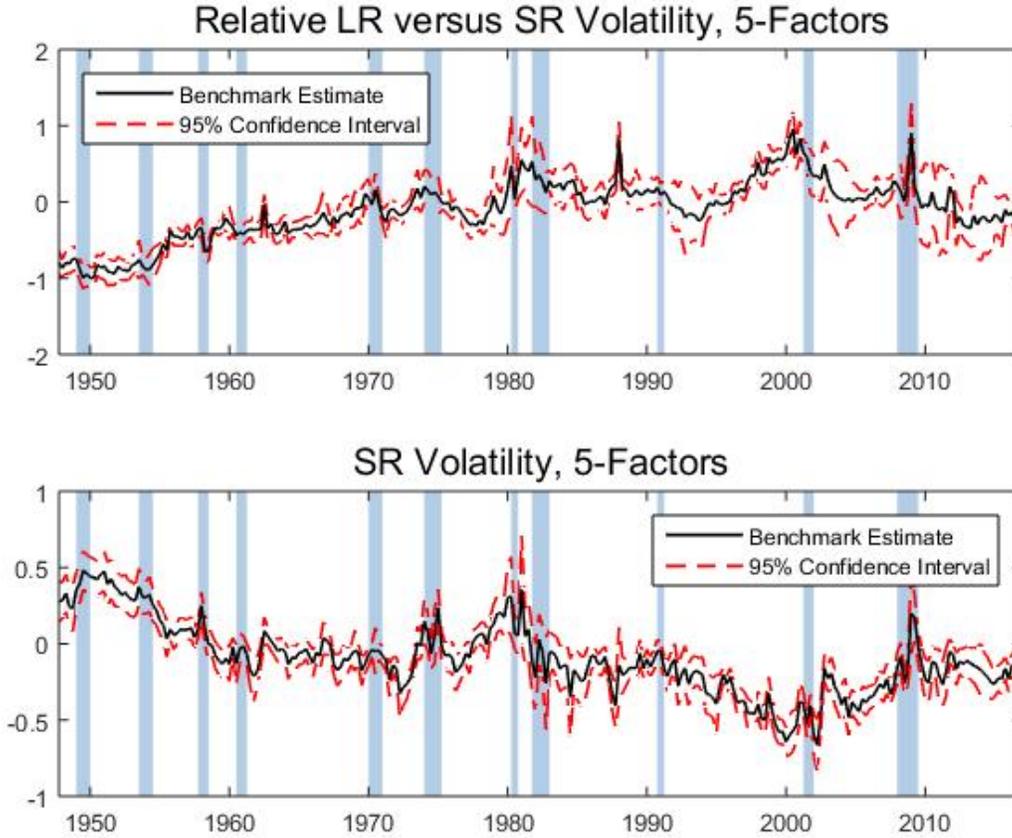
In figure 3, we plot our constructed time series of  $\zeta_t$  (top panel) and  $\sigma_{a,t}$  (bottom panel) and use shaded areas to indicate NBER-defined recessions. Our second result refers to the counter-cyclicality of  $\zeta$ , that is, our relative volatility process tends to spike up right before recessions, and it negatively predicts future economic growth. In table 3, we report the results of a regression of future output growth,  $\Delta y_{t+j} \equiv \ln Y_{t+j} - \ln Y_t$ , on  $\zeta_t$ . The regression coefficient is unambiguously negative and statistically significant, consistent with the patterns depicted in figure 3.

Third, the estimated  $x_t$  and  $\zeta_t$  processes are persistent and highly correlated with each other (see figure C.1 for our estimated  $x$ ). In our benchmark estimation, the sample autocorrelations of  $x_t$  and  $\zeta_t$  are 0.80 and 0.90, respectively. The Wald test of the hypothesis that these autocorrelation coefficients are zero has a  $p$ -value smaller than 1% (see table C.1). The high correlation between relative volatility  $\zeta$  and expected growth  $x$  is reflected by the fact that when we project  $\zeta_t$  on  $x_t$ , we can explain 75% of the standard deviation of relative volatility (table 3). Given that both  $x_t$  and  $\zeta_t$  have significant predictive powers for future economic growth and that they are highly correlated, our specification of  $\zeta_t$  in equation (32) is an efficient way to summarize the dynamics of relative volatility without introducing additional state variables into the model.

We use these estimation results to guide our calibration. Specifically, we set  $\beta_{\phi|x} = -30.7$  according to its point estimate. We also calibrate  $\rho_\sigma = 0.91$  and  $\sigma_\sigma = 0.09$  to match the point estimates of the autocorrelation coefficient and volatility for the  $\sigma_{a,t}$  process. We now turn to the quantitative implications of our model.

---

1960:Q1 to 2011:Q4 on S. Ludvigson's webpage.



**Fig. 3: Volatility Factors in Productivity ( $\zeta_t$  and  $\sigma_{a,t}$ )**

This figure shows the estimated relative volatility process,  $\zeta_t$  (top panel), and the conditional volatility of the productivity short-run shock,  $\sigma_{a,t}$  (bottom panel), obtained through the methods described in section 4.1. The main empirical features of these processes are reported in table 3. Grey bars denote NBER recession periods.

## 4.2 Quantitative results

**The slope of the term structure of equity returns.** To illustrate the relationship between the conditional volatility of productivities and the slope of the term structure, we report the implications of our model for the slope of the term structure for different combinations of the conditional volatilities in table 4.  $RP(7) - RP(2)$  is the levered model-implied spread between the expected return on a zero-coupon equity with a seven-year maturity and that with a two-year maturity. We also report the market equity premium as  $E_t[r^{Lev}]$ , assuming a financial leverage of 3.

We choose two levels of the volatility of the contemporaneous productivity shock, 5.65% for

**Table 4: Productivity Volatility Factors in Our Model with Learning**

Panel A: Conditional Risk Premia					
Short-run vol.	Low		High		
Long-run relative vol.	Low	High	Low	High	
$RP_t(7) - RP_t(2)$	1.82	-1.55	5.23	-6.00	
$E_t[r^{L,ex}]$	1.16	1.35	4.59	5.38	
$E_t[r_{K,t+1}^L - r_{S,t+1}^L]$	1.10	1.31	4.34	5.23	
$MD^S/MD^K$	1.9	1.9	1.9	1.9	
Panel B: Conditional Second Moments and CAPM					
Maturity	Vol.	SR	$\alpha$	$\beta$	$\partial\beta/\partial x$
		Spot Equity Excess Returns			
2	8.46	0.37	4.21	-0.81	-257.8
7	5.73	0.23	0.31	0.73	-46.7
		Forward Equity Excess Returns			
2	8.37	0.38	4.22	-0.79	-255.5
7	5.71	0.26	0.42	0.78	-31.7
		Bonds Excess Returns			
2	0.16	-0.26	-0.02	-0.02	-2.3
7	0.70	-0.25	-0.18	-0.05	-15.0

All entries are obtained from our benchmark model augmented by time-varying volatility factors, as described in section 4. Our baseline annual calibration is reported in table 1, the additional parameters are specified in section 4. Excess returns are levered by a factor of three, consistent with Garcia-Feijoo and Jorgensen (2010). In panel B, all entries refer to the case of low short-run volatility risk and high relative volatility. SR denotes the Sharpe-ratio, and  $\alpha$  and  $\beta$  are obtained from a conditional CAPM regression. The forward equity excess return is obtained by going long in zero-coupon equity and short in a bond of the same maturity.

the high-volatility regime, which corresponds to the average volatility for the pre-World War II period, and 2.83% for the low-volatility regime, which corresponds to the post-World War II period. We set the low relative volatility to be 4.5% and the high relative volatility to be 5% to illustrate the effect of relative volatility.

The first column of table 4, panel A (“Low/Low”), refers to a case of moderation in both short-run volatility and long-run risk. In this setting, our model delivers an upward-sloping term structure over the seven-year horizon, as long-run news is not sizeable enough to make the term structure downward sloping over short maturities. Both the conditional value premium and the

conditional aggregate equity premium are below average due to the assumed moderation scenario.

In the second column of table 4, panel A (“Low/High”), we consider a scenario in which the relative volatility of long-run risk is 25% higher than in the low state. In this case, the long-run shocks are sizable enough to make short-term dividends riskier than dividends with a maturity of seven years, that is, the term structure slopes downward. Both the conditional value premium and the conditional equity premium, in contrast, increase with respect to the figures reported in the first column.

The next two columns in panel A focus on the case of higher short-run volatility. An increase in short-run volatility simultaneously (i) magnifies both the equity and the value premium, and (ii) expands the absolute value of the term structure spread. Hence the sign of the term structure spread depends solely on relative volatility, whereas the magnitude of the spread depends on the amount of short-run volatility. We formally test this model-implied relationship between relative volatility and the slope of the term structure in section 5 of this paper.

**Implications for the cross section.** Empirically, high book-to-market-ratio stocks (value stocks) earn a higher average return than low book-to-market stocks (growth stocks). Because the cash flow of value stocks has a shorter duration than that of growth stocks (Da (2009), Dechow et al. (2004)), in endowment economies where value and growth are both claims to shares of aggregate dividend, the existence of the value premium requires a downward-sloping term structure of equity returns (Lettau and Wachter (2007), Santos and Veronesi (2010), Lettau and Wachter (2011), Croce et al. (2015)). For these models, the Binsbergen et al. (2013) evidence on the time-varying *sign* of the slope of the term structure of equity return is challenging. Specifically, the upward-sloping term structure observed during boom periods would imply a growth premium, as opposed to the value premium that we observe in the data.

Our production-based economy is not subject to this problem, as the value premium depends on endogenous, heterogenous exposure of tangible and intangible capital to fundamental shocks. In

contrast to prior literature, duration is not the key determinant of risk premia. More specifically, as we have shown in section 3, our model generates a value premium because growth options have endogenously lower exposure to news shocks than value stocks. This feature remains unchanged with time-varying volatility.

In panel A of table 4, we report our model-implied spread between value and growth stocks,  $E[r_K^L - r_S^L]$ , for different combinations of conditional volatilities. We also report the Macaulay duration of the cash flows of value and growth portfolios implied by our model, where the Macaulay duration is calculated using the steady-state discount rate. The existence of the value premium and the negative relation between expected returns and duration in the cross section of stocks are a robust outcome across all scenarios, regardless of the sign of the slope of the term structure.

**Conditional second moments and CAPM.** Binsbergen et al. (2012), Binsbergen et al. (2013), and Binsbergen and Koijen (2017) document several facts regarding the term structure of equity returns. First, both the risk premium and Sharpe ratio for short-maturity claims to zero-coupon equity are higher than for the aggregate stock market (Binsbergen and Koijen (2017)). Second, the returns on short-term dividend claims are risky as measured by volatility, but safe as measured by market betas (Binsbergen and Koijen (2017)). Third, the CAPM  $\beta$  of claims to aggregate dividends is countercyclical and this time variation of  $\beta$  decreases with maturity (Binsbergen et al. (2013)).

These results are documented around the period of the Great Recession, which according to our estimation is a period with low volatility of short-run productivity shocks (post–World War II) and higher long-run volatility. For consistency, we simulate our model conditioning on this combination of values for our volatility state processes, and we report the model-implied moments in panel B of table 4.

Binsbergen and Koijen (2017) report data on both the returns on claims to equity dividend strips (spot equity returns) and those on their futures contract (forward equity returns). Forward

returns are just the spot equity returns less the returns of a bond of equal maturity, and they help in separating the effect of the term structure of interest rates from the effect of the term structure of the equity premium. We report key moments for all of these returns in panel B of table 4 and make the following observations.

First, short-term dividends in our model have a higher risk premium as well as a higher Sharpe ratio, consistent with the pattern reported in Binsbergen and Kojien (2017). These features of our model are due to the fact that short-term dividends have a larger exposure to news shocks, which require a higher market price of risk than contemporaneous productivity shocks.

Second, short-term dividends have a higher return volatility, but a lower CAPM  $\beta$ . In our model, short-term dividends have a higher volatility because the effect of productivity shocks on cash flows decays over time, as shown in the impulse response functions in figure 4. In addition, the failure of CAPM can be explained by the presence of multiple shocks in our model. Note that under recursive preferences, all three shocks  $[\varepsilon_{a,t+1}, \varepsilon_{x,t+1}, \varepsilon_{\sigma,t+1}]$  carry a risk premium, but they are independent of each other and have different market prices of risk. Short-term dividends are more sensitive to news shocks, which carry a high market price of risk but do not generate very volatile responses in the returns. As a result, short-term dividends have a high Sharpe ratio, but a low CAPM  $\beta$  and high  $\alpha$  compared to longer-maturity cash flows.

Third, consistent with the evidence provided by Binsbergen et al. (2013), the CAPM  $\beta$ s in our model are countercyclical. In the rightmost column of panel B, Table 4, we report the model-implied sensitivity of CAPM  $\beta$  with respect to the expected growth rate of the economy. All partial derivatives are clearly negative for equities. They also decrease with respect to maturity, as in Binsbergen et al. (2013), because the effect of news shocks decays over time, as shown in the impulse response functions in figure 2.

Consistent with the evidence provided in Binsbergen and Kojien (2017), these results also apply to forward equity excess returns, because they are driven by the term structure of the risk premium and not by that of the risk-free interest rate.

## 5 Testable Implications of the Learning Mechanism

The key features that distinguish our setting from the standard RBC model are the learning mechanism and time-varying relative volatility of long-run versus short-run shocks. In this section, we formally test several implications of these features of our model.

**News shocks and payouts.** The key implication of the learning mechanism in our model is the response of investment and dividends to news shocks. In the standard RBC model, investment responds positively, and therefore dividends respond negatively, to news shocks. The opposite happens in our model for the term structure of equity returns. In this section, we directly test this implication of our model using evidence on macroeconomic quantities. We show that the aggregate payout has a negative exposure to short-run shocks but a positive exposure to growth news shocks, as in our setting with learning.

We proceed in two steps. First, we measure aggregate dividends using the accounting identity implied by our model,  $D_t = Y_t - I_t - J_t - W_t N_t$ . Because our model abstracts away from leverage and capital structure decisions,  $D_t$  in our model cannot be directly compared to stock market dividends. We therefore use the model to guide our empirical measurement. Both output,  $Y_t$ , and investment,  $I_t + J_t$ , are from table 1.1.5 of the NIPA system. We exclude both government expenditure and net exports, to be consistent with the model, and use the CPI index for all urban consumers to obtain real values. As in Choi and Rois-Rull (2009), we estimate labor income to be 65% of total output. For robustness, we also consider the aggregate dividends reported in the Flow of Funds Accounts dataset for nonfinancial firms over the sample period 1952:Q1–2016:Q4, in the spirit of Belo et al. (2014).<sup>9</sup>

Due to data limitations, we focus on quarterly observations that are available only starting from 1947:Q1. To maximize sample length, our data include observations through 2016:Q4. Our

---

<sup>9</sup>We use Table F.103 (quarterly): Balance Sheet of Nonfinancial Corporate Business. Cash dividend is constructed as the net dividends, line 2.

main results are robust to the exclusion of the Great Recession period.

In our second step, we estimate the following equation:

$$\begin{aligned} Z_t - E_{t-1}[Z_t] &= \beta_{srr} e^{\sigma_{a,t-1}} \epsilon_{a,t} + \beta_{lrr} e^{\sigma_{a,t-1} + \zeta_{t-1}} \epsilon_{x,t} + \beta_{vol} \epsilon_{\sigma,t} + \beta_{rel\_vol} \epsilon_{\zeta,t} + resid_t, \\ E_{t-1}[Z_t] &= \beta_0 + \rho Z_{t-1} + \beta_x x_{t-1} + \beta_\sigma \sigma_{a,t-1} + \beta_\zeta \zeta_{t-1}, \end{aligned} \quad (37)$$

where  $Z_t$  is either the investment-to-output ratio,  $\frac{I_t}{Y_t}$ , or the dividends-to-output ratio,  $\frac{D_t}{Y_t}$ . We divide our main variables by output for three reasons: (i) since  $D_t$  can be negative, we cannot just focus on growth rates; (ii) this is a common way to detrend our variables; and (iii) according to the model, it does not affect our ability to identify the sensitivity of our variables to news shocks, as total output is nearly invariant upon the arrival of pure news shocks.

In the model, a linear approximation of the equilibrium dividend and investment processes suggests the dependence of these variables on both contemporaneous productivity innovations  $(\epsilon_{a,t}, \epsilon_{x,t}, \epsilon_{\sigma,t}, \epsilon_{\zeta,t})$  and predetermined state variables. For the sake of parsimony, we use the lagged values of either  $\frac{I_{t-1}}{Y_{t-1}}$  or  $\frac{D_{t-1}}{Y_{t-1}}$  to capture the role of the endogenous state variables (i.e., capital stocks) 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,  $x_{t-1}$ , relative long-run volatility,  $\zeta_{t-1}$ , and short-run conditional volatility,  $\sigma_{a,t-1}$ .

Our main findings are reported in panel A of table 5. Since neither our dividends series nor our regressors are standardized, magnitudes are not directly comparable. As a result, we only discuss the sign of our estimates. The data suggest that the response of the aggregate payout to short-run news is negative, as predicted by standard production-economy models. Most importantly, the immediate response of aggregate investment to long-run news is negative, implying that the aggregate payout increases with positive new shocks, consistent with our model.

We point out two additional empirical results that broadly support the validity of our empirical methods. First, according to our estimation, cash dividends feature a positive response to news

Table 5: News Shocks, Payout, and Asset Prices

Panel A: Payout Exposures					
	$\beta_{srr}$	$\beta_{lrr}$	$\beta_{vol}$	$\beta_{rel.vol}$	Adj. $R^2$
Aggregate Investment	0.078*** (0.014)	-0.150*** (0.064)	-0.097*** (0.033)	-0.090*** (0.039)	0.937
Aggregate Payout	-0.051*** (0.015)	0.160*** (0.064)	0.098*** (0.034)	0.099*** (0.040)	0.938
FoF Dividends	-0.020** (0.012)	0.022* (0.035)	0.004 (0.025)	-0.007 (0.027)	0.884
Panel B: Time-Varying Volatility Factors and Asset Prices					
	$R_{mkt}^{ex}$	$HML$	$TSS$		
Rel.Vol.	0.423* (0.273)	0.080 (0.219)	-0.630*** (0.092)		
SR-Vol.	0.419 *** (0.173)	0.095 (0.144)	0.528*** (0.045)		
Adj. $R^2$	0.027	-0.009	0.699		
Adj. $R^2$ (no SR-vol.)	0.010	-0.006	0.463		
Sample:	1947:Q3–2016:Q4	1947:Q3–2016:Q4	2002:Q4–2010:Q4		

In panel A, data are from the US and span the sample 1947:Q1–2016:Q4. The statistics reported refer to the regression specified in equation (37). In panel B, we report the estimates from the following regressions:

$$\begin{aligned}
 Z_{t+1} &= const. + \beta_x^z x_t + \beta_\phi^z \phi_t + \beta_{\sigma_a}^z \sigma_{a,t} + resid, \quad Z \in \{R_{mkt}^{ex}; HML\}, \\
 TSS_t &= const. + \beta_x^{TSS} x_t + \beta_\phi^{TSS} \phi_t + \beta_{\sigma_a}^{TSS} \sigma_{a,t} \cdot \text{sign}(TSS_t) + resid,
 \end{aligned}$$

where  $R_{mkt}^{ex}$  and  $HML$  are the Fama-French quarterly market excess return and HML factors, respectively.  $TSS$  denotes the term structure spread between 7-year and 2-year zero-coupon equities (Binsbergen et al. 2013). The factors  $x_t$ ,  $\sigma_{a,t}$ , and  $\zeta_t$  are estimated according to the procedure described in section 4. Our estimates are from the five-factor specification with volatility estimated through projection methods. The  $\text{sign}(TSS_t)$  term accounts for the opposite impact of volatility on the spread depending on whether the term structure is either upward or downward sloping. For each regression, we report (i) our estimates for the exposure to both short-run conditional volatility ( $\sigma_{a,t}$ ) and relative volatility ( $\zeta_t$ ); (ii) the  $p$ -value associated with the null that the signs of our exposure coefficients are opposite to those estimated (we denote  $p$ -values smaller than 1%, 5%, and 10% by \*\*\*, \*\*, and \*, respectively); and (iii) the adjusted  $R^2$  from each regression with and without the inclusion of short-run volatility ( $\sigma_t$ ). Numbers in parentheses are GMM Newey-West adjusted standard errors.

shocks, as the aggregate payout. Second, the data suggest that adverse volatility shocks to either short- or long-run shocks are associated with lower investment. Consistent with prior studies (see, for example, Bloom et al. (2007) and Bloom (2009)), our volatility shocks are contractionary for

investment.<sup>10</sup>

**Volatility factors and asset prices in the data.** The second key feature of our model is the time-varying relative volatility of news shocks and contemporaneous productivity shocks. Using our proxies for the volatility of contemporaneous productivity shocks and relative volatility developed in section 4.1, we test the implications for our model on the relationship between the market equity premium, the value-growth spread, the slope of the term structure, and the measured volatility processes in the data.

In the data, the market excess return and the HML factor are from the K. French’s webpage. We interpret the latter as a proxy of the return differential between assets in place and growth options (see, among several others, Ai et al. (2012)). The proxy for the slope of the term structure ( $TSS_t$ ) is obtained through a quarterly interpolation of the data reported by Binsbergen et al. (2013), figure 1, maturity 7–2. We use the volatility processes estimated from our five-factor empirical model in order to work with a longer sample. We then run standard regressions that we detail in panel B of table 5.

First of all, we note that both short-run volatility and relative volatility have a statistically significant positive impact on the aggregate risk premium, consistent with our model. Removing short-run volatility produces just a marginal deterioration of the adjusted  $R^2$ , suggesting that relative long-run volatility carries a significantly higher market price of risk, as in our model.

Second, our model implies that the value spread increases with both the volatility of short-run shocks and the relative volatility, because the former enhances the overall risk compensation and the latter strengthens the effect of learning. In table 5, both of our volatility processes have on average a positive coefficient. Unfortunately, in this specification the inference is not sharp

---

<sup>10</sup>We note that our model cannot replicate these findings, as at the equilibrium total investment equals total savings and increases because of precautionary motives. This limitation is common to other production economies with zero government expenditure and no external trade. Explicitly accounting for countercyclical government expenditure and for current account adjustments can solve this problem. We leave these extensions for future studies.

enough. This result is consistent with the findings of Bansal et al. (2005), Bansal et al. (2007), and Hansen et al. (2006, 2008): in a short sample, it is nearly impossible to obtain statistically different risk exposures in the cross-section of returns. In the spirit of prior empirical literature, we take seriously our point estimates and interpret them as a sign of the positive dependence of the expected value premium on both short- and long-run volatility. Based on untabulated results, we note that if we use a GARCH approach to estimate volatilities, both of these coefficients stay positive and become statistically significant.

Third, as we highlight in earlier sections, in our model, the slope of the term structure is decreasing in the relative volatility, while an increase in the volatility of the contemporaneous productivity shock enhances the magnitude of risk compensation and therefore that of the slope of the term structure ( $TSS_t$ ). This suggests the following regression:

$$TSS_t = const. + \beta_x^{TSS} x_t + \beta_\zeta^{TSS} \zeta_t + \beta_{\sigma_a}^{TSS+} \sigma_{a,t} I(TSS_t > 0) + \beta_{\sigma_a}^{TSS-} \sigma_{a,t} I(TSS_t \leq 0) + resid, \quad (38)$$

where  $I$  is an indicator function. Since, we cannot reject the restriction  $\beta_{\sigma_a}^{TSS+} = -\beta_{\sigma_a}^{TSS-}$ , we have imposed it both for parsimony and to sharpen our inference.<sup>11</sup>

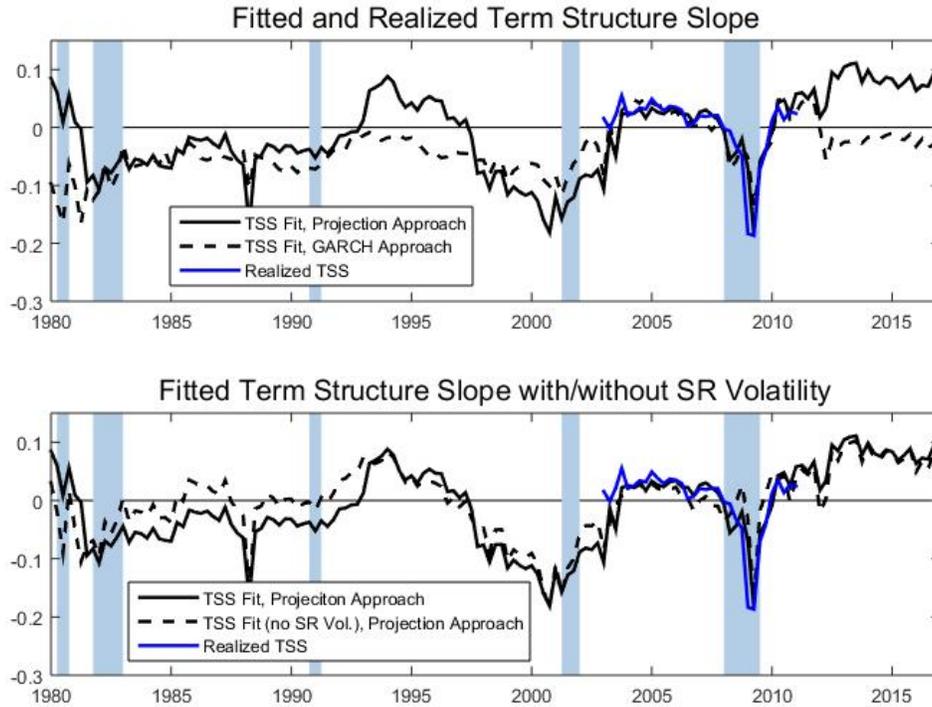
Because the relative volatility  $\zeta_t$  lowers the slope of the term structure, we expect  $\beta_\zeta^{TSS}$  to be negative. In addition, we expect  $\beta_{\sigma_a}^{TSS+}$  to be positive. These implications of the model are confirmed in table 5. It is also important to note that relative volatility is a key explanatory variable for the term spread, as denoted by the significant increase in the adjusted  $R^2$  of our regression. When both relative long-run volatility and short-run volatility are included as explanatory variables, our  $R^2$  is very sizable, with a value of 70%.

To better illustrate these points, in the bottom panel of figure 4 we depict both the realized and the fitted term structure slope with and without accounting for short-run risk volatility. Our

---

<sup>11</sup>Note that  $\text{sign}(x) = I(x > 0) - I(x \leq 0)$ , and hence we can estimate:

$$TSS_t = const. + \beta_x^{TSS} x_t + \beta_\phi^{TSS} \phi_t + \beta_{\sigma_a}^{TSS+} \sigma_{a,t} \text{sign}(TSS_t) + resid.$$

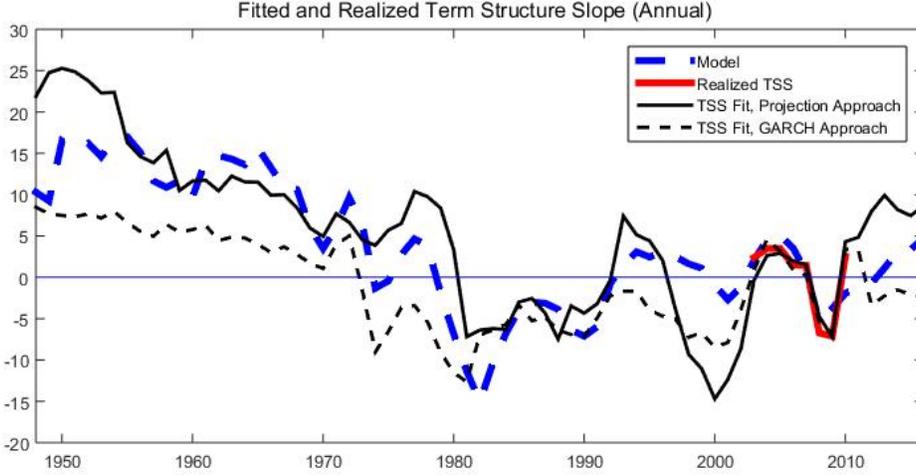


**Fig. 4: Volatility Factors and Term Structure Slope**

This figure shows both the realized (Binsbergen et al. (2013)) and the estimated term structure spread (TSS) between 7-year and 2-year zero-coupon equities. The fitted TSS values are obtained by estimating different versions of equation (38) over the sample 2002:Q4–2010:Q4. Values outside of this sample period are based on extrapolations. The relative volatility process,  $\zeta_t$ , and the conditional volatility of the productivity short-run shock,  $\sigma_{a,t}$  (denoted as SR Vol.), are obtained through the methods described in section 4. Grey bars denote NBER recession periods.

relative volatility factor explains most of the variability of the term structure slope. In the top panel of figure 4, we consider a sensitivity exercise and compare our results across the projection-based and GARCH(1,1)-based volatility measures. Both methods yield very similar in-sample results. Out of sample, in contrast, the two methods predict different signs for the term structure slope, especially prior to the mid-nineties. Specifically, the GARCH(1,1)-based volatility measures suggest that unconditionally the term structure may be downward sloping, and it becomes flatter with long-run risk moderation.

Given our short sample, however, these extrapolations are just suggestive and leave substantial



**Fig. 5: Annual Predictions on Term Structure Slope (TSS)**

This figure shows both the realized (Binsbergen et al. (2013)) and the estimated annual term structure spread (TSS) between 7-year and 2-year zero-coupon equities. The fitted TSS values are obtained by estimating different versions of equation (38) over the sample 2002:Q4–2010:Q4. Values outside of this sample period are based on extrapolations. The relative volatility process,  $\zeta_t$ , and the conditional volatility of the productivity short-run shock,  $\sigma_{a,t}$  (denoted as SR Vol.), are obtained through the methods described in section 4. The thick-dashed line refers to the output of our equilibrium model when we fit our estimated annual shocks.

uncertainty regarding the sign of the average slope of the equity term structure. On the positive side, our analysis contributes to the literature by linking the conditional slope of the term structure to macroeconomic fundamentals, specifically, conditional moments of productivity growth and investment dynamics.

**Model-implied  $TSS$ .** In figure 5, we show the realized annual  $TSS$  that we obtain by compounding the quarterly  $TSS$  over each calendar year. Similarly to figure 4, we also show the predictions from our empirical procedures compounded to an annual frequency. Furthermore, we show the  $TSS$  implied by our equilibrium model starting from 1948.<sup>12</sup>

<sup>12</sup>In order to connect our annual model with our quarterly data, we input annual equivalents of our exogenous variables in our equilibrium model. Specifically, (i) we annualize productivity growth by compounding quarterly rates during the year, and (ii) we sample the quarterly long-run growth component at the end of each year. We then recover both the annual short-run and the annual long-run shocks to match perfectly the dynamics of  $\Delta a_t$  and  $x_t$  at the annual frequency. The model perfectly match the annual time-series of both expected growth ( $x_t$ ) and realized growth ( $\Delta a_{t+1}$ ) with  $\rho_x = 0.77$  and  $\sigma = 2.8\%$ . We set all endogenous variables to their steady state value in 1948, and then we feed in the subsequent exogenous shocks that we obtain from the data.

We find two important results. First of all, our model replicates very well the pattern of  $TSS$  observed in the data by Binsbergen et al. (2013). Second, the time-series of  $TSS$  from our model features fluctuations consistent with those estimated through our regression approach. In particular, the moderation of short-run shocks causes the model-implied  $TSS$  to steadily decay until 1980 as in the data. Our model also implies that the boom episodes that have followed the last three recessions should also have a positive  $TSS$  because of significant reduction of long-run volatility.

## 6 Conclusions

We propose a production-based general equilibrium model to provide a unified explanation of the relationship between the timing of cash flows and their expected returns both for aggregate stock market dividends and for the cross section of book-to-market-sorted portfolios. The key mechanism in our paper is based on the interplay of learning about exposure to aggregate shocks, and the time-varying volatility of news regarding future productivity shocks. We show that our model is able to explain stylized facts about the time-varying slope of the term structure of dividend strips, as well as the negative relationship between cash-flow duration and expected returns in the cross section of equity returns. We also provide a novel empirical analysis linking news shocks, time variation in long-run news uncertainty, aggregate payouts, and equity term structure slope.

Our analysis abstracts away from optimal choice of financial leverage. A fully specified general equilibrium model with endogenous capital structure choices is beyond the scope of this paper, but this represents an important topic for future research. Future studies should also analyze the term structure of equity in a multicountry version of our intangible capital model in order to shed light on the international comovements documented in Binsbergen et al. (2013).

## References

- Ai, H., M. M. Croce, and K. Li. 2012. Toward a Quantitative General Equilibrium Asset Pricing Model with Intangible Capital. *Review of Financial Studies*. Forthcoming.
- Andries, M., T. Eisenbach, and M. C. Schmalz. Horizon-dependent risk aversion and the timing and pricing of uncertainty. Federal Reserve Bank of New York Working Paper 2017.
- Bansal, R., and I. Shaliastovich. 2013. A long-run risks explanation of predictability puzzles in bond and currency markets. *Review of Financial Studies* 26(1):1–33.
- Bansal, R., and A. Yaron. 2004. Risk for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance* 59:1481–509.
- Bansal, R., R. Dittmar, and C. Lundblad. 2005. Consumption, dividends, and the cross section of equity returns. *Journal of Finance* 60:1639–672.
- Bansal, R., R. F. Dittmar, and D. Kiku. 2007. Cointegration and consumption risks in asset returns. *Review of Financial Studies*, forthcoming.
- Belo, F., P. Colin-Dufresne, and R. Goldstein. 2014. Dividend dynamics and the term structure of dividend strips. *The Journal of Finance* 70(3).
- Berk, J., R. C. Green, and V. Naik. 1999. Optimal investment, growth options, and security returns. *Journal of Finance* 54:1553–608.
- Binsbergen, J., and R. S. Koijen. 2016. On the timing and pricing of dividends – reply. *American Economic Review* 106(10):3224–37.
- Binsbergen, J., and R. S. Koijen. 2017. The term structure of returns: facts and theory. *Journal of Financial Economics* 124(1):1–21.

- Binsbergen, J., M. W. Brandt, and R. S. Koijen. 2012. On the timing and pricing of dividends. *American Economic Review* 102(4):1596–618.
- Binsbergen, J., W. H. Hueskes, R. S. Koijen, and E. B. Vrugt. 2013. Equity yields. *Journal of Financial Economics* 110(3):503–19.
- Bloom, N. 2009. The impact of uncertainty shocks. *Econometrica* 77(3):623–685. ISSN 1468-0262.
- Bloom, N., S. Bond, and J. Van Reenen. 2007. Uncertainty and investment dynamics. *The Review of Economic Studies* 74(2):391–415.
- Blundell, R. W., and S. R. Bond. 2000. GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews* 19:321–340.
- Boguth, O., M. Carlson, A. J. Fisher, and M. Simutin. 2012. Leverage and the Limits of Arbitrage Pricing: Implications for Dividend Strips and the Term Structure of Equity Risk Premia. Working Paper.
- Carlson, M., A. Fisher, and R. Giammarino. 2004. Corporate investment and asset price dynamics: Implications for the cross-section of returns. *Journal of Finance* 59:2577–603.
- Choi, S., and J.-V. Rois-Rull. 2009. Understanding the dynamics of labor share: The role of noncompetitive factor prices. *Annals of Economics and Statistics*.
- Cooper, I. 2006. Asset pricing implications of nonconvex adjustment costs and irreversibility of investment. *Journal of Finance* 61:139–70.
- Croce, M. M. 2014. Long-run productivity risk: A new hope for production-based asset pricing. *Journal of Monetary Economics* 66:13–31.
- Croce, M. M., M. Lettau, and S. C. Ludvigson. Investor information, long-run risk, and the duration of risky cash-flows. 2015.

- Da, Z. 2009. Cash flow, consumption risk, and the cross-section of stock returns. *Journal of Finance* 64:923–56.
- David, A. 1997. Fluctuating confidence in stock markets: Implications for returns and volatility. *Journal of Financial and Quantitative Analysis* 32:427–462.
- Dechow, P., R. Sloan, and M. Soliman. 2004. Implied equity duration: A new measure of equity risk. *Review of Accounting Studies* 9:197–228.
- Epstein, L., and S. E. Zin. 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica* 57:937–69.
- Favilukis, J., and X. Lin. 2013. Wage rigidity: A quantitative solution to several asset pricing puzzles. *LSE Working Paper*.
- Gala, V. 2005. Investment and returns. Working Paper, London Business School.
- Garcia-Feijo, L., and R. D. Jorgensen. 2010. Can operating leverage be the cause of the value premium? *Financial Management* 39:1127–54.
- Gomes, J., L. Kogan, and L. Zhang. 2003. Equilibrium cross section of returns. *Journal of Political Economy* 111:693–732.
- Gourio, F. 2012. Disaster risk and business cycles. *American Economic Review*.
- Hansen, L. P., J. C. Heaton, and N. Li. 2006. Consumption strikes back. Working Paper, University of Chicago.
- Hansen, L. P., J. C. Heaton, and N. Li. 2008. Consumption strikes back? Measuring long-run risk. *Journal of Political Economy* 116:260–302.
- Hasler, M., and R. Marfe. 2017. Disaster recovery and the term structure of dividend strips. *Journal of Financial Economics* 122(1).

- Hsieh, C.-T., and P. J. Klenow. 2009. Misallocation and manufacturing tfp in china and india. *The Quaterly Journal of Economics* CXXIV(4):1403 – 1448.
- Jurado, K., S. Ng, and S. Ludvigson. 2015. Measuring uncertainty. *American Economic Review* 105(3):1177–215.
- Kogan, L., and D. Papanikolaou. 2010. Growth opportunities and technology shocks. *American Economic Review, P&P* 100:532:36.
- Kogan, L., and D. Papanikolaou. 2012. A theory of firm characteristics and stock returns: The role of investment-specific technology shocks. Working Paper.
- Kogan, L., and D. Papanikolaou. 2014. Growth opportunities and technology shocks. *Journal of Finance* 69:675–718.
- Kogan, L., D. Papanikolaou, and N. Stoffman. 2013. Technological innovation: Winners and losers. Working Paper.
- Kreps, D. M., and E. L. Porteus. 1978. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46:185–200.
- Lettau, M., and J. 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.
- Lettau, M., and J. A. Wachter. 2011. The Term Structure of Equity and Interest Rates. *Journal of Financial Economics* 101:90–113.
- Lin, X., and L. Zhang. 2007. The investment manifesto. *Journal of Monetary Economics*.
- Liu, L. X., T. M. Whited, and L. Zhang. 2009. Investment-based expected sotck returns. *Journal of Political Economy*.

- Marfe, R. 2017. Income insurance and the equilibrium term structure of equity. *The Journal of Finance*. forthcoming.
- Melitz, M. J. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71:1965 – 1725.
- Papanikolaou, D. 2011. Investment-specific shocks and asset prices. *Journal of Political Economy*. Forthcoming.
- Pastor, L., and P. Veronesi. 2009. Technological revolutions and stock prices. *American Economic Review* 99:1451–1483.
- Santos, T., and P. Veronesi. 2010. Habit formation, the cross section of stock returns and the cash flow risk puzzle. *Journal of Financial Economics* 98:385–413.
- Veronesi, P. 2000. How does information quality affect stock returns. *The Journal of Finance* 55 (2).
- Zhang, L. 2005. The value premium. *Journal of Finance* 60:67–103.

## Appendix A: Aggregation with learning

**Proof of Lemma 1.** Consider the resource allocation problem in (6). Suppose firms do not know  $A_j$  with certainty, but instead observe a noisy signal of it, denoted  $s$ . The expected output conditioning on  $s$  is  $E_s \left[ (Ak(s)^\alpha n(s)^{1-\alpha})^{\frac{\eta-1}{\eta}} \right]$ , where  $E_s$  denotes the belief about  $A_j$  given signal  $s$ . Note that firm  $j$ 's choice must be a function of its information. We use notations  $k(s)$  and  $n(s)$  to indicate that capital and labor input must be measurable functions of the signal  $s$ . Because there is a continuum of firms, we can assume that a version of the law of large numbers holds and compute the total output of the economy as

$$\left\{ \int E_s \left[ (Ak(s)^\alpha n(s)^{1-\alpha})^{\frac{\eta-1}{\eta}} \right] ds \right\}^{\frac{\eta}{\eta-1}}.$$

Therefore, maximization of total output can be written as

$$\begin{aligned} Y = \max \left\{ \int E_s \left[ (Ak(s)^\alpha n(s)^{1-\alpha})^{\frac{\eta-1}{\eta}} \right] ds \right\}^{\frac{\eta}{\eta-1}}. \\ \int k(s) ds = K, \\ \int n(s) ds = N. \end{aligned} \tag{A1}$$

The optimal policy of the above problem is given by the following lemma:

**Lemma 3** *The aggregate production function in (A1) can be written as*

$$Y = \mathbf{A}K^\alpha N^{1-\alpha}, \quad \text{where} \quad \mathbf{A} = \left[ \int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds \right]^{\frac{1}{\eta-1}}.$$

**Proof.** *Given  $s$ , firms maximize expected profit:*

$$\max E_s \left[ (Ak_s^\alpha n_s^{1-\alpha})^{\frac{\eta-1}{\eta}} \right] - Rk_s - Wn_s.$$

Optimality implies that the expected marginal product of capital must be equalized across firms:

$$\frac{E_s \left[ A^{\frac{\eta-1}{\eta}} \right] k_s^{\frac{\eta-1}{\eta}-1}}{E_{s'} \left[ A^{\frac{\eta-1}{\eta}} \right] k_{s'}^{\frac{\eta-1}{\eta}-1}} = 1$$

That is,

$$\frac{k_s}{k_{s'}} = \frac{n_s}{n_{s'}} = \left( \frac{E_s \left[ A^{\frac{\eta-1}{\eta}} \right]}{E_{s'} \left[ A^{\frac{\eta-1}{\eta}} \right]} \right)^\eta.$$

Therefore, the optimal choices are capital and labor must satisfy

$$\begin{aligned} k_s &= \frac{\left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta}{\int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds} K, & n_s &= \frac{\left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta}{\int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds} N, \\ E_s \left[ \left( A k_s^\alpha n_s^{1-\alpha} \right)^{\frac{\eta-1}{\eta}} \right] &= E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \left[ \left( \frac{\left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta}{\int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds} K \right)^\alpha \left( \frac{\left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta}{\int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds} N \right)^{1-\alpha} \right]^{\frac{\eta-1}{\eta}} \\ &= \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta \left( \frac{K^\alpha N^{1-\alpha}}{\int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds} \right)^{\frac{\eta-1}{\eta}}. \end{aligned}$$

As a result, total output can be written as:

$$\begin{aligned} \left[ \int (A_j k_j^\alpha n_j^{1-\alpha})^{\frac{\eta-1}{\eta}} dj \right]^{\frac{\eta}{\eta-1}} &= \left[ \int E_s \left\{ (A k_s^\alpha n_s^{1-\alpha})^{\frac{\eta-1}{\eta}} \right\} ds \right]^{\frac{\eta}{\eta-1}} \\ &= \left[ \int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds \right]^{\frac{\eta}{\eta-1}} \frac{K^\alpha N^{1-\alpha}}{\int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds} \\ &= \left[ \int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds \right]^{\frac{1}{\eta-1}} K^\alpha N^{1-\alpha}, \end{aligned}$$

as needed. ■

To prove Lemma 1 of the paper, note that given  $A_j = e^{\beta_j \Delta a}$ , and  $\beta_j \sim N(\mu, \frac{1}{\Delta a} \sigma^2)$ , under our assumption of the information structure, (3), the posterior distribution of  $\beta_j$  is normal with

$$\begin{aligned} \text{Var}_s[\beta] &= \frac{1}{\frac{1}{\Delta a} \sigma^2 + \frac{1}{\Delta a} \tau^2} = \frac{1}{\Delta a} \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}; \\ E_s[\beta] &= \text{Var}_s[\beta] \left[ \frac{1}{\frac{1}{\Delta a} \sigma^2} \mu + \frac{1}{\frac{1}{\Delta a} \tau^2} s \right] = \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}} \left[ \frac{1}{\sigma^2} \mu + \frac{1}{\tau^2} s \right] = \frac{1}{\tau^2 + \sigma^2} [\tau^2 \mu + \sigma^2 s]. \end{aligned}$$

Also, the signal  $s$  follows a normal distribution with mean  $\mu$  and variance  $\frac{1}{\Delta a} [\sigma^2 + \tau^2]$ . Therefore,

$$\begin{aligned} E_s \left[ A^{\frac{\eta-1}{\eta}} \right] &= E_s \left[ e^{\frac{\eta-1}{\eta} \Delta a \beta} \right] \\ &= e^{\frac{\eta-1}{\eta} \left\{ \Delta a \frac{\tau^2}{\tau^2 + \sigma^2} \mu + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \Delta a \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2} \right\} + \frac{\eta-1}{\eta} \Delta a \frac{\sigma^2}{\tau^2 + \sigma^2} s}, \end{aligned}$$

and

$$\left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta = e^{(\eta-1) \left\{ \Delta a \frac{\tau^2}{\tau^2 + \sigma^2} \mu + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \Delta a \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2} \right\} + (\eta-1) \Delta a \frac{\sigma^2}{\tau^2 + \sigma^2} s}.$$

We have

$$\begin{aligned} \int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds &= e^{(\eta-1) \left\{ \Delta a \frac{\tau^2}{\tau^2 + \sigma^2} \mu + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \Delta a \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2} \right\}} e^{(\eta-1) \Delta a \frac{\sigma^2}{\tau^2 + \sigma^2} \mu + \frac{1}{2} (\eta-1)^2 \Delta a \left( \frac{\sigma^2}{\tau^2 + \sigma^2} \right)^2 [\sigma^2 + \tau^2]} \\ &= e^{(\eta-1) \Delta a \left\{ \mu + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \left[ \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2} + \eta \frac{\sigma^4}{\tau^2 + \sigma^2} \right] \right\}} \end{aligned}$$

As a result, the aggregate productivity is

$$\mathbf{A} = \left[ \int \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta ds \right]^{\frac{1}{\eta-1}} = e^{\Delta a \left\{ \mu + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \frac{(1+\eta) \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2}}{\tau^2 + \sigma^2} \right\}}.$$

Under our normalization condition,  $\mu + \frac{1}{2} \frac{\eta-1}{\eta} \sigma^2 = 1$ , which implies  $\mu = 1 - \frac{1}{2} \frac{\eta-1}{\eta} \sigma^2$ , aggregate productivity can be written as:

$$\ln \mathbf{A} = \Delta a \left[ \mu + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \frac{\left( 1 + \eta \frac{\tau^{-2}}{\sigma^{-2}} \right)}{\tau^{-2} + \sigma^{-2}} \right],$$

where

$$\mu + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \frac{\left( 1 + \eta \frac{\tau^{-2}}{\sigma^{-2}} \right)}{\tau^{-2} + \sigma^{-2}} = 1 - \frac{1}{2} \frac{\eta-1}{\eta} \sigma^2 + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \frac{\left( 1 + \eta \frac{\tau^{-2}}{\sigma^{-2}} \right)}{\tau^{-2} + \sigma^{-2}} \quad (\text{A2})$$

$$= 1 + \frac{1}{2} \frac{(\eta-1)^2}{\eta} \frac{\sigma^4}{\tau^2 + \sigma^2}. \quad (\text{A3})$$

This completes the proof of Lemma 1.

**Proof of Lemma 2.** In the dynamic setup,  $\ln A_{j,t} = \sum_{i=1}^t \beta_{j,i} \Delta a_i$ . To save notation, we suppress subscripts  $t$  and  $j$  and write  $\ln A = \sum_{i=1}^t \beta_i \Delta a_i$ . Suppose the posterior variance for  $\beta_i$  is  $\frac{1}{\Delta a_i} \frac{1}{\sigma^{-2} + \tau_i^{-2}}$ . To prove Lemma 2, we apply the formula in Lemma 3 and first compute  $E_s \left[ A^{\frac{\eta-1}{\eta}} \right]$ .

Note that

$$E_s \left[ \frac{\eta-1}{\eta} \ln A \right] = \frac{\eta-1}{\eta} \left[ \sum_{i=1}^t \Delta a_i E_s(\beta_i) \right],$$

and

$$\begin{aligned} \text{Var}_s \left[ \frac{\eta-1}{\eta} \ln A \right] &= \left( \frac{\eta-1}{\eta} \right)^2 \sum_{i=1}^t \Delta a_i^2 \text{Var}_s(\beta_i), \\ &= \left( \frac{\eta-1}{\eta} \right)^2 \sum_{i=1}^t \Delta a_i \frac{1}{\sigma^{-2} + \tau_i^{-2}}. \end{aligned}$$

Therefore,

$$\left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta = e^{(\eta-1) \left[ \sum_{i=1}^t \Delta a_i E_s(\beta_i) \right] + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) (\eta-1) \sum_{i=1}^t \Delta a_i \frac{1}{\sigma^{-2} + \tau_i^{-2}}}.$$

We now need to compute  $\left\{ E \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta \right\}^{\frac{1}{\eta-1}}$ . Using the law of iterated expectations, we have

$$\begin{aligned} & E \left\{ (\eta-1) \left[ \sum_{i=1}^t \Delta a_i E_s(\beta_i) \right] + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) (\eta-1) \sum_{i=1}^t \Delta a_i \frac{1}{\sigma^{-2} + \tau_i^{-2}} \right\} \\ &= (\eta-1) \left[ \sum_{i=1}^t \Delta a_i \mu \right] + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) (\eta-1) \sum_{i=1}^t \Delta a_i \frac{1}{\sigma^{-2} + \tau_i^{-2}} \end{aligned}$$

Also, variance decomposition implies  $Var[\beta_i] = Var[\beta_i|s] + Var[E(\beta_i|s)]$ . Therefore,  $Var[E(\beta_i|s)] = \frac{1}{\Delta a_i} \left[ \sigma^2 - \frac{1}{\sigma^{-2} + \tau_i^{-2}} \right] = \frac{1}{\Delta a_i} \frac{\frac{\tau_i^{-2}}{\sigma_i^{-2}}}{\sigma^{-2} + \tau_i^{-2}}$ . We have

$$\begin{aligned} & Var \left\{ (\eta-1) \left[ \sum_{i=1}^t \Delta a_i E_s(\beta_i) \right] + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) (\eta-1) \sum_{i=1}^t \Delta a_i \frac{1}{\sigma^{-2} + \tau_i^{-2}} \right\} \\ &= (\eta-1)^2 \sum_{i=1}^t \Delta a_i \cdot \frac{\tau_i^{-2}/\sigma_i^{-2}}{\sigma^{-2} + \tau_i^{-2}}. \end{aligned}$$

As a result,

$$E \left[ \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta \right] = e^{Term},$$

where

$$\begin{aligned} Term &= (\eta-1) \left[ \sum_{i=1}^t \Delta a_i \mu \right] + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) (\eta-1) \sum_{i=1}^t \Delta a_i \frac{1}{\sigma^{-2} + \tau_i^{-2}} \\ &\quad + \frac{1}{2} (\eta-1)^2 \sum_{i=1}^t \Delta a_i \cdot \frac{\tau_i^{-2}/\sigma_i^{-2}}{\sigma^{-2} + \tau_i^{-2}} \\ &= (\eta-1) \left[ \sum_{i=1}^t \Delta a_i \mu \right] + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) (\eta-1) \sum_{i=1}^t \Delta a_i \frac{1 + \eta \tau_i^{-2}/\sigma_i^{-2}}{\sigma^{-2} + \tau_i^{-2}} \end{aligned}$$

Using Lemma 3, we have:

$$\begin{aligned} \mathbf{A} &= \left\{ E \left( E_s \left[ A^{\frac{\eta-1}{\eta}} \right] \right)^\eta \right\}^{\frac{1}{\eta-1}} \\ &= \exp \left\{ \sum_{i=0}^t \left( \mu + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \frac{1 + \eta \tau_i^{-2} / \sigma_i^{-2}}{\sigma^{-2} + \tau_i^{-2}} \right) a_i \right\}. \end{aligned} \quad (\text{A4})$$

Equations (12) and (13) can be obtained by using the definitions of  $\lambda^*$  and  $\lambda(\tau^2)$  and equation (A3) to simplify equation (A4) above.

**Recursive representation of the perpetual learning model.** To derive a recursive representation of the productivity of adolescent firms, we first prove the following lemma.

**Lemma 4** *Let the individual firms' productivity be given by (10). Suppose that at time  $t$ , for  $s = 1, 2, \dots, t$ , the posterior distribution of  $\beta_s$  is*

$$N \left( \frac{1}{\tau_s^2 + \sigma^2} [\tau_s^2 \mu + \sigma^2 s], \frac{1}{\Delta a_s} \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau_s^2}} \right).$$

*Suppose also, at time  $t+1$ , adolescent firms obtain a signal for all  $\Delta a_s$  with  $s = 0, 1, 2, \dots, t$  with  $e_s \sim N \left( 0, \frac{1}{\Delta a_s} \varpi_s^2 \right)$ . Then the aggregate productivity of all adolescent firms satisfy:*

$$\ln \bar{\mathbf{A}}_{t+1} - \ln \bar{\mathbf{A}}_t = \sum_{s=0}^t \xi_{t-s} \Delta a_s + \Delta a_{t+1}, \quad (\text{A5})$$

where

$$\xi_{t-s} = \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \frac{(\eta-1) \varpi_s^{-2}}{(\sigma^{-2} + \tau_s^{-2} + \varpi_s^{-2}) (\sigma^{-2} + \tau_s^{-2})}. \quad (\text{A6})$$

**Proof.** By our aggregation result in Lemma 2, the aggregate productivity of all adolescent firms is given by equation (A4). At time  $t+1$ , because adolescent firms have no information about  $\beta_{t+1}$ ; their posterior distribution of  $\beta_{t+1}$  is just  $N \left( \mu, \frac{1}{a_{t+1}} \sigma^2 \right)$ . Apply Lemma 2 again, the aggregate

productivity for all adolescent firms is

$$\bar{\mathbf{A}}_{t+1} = \exp \left\{ \sum_{i=0}^t \left( \mu + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \frac{1 + \eta (\tau_i^{-2} + \varpi_i^{-2}) / \sigma_i^{-2}}{\sigma^{-2} + \tau_i^{-2} + \varpi_i^{-2}} \right) a_i + \left( \mu + \frac{1}{2} \frac{\eta-1}{\eta} \sigma^2 \right) a_{t+1} \right\}. \quad (\text{A7})$$

Equation (A5) can be obtained by comparing (A7) and (A4) and setting

$$\begin{aligned} \xi_{t-s} &= \left( \mu + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \frac{1 + \eta (\tau_s^{-2} + \varpi_s^{-2}) / \sigma_s^{-2}}{\sigma^{-2} + \tau_s^{-2} + \varpi_s^{-2}} \right) - \left( \mu + \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \frac{1 + \eta \tau_s^{-2} / \sigma_s^{-2}}{\sigma^{-2} + \tau_s^{-2}} \right) \\ &= \frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \frac{(\eta-1) \varpi_s^{-2}}{(\sigma^{-2} + \tau_s^{-2} + \varpi_s^{-2}) (\sigma^{-2} + \tau_s^{-2})}, \end{aligned}$$

as needed. ■

In what follows, to save notation, we suppress the firm subscript  $j$ , use regular font for individual firm productivity, and use bold face for aggregate productivity. To derive the recursive representation in (17) and (16), we construct the time series of the quality of signals in our model as follows.

1. In period 0,  $\ln A_0 = \beta_0 \Delta a_0$ .  $\beta_0 \sim N \left( \mu, \frac{1}{\Delta a_0} \sigma^2 \right)$ . Therefore,  $\ln \bar{\mathbf{A}}_0 = \left( \mu + \frac{1}{2} \frac{\eta-1}{\eta} \sigma^2 \right) \Delta a_0 = \Delta a_0$ .
2. In period 1,  $\ln A_1 = \beta_1 \Delta a_1 + \beta_0 \Delta a_0$ , the firm drew a new signal  $\beta_0 + \varepsilon_0^1$ , where  $\varepsilon_0^1 \sim N \left( 0, \frac{1}{\Delta a_0} \tau_0^2 \right)$ . As a result, the posterior distributions are:  $\beta_1 \sim N \left( \mu, \frac{1}{\Delta a_1} \sigma^2 \right)$  and  $\beta_0 \sim N \left( E[\beta_0 | e_1(0)], \frac{1}{\Delta a_0} \frac{1}{\sigma^{-2} + \tau_0^{-2}} \right)$ . By (A5),

$$\ln \bar{\mathbf{A}}_1 - \ln \bar{\mathbf{A}}_0 = \xi_0 \Delta a_0 + \Delta a_1.$$

Given  $\rho_s \in (0, 1)$ , we can choose  $\tau_0$  so that

$$\frac{1}{2} \left( \frac{\eta-1}{\eta} \right) \frac{(\eta-1) \tau_0^{-2}}{(\sigma^{-2} + \tau_0^{-2}) \sigma^{-2}} = (1 - \rho_s) (\lambda^* - 1). \quad (\text{A8})$$

By setting  $t = s = 0$  and  $\varpi_0 = \infty$  in the definition of  $\xi_{t-s}$  in equation (A6), we have  $\xi_0 = (1 - \rho_s)(\lambda^* - 1)$  and

$$\ln \bar{\mathbf{A}}_1 - \ln \bar{\mathbf{A}}_0 = \xi_0 a_0 + a_1 = (1 - \rho_s)(\lambda^* - 1) \Delta a_0 + \Delta a_1.$$

3. In general, in period  $t + 1$ , the firm observes the following sequence of signals:

$$\begin{aligned} \beta_0 + \varepsilon_0^{t+1} & \quad \varepsilon_0^{t+1} \sim N\left(0, \frac{1}{\Delta a_0} \tau_t^2\right) \\ \beta_1 + \varepsilon_1^{t+1} & \quad \varepsilon_1^{t+1} \sim N\left(0, \frac{1}{\Delta a_1} \tau_{t-1}^2\right) \\ & \quad \dots \\ \beta_{t-1} + \varepsilon_{t-1}^{t+1} & \quad \varepsilon_{t-1}^{t+1} \sim N\left(0, \frac{1}{\Delta a_{t-1}} \tau_1^2\right) \\ \beta_t + \varepsilon_t^{t+1} & \quad \varepsilon_t^{t+1} \sim N\left(0, \frac{1}{\Delta a_t} \tau_0^2\right), \end{aligned}$$

Aggregation implies that  $\ln \bar{\mathbf{A}}_{t+1} - \ln \bar{\mathbf{A}}_t = \sum_{i=0}^t \xi_{t-i} a_i + a_{t+1}$  (which is just equation (A5)), where  $\xi_0$  is given by (A8), and in general,

$$\xi_{j+1} = \frac{1}{2} \left( \frac{\eta - 1}{\eta} \right) \frac{(\eta - 1) \tau_{j+1}^{-2}}{\left( \sigma^{-2} + \sum_{i=0}^j \tau_i^{-2} + \tau_{j+1}^{-2} \right) \left( \sigma^{-2} + \sum_{i=1}^j \tau_i^{-2} \right)}.$$

If we construct the sequence of  $\tau_j$  recursively as follows:  $\tau_0$  is defined by (A8) and  $\tau_{j+1}$  satisfies

$$\frac{1}{2} \left( \frac{\eta - 1}{\eta} \right) \frac{(\eta - 1) \tau_{j+1}^{-2}}{\left( \sigma^{-2} + \sum_{i=1}^j \tau_i^{-2} + \tau_{j+1}^{-2} \right) \left( \sigma^{-2} + \sum_{i=1}^j \tau_i^{-2} \right)} = \rho_s^{j+1} (1 - \rho_s)(\lambda^* - 1), \quad (\text{A9})$$

then the law of motion of  $\ln \bar{\mathbf{A}}_t$  can be written as:

$$\ln \bar{\mathbf{A}}_{t+1} - \ln \bar{\mathbf{A}}_t = \sum_{j=0}^t \xi_{t-j} \Delta a_j + \Delta a_{t+1} = \sum_{j=0}^t \rho_s^{t-j} (1 - \rho_s)(\lambda^* - 1) \Delta a_j + \Delta a_{t+1}. \quad (\text{A10})$$

Finally, define  $\chi_{t+1} = \ln \left( \frac{\hat{\mathbf{A}}_{t+1}}{\mathbf{A}_{t+1}} \right)$ . Equations (14) and (A10) together imply that  $\chi_t$  must satisfy the recursion (16) and  $\bar{\mathbf{A}}_t$  satisfies (17).

[Hengjie revision ends here.]

## Appendix B: Calibration of the Learning Parameters

In our model, the exposure of mature firms to the common productivity shock  $\Delta a_{t+1}$  ( $\lambda$ ) and the probability of transition from adolescence to maturity ( $\phi$ ) together determine firms' exposure to aggregate productivity shocks as a function of firms' capital age. Firms' capital age is what we measure in the data. In this section, we describe the exposure-age relationship in the model and the empirical procedure that we use to measure this moment in the data.

In our model, the firms' survival rate is  $1 - \delta$  per year. Therefore, the total measure of firms with age  $t$  is  $(1 - \delta)^{t-1}$ . Because firms become mature at the rate  $\phi$  per period, the fraction of adolescent firms among firms of age  $t$  is  $(1 - \phi)^{t-1}$ , and the fraction of mature firms is  $1 - (1 - \phi)^{t-1}$ . Since the exposure of mature firms is  $\lambda$  and the exposure of young firms is 1, the exposure of aggregate productivity with respect to  $\Delta a_{t+1}$  is a weighted average of the exposure of young and mature firms:

$$\delta \times \sum_{t=1}^{\infty} (1 - \delta)^{t-1} [(1 - \phi)^{t-1} + (1 - (1 - \phi)^{t-1}) \lambda].$$

Therefore, the exposure of firms of age  $t$  with respect to measured aggregate productivity shock is

$$\frac{(1 - \phi)^{t-1} + (1 - (1 - \phi)^{t-1}) \lambda}{\delta \times \sum_{t=1}^{\infty} (1 - \delta)^{t-1} [(1 - \phi)^{t-1} + (1 - (1 - \phi)^{t-1}) \lambda]}.$$

In order to empirically measure the exposure-age relationship in the data, we follow Ai et al. (2012) and use the annual data of publicly traded companies of US stock exchanges listed in both the COMPUSTAT and CRSP databases for the period 1950–2008. Our main goal is to pin down two parameters,  $\lambda$  and  $\phi$ , by matching moments of the exposure-age relationship in the data. We adopt the following empirical procedure to achieve this goal. In the first stage, we estimate the firm-level productivity,  $A_{i,j,t}$ , from the Cobb-Douglas production function. Here, we use  $i$  and  $j$  to index firm and industry, and we use  $t$  to denote time. The detailed estimation procedure is described in appendix 3.1.2 of Ai et al. (2012).

**Table B.1: Exposure to Aggregate Productivity Shocks by Age Groups**

Regression	$\xi_1$	
	$Age < 4$	$Age \geq 4$
(1)	0.63** (0.28)	1.08*** (0.11)
(2)	0.58** (0.26)	1.09*** (0.10)
Model	0.67	1.15

This table reports the aggregate productivity exposures of two firm groups based on a capital-age cutoff of four years. All estimates are based on the following regression (equation (B3)):

$$\Delta \ln A_{i,j,t} = \xi_{0i} + \xi_1 \Delta \ln \bar{A}_t + \tilde{\varepsilon}_{i,j,t}. \quad (\text{B2})$$

The exposures are normalized so that the firm exposure of the whole sample regression is equal to 1. Regressions (1) and (2) differ in that they use two alternative estimation methods in the first stage to estimate  $\Delta \ln A_{i,j,t}$ . Regression (1) is based on the fixed effect procedure, whereas regression (2) is based on the dynamic error component method of Blundell and Bond (2000). These estimation methods are described in appendix B of Ai et al. (2012). Numbers in parentheses are standard errors, and they are heteroskedasticity consistent and clustered at the firm level. We use \*, \*\*, and \*\*\* to indicate  $p$ -values smaller than 0.10, 0.05, and 0.01, respectively. In the last row (“Model”), we report the model-implied  $\xi_1$  based on our calibrated parameters,  $\lambda$  and  $\phi$ .

Second, as in Ai et al. (2012), we construct a firm-specific capital age measure,  $KAge_{i,t}$ , as a weighted average age of its capital vintages, i.e., investments ( $I_{i,t}$ ):

$$KAge_{i,t} = \frac{\sum_{l=1}^T (1 - \delta_i)^l \cdot I_{i,t-l} \cdot l}{\sum_{l=1}^T (1 - \delta_i)^l \cdot I_{i,t-l}}. \quad (\text{B1})$$

We choose  $T = 15$  but we obtain comparable results for different values of  $T$ , such as  $T = 5$  and  $T = 8$ .

Lastly, we estimate the exposure of firms’ productivity with respect to the aggregate productivity by different capital age groups using the following regression:

$$\Delta \ln A_{i,j,t} = \xi_{0i} + \xi_1 \Delta \ln \bar{A}_t + \tilde{\varepsilon}_{i,j,t}, \quad (\text{B3})$$

where  $\xi_{0i}$  controls for the firm-specific fixed effect, and  $\Delta \ln \bar{A}_t$  is the growth rate of aggregate

productivity as measured by the US Bureau of Labor Statistics (BLS). In particular, in order to determine our two parameters, we divide all firms into two groups based on a capital-age cutoff of four years, and use their group-specific aggregate productivity exposures to guide our calibration.

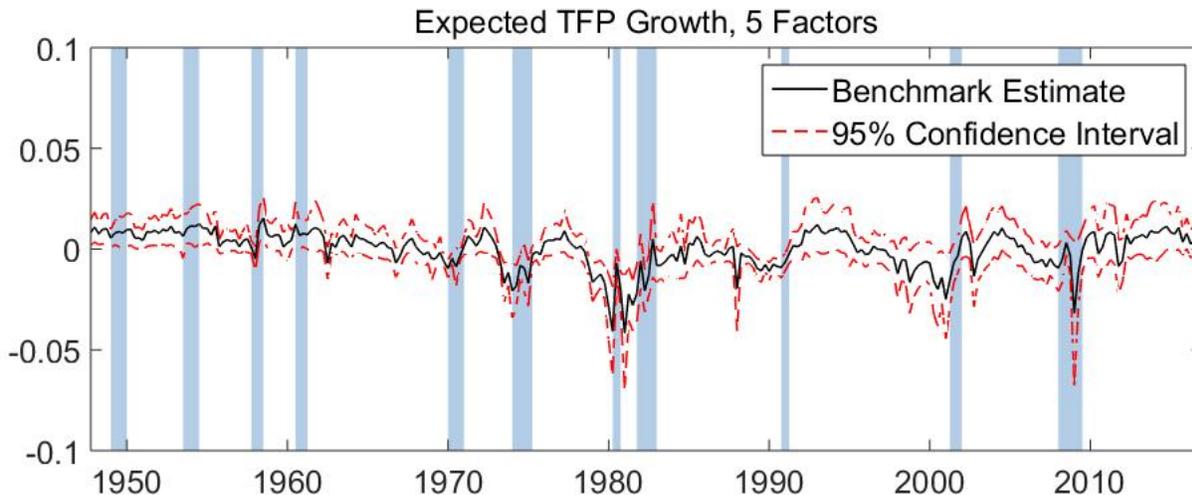
We report our estimation results in table B.1. To summarize the empirical results, we find that the firm group with the higher capital age has a significantly higher exposure to the aggregate productivity growth, consistent with the learning mechanism we emphasize in this manuscript. We calibrate the two parameters,  $\lambda = 6$  and  $\phi = 0.7$ , to target the point estimates of  $\xi_1$  obtained for our two different age groups. In the last row of table B.1 we report the model-implied exposures under our benchmark calibration, and we note that they are consistent with their empirical counterparts.

## Appendix C: Additional Empirical Results

**Table C.1: Time-Varying Volatility in Productivity (II)**

Volatility:	Hypothesis Testing			
	Factor-Based		GARCH-Based	
Mean:	5 Factors	13 Factors	5 Factors	13 Factors
$H_0 : \beta^x = 0$	18.73 (0.00)	267.05 (0.00)	18.31 (0.00)	269.06 (0.00)
$H_0 : \beta_x^\sigma = \beta_a^\sigma$	126.30 (0.00)	153.27 (0.00)	— (0.00)	— (0.00)
$H_0 : \rho_x = 0$	120.65 (0.00)	18.95 (0.00)	124.81 (0.00)	19.72 (0.00)
$H_0 : \rho_\zeta = 0$	142.85 (0.00)	12.66 (0.00)	34.80 (0.00)	12.56 (0.00)

We jointly estimate the set of equations (30)–(36) and report Wald tests as well as  $p$ -values (in parentheses) for the null hypotheses listed in the first column. We use US data spanning the sample 1947:Q1–2016:Q4, and all  $p$ -values are based on GMM Newey-West adjusted standard errors. Our log-volatility processes,  $\sigma_{a,t}$  and  $\sigma_{x,t}$ , are identified by either using the factor representation in equations (35)–(36) or by adopting two GARCH(1,1) models. Our five factors are the price-dividend ratio, three-month Treasury bill yield, three-year and five-year Treasury bond yields, and the integrated volatility of stock market returns. Our 13 factors are the principal components extracted by Jurado et al. (2015) from a wide cross section of macroeconomic and financial indicators.



**Fig. C.1: Fitted Long-Run Productivity Risk**

This figure shows the expected productivity growth estimated as detailed in section 4. The estimation is based on the benchmark specification with the four factors of Bansal and Shaliastovich (2013) and integrated stock market volatility. Quarterly US data start in 1947:Q1 and end in 2016:Q4. Grey bars denote NBER recession periods.

## Appendix D: Sensitivity Analysis

In this section, we address only moments that change significantly from our benchmark or that are important for our analysis. All results are summarized in table D.1.

**The role of information diffusion ( $\rho_s$ ).** The parameter  $\rho_s$  is tightly related to the ability of an adolescent firm to learn about its past productivity exposures. Most importantly, it determines the half-life of the productivity gap between mature and adolescent firms. We vary this parameter in order to change the half-life by  $\pm 20\%$  and offer the following remarks.

First, the sensitivity of our main results with respect to this parameter is very limited. Increasing  $\rho_s$  makes our learning friction more powerful, and as a result it increases slightly both the equity premium and the value premium. Second, focusing on macroeconomic aggregates, a higher  $\rho_s$  predicts a higher volatility of investment and a lower correlation with consumption, consistent with the data.

**The role of learning speed ( $\phi$ ).** Increasing the probability of becoming a mature firm, i.e., a firm with full information, is equivalent to speeding up the completion of the learning process. When we increase  $\phi$ , we reduce the fraction of capital allocated to adolescent firms, but the average age of aggregate capital remains unchanged because the age of both mature and adolescent firms declines.

According to our simulations, decreasing the share of young capital through a higher  $\phi$  makes the term structure even higher over short maturities because it mitigates the substitution effect even further compared to our benchmark. Equivalently, positive news shocks are associated with an even stronger income effect because all firms are expected to quickly take advantage of technological progress.

This improvement, however, comes with a lower risk premium for physical capital held by mature firms, because in the absence of adjustment costs a higher  $\phi$  makes the relative price of

**Table D.1: Sensitivity Analysis**

Panel A: the role of information diffusion ( $\rho_s$ )				
	Data	Benchmark	80%HL	120%HL
$\sigma(\Delta c)$	02.53 (00.56)	2.98	3.04	2.94
$\sigma(\Delta i)/\sigma(\Delta c)$	05.29 (00.50)	4.03	3.63	4.25
$\rho_{\Delta c, \Delta i}$	00.39 (00.15)	0.69	0.74	0.65
$E[r^f]$	00.89 (00.97)	0.44	0.38	0.40
$E[r_K^{L,ex}]$	05.70 (02.25)	4.05	3.91	4.13
$E[r_K^L - r_S^L]$	04.32 (01.39)	3.83	3.42	4.13
Panel B: the role of learning speed ( $\phi$ )				
	Data	Benchmark	$\phi = 0.6$	$\phi = 0.8$
$\sigma(\Delta c)$	02.53 (00.56)	2.98	2.95	3.02
$\sigma(\Delta i)/\sigma(\Delta c)$	05.29 (00.50)	4.03	3.92	4.18
$E[r^f]$	00.89 (00.97)	0.44	0.56	0.32
$E[r_K^{L,ex}]$	05.70 (02.25)	4.05	4.57	3.48
$E[r_K^L - r_S^L]$	04.32 (01.39)	3.83	4.13	2.98
RP(2)	10.08 (05.04)	6.77	3.39	9.46
Age $\bar{K}$		1.33	1.54	1.17
Age $\hat{K}$		11.50	11.74	11.31
$\bar{K}/K^{tot}$		15%	18%	14%
Age $K^{tot}$		9.90	9.91	9.89
Panel C: the role of intangible capital ( $\nu$ )				
	Data	Benchmark	No Intang.	
$E[I/Y]$	00.15 (00.05)	0.17	0.31	
$\sigma(\Delta c)$	02.53 (00.56)	2.98	2.96	
$\sigma(\Delta i)/\sigma(\Delta c)$	05.29 (00.50)	4.03	4.23	
$\sigma(\Delta n)$	02.07 (00.21)	1.51	1.63	
$AC_1(\Delta C)$	0.49 (00.15)	0.40	0.41	
$\rho_{\Delta c, \Delta n}$	00.28 (00.07)	0.55	0.49	
$\rho_{\Delta c, \Delta i}$	00.39 (00.15)	0.69	0.64	
$E[r^f]$	00.89 (00.97)	0.44	0.40	
$E[r_K^{L,ex}]$	05.70 (02.25)	4.05	3.96	
RP(2)	10.08 (05.04)	6.77	8.30	

All entries for the models and for the data are obtained as in table 2. In panel A, we vary the parameter  $\rho_s$  so that the half-life (HL) of the cointegration residual  $\chi_t$  is modified by  $\pm 20\%$  relative to the benchmark. In panel B, we change the parameter  $\phi$ , and in panel C we remove intangible capital from the model by setting  $\nu = 1$ .

adolescent and mature firms less volatile and close to 1. As a natural byproduct of this phenomenon, we also observe a decline in our value premium.

**The role of intangible capital ( $\nu$ ).** In panel C of table D.1, we show that intangible capital in our setting is important only in order to have a well-defined concept of the value premium. Absent an interest in the relation between equity excess returns and duration in the cross section of book-to-market-sorted firms, intangible capital does not play a crucial role. The only change worthy of notice is the increased  $RP(2)$ . If we remove growth options from the model,  $RP(2)$  increases to 8.30, a result still within the available empirical confidence intervals.