

NBER WORKING PAPER SERIES

DIVIDEND DYNAMICS, LEARNING, AND EXPECTED STOCK INDEX RETURNS

Ravi Jagannathan  
Binying Liu

Working Paper 21557  
<http://www.nber.org/papers/w21557>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
September 2015

We thank Jonathan Berk, Jules van Binsbergen, Wayne Ferson, Lawrence Harris, Gerard Hoberg, Kai Li, Narayan Naik, and seminar participants at HKUST, London Business School, Purdue University, Texas A&M University, University of Southern California, for helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Ravi Jagannathan and Binying Liu. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Dividend Dynamics, Learning, and Expected Stock Index Returns  
Ravi Jagannathan and Binying Liu  
NBER Working Paper No. 21557  
September 2015, Revised October 2106  
JEL No. G10,G11,G12

**ABSTRACT**

We show that, in a frictionless and efficient market, an asset pricing model that better describes investors' behavior should better forecast stock index returns. We propose a dividend model that predicts, out-of-sample, 31.3% of the variation in annual dividend growth rates (1976-2015). Further, when learning about dividend dynamics is incorporated into a long-run risks model, the model predicts, out-of-sample, 22.4% of the variation in annual stock index returns (1976-2015). This supports the view that both investors' aversion to long-run risks and learning about these risks are important in determining asset prices and expected returns.

Ravi Jagannathan  
Kellogg Graduate School of Management  
Northwestern University  
2001 Sheridan Road  
Leverone/Anderson Complex  
Evanston, IL 60208-2001  
and NBER  
rjaganna@kellogg.northwestern.edu

Binying Liu  
Northwestern University  
Kellogg School of Management  
2001 Sheridan Road  
Evanston, IL 60208  
b-liu@kellogg.northwestern.edu

The average return on equities has been substantially higher than the average return on risk free bonds over long periods of time. Between 1946 and 2015, the S&P500 earned 63 basis points per month more than 30 days T-bills (i.e. over 7% annualized). Over the years, many dynamic equilibrium asset pricing models have been proposed in the literature to understand the nature of risks in equities that require such a large premium and why risk free rates are so low. A common feature in most of these models is that risk premium on equities does not remain constant over time, but varies in a systematic and stochastic manner. A large number of academic studies have found support for such predictable variation in equity premium.<sup>1</sup> This led Lettau and Ludvigson (2001) to conclude "it is now widely accepted that excess returns are predictable by variables such as price-to-dividend ratios."

Goyal and Welch (2008) argue that variables such as price-to-dividend ratios, although successful in predicting stock index returns in-sample, fail to predict returns out-of-sample. The difference between in-sample and out-of-sample prediction is the assumption made on investors' information set. Traditional dynamic equilibrium asset pricing models assume that, while investors' beliefs about investment opportunities and economic conditions change over time and drive the variation in stock index prices and expected returns, these investors nevertheless have complete knowledge of the parameters describing the economy. For example, these models assume that they know the true model and model parameters governing consumption and dividend dynamics. However, as Hansen (2007) argues, "this assumption has been only a matter of analytical convenience" and is unrealistic in that it requires us to "burden the investors with some of the specification problems that challenge the econometrician". Motivated by this insight, a recent but growing literature has focused on the role of learning in asset pricing models. Timmermann (1993) and Lewellen and Shanken (2002) demonstrate, via simulations, that parameter uncertainty can lead to excess predictability and volatility in stock returns. Johannes, Lochstoer, and Mou (2014) propose a Markov-switching model for consumption dynamics and show that learning about the consumption process is reflected in asset prices and expected returns. Croce, Lettau, and Ludvigson (2014) show that a bounded rationality limited information long-run risks model can generate a downward-sloping equity term structure. Collin-Dufresne, Johannes, and Lochstoer (2016) provide the theoretical foundation that

---

<sup>1</sup>See, among others, Campbell and Shiller (1988b), Fama and French (1993), Lamont (1998), Baker and Wurgler (2000), Lettau and Ludvigson (2001), Campbell and Vuolteenaho (2004), Lettau and Ludvigson (2005), Polk, Thompson, and Vuolteenaho (2006), Ang and Bekaert (2007), van Binsbergen and Kojen (2010), Kelly and Pruitt (2013), van Binsbergen, Hueskes, Kojen, and Vrugt (2013), Li, Ng, and Swaminathan (2013), and Da, Jagannathan, and Shen (2014).

parameter learning can be a source of long-run risks under Bayesian learning.<sup>2</sup> We add to this literature.

The main contributions of our paper, which distinguish it from the existing literature on the interaction between learning and asset pricing, is as follows. First, we show that, when equity markets are frictionless and efficient, an asset pricing model that is closer to the true asset pricing model, i.e. the model that better describes investors' behavior, should better forecast stock index returns. This provides the theoretical foundation for the economic significance of an asset pricing model's performance in forecasting annual stock index returns as a measure to assess that model. Then, we show that, when learning about dividend dynamics is incorporated into a long-run risks model, the model predicts as much as 22.4 percent of the variation in annual stock index returns between 1976 and 2015 out-of-sample. This not only addresses the Goyal and Welch (2008) critique and significantly revises upward the degree of return predictability documented in the existing literature, but also lends support to the view that both investors' aversion to long-run risks and learning about these risks play important roles in determining asset prices and expected returns.<sup>34</sup>

To study how learning about dividend dynamics affects stock index prices and expected returns, we first need a dividend model that is able to realistically capture how investors form expectations about future dividends. Inspired by Campbell and Shiller (1988b), we put forth a model of dividend growth rates that incorporates information in corporate payout policy into the latent variable model used in Cochrane (2008), van Binsbergen and Koijen (2010), and others. Our model explains serial correlations in annual dividend growth rates up to 5 years. Further, our model predicts 42.8 percent of the variation in annual dividend growth rates between 1946 and 2015 in-sample and predicts 33.4 percent of the variation in annual dividend growth rates between 1976 and 2015 out-of-sample. Based on these results, we comfortably reject the null that expected dividend growth rates are constant and demonstrate that the superior performance of our dividend model over alternative models in predicting annual dividend growth rates is significant.

We document that uncertainties about parameters in our dividend model, especially parameters surrounding the persistent latent variable, are high and resolve slowly. That

---

<sup>2</sup>Instead of learning, an alternative path that researchers have taken is through introducing preferences shocks. See Albuquerque, Eichenbaum, and Rebelo (2015).

<sup>3</sup>Our paper is also consistent with the argument of Lettau and Van Nieuwerburgh (2008) that steady-state economic fundamentals, or in our interpretation, investors beliefs about these fundamentals, vary over time and these variations are critical in determining asset prices and expected returns.

<sup>4</sup>Following the existing literature, we adopt the stock index as a proxy for the market portfolio.

is, these uncertainties remain substantial even at the end of our 70 years data sample, suggesting that learning about dividend dynamics is a difficult and slow process. Further, when our dividend model is estimated at each point in time based on data available at the time, model parameter estimates fluctuate, some significantly, over time as more data become available. In other words, if investors estimate dividend dynamics using our model, we expect their beliefs about the parameters governing the dividend process to vary significantly over time. We then show that these changes in investors' beliefs can have large effects on their expectations of future dividends. Through this channel, changes in investors' beliefs about the parameters governing the dividend process can contribute significantly to the variation in stock prices and expected returns.

We provide evidence that investors behave as if they learn about dividend dynamics and price stocks using our model. First, we define stock yields as discount rates that equate the present value of expected future dividends to the current prices of the stock index. From the log-linearized present value relationship of Campbell and Shiller (1988), we write stock yields as a function of price-to-dividend ratios and long-run dividend growth expectations. We show that, assuming that investors learn about dividend dynamics, these stock yields explain 14.9 percent of the variation in annual stock index returns between 1976 and 2015. In comparison, stock yields, assuming full information, predict only 9.9 percent of the same variation. Next, we embed our dividend model into an dynamic equilibrium asset pricing model that features Epstein and Zin (1989) preferences, which capture preferences for the early resolution of uncertainty, and consumption dynamics from the long-run risks model of Bansal and Yaron (2004). We refer to this model as our long-run risks model. We find that, assuming learning, our long-run risks model predicts 22.4 percent of the variation in annual stock index returns between 1976 and 2015. Learning accounts for over half of the 22.4 percent. Both the model's forecasting performance and the incremental contribution of learning to this performance are significant.

Our results suggest that, aside from a common persistent component in consumption and dividend growth rates, the assumption that investors hold Epstein and Zin (1989) preferences with early resolution of uncertainty, a critical component of any long-run risks model, is essential to the model's strong performance in predicting annual stock index returns. More specifically, we find that, replacing Epstein and Zin (1989) preferences with constant relative risk aversion (CRRA) preferences,  $R$ -square value for predicting annual stock index returns, between 1976 and 2015, drops from 6.4 percent to 4.9 percent

for the case of full information and from 22.4 percent to 11.1 percent after learning is incorporated. This substantial deterioration in forecasting performance is evidence that the assumption of early resolution of uncertainty, as modeled through Epstein and Zin (1989), brings us steps closer to discovering the true asset pricing model.

We follow Cogley and Sargent (2008), Piazzesi and Schneider (2010), and Johannes, Lochstoer, and Mou (2014), and define learning based on the anticipated utility of Kreps (1998), where agents update using Bayes' law but optimize myopically in that they do not take into account uncertainties associated with learning in their decision making process. That is, anticipated utility assumes agents form expectations not knowing that their beliefs will continue to evolve going forward in time as the model keeps updating. Given the relative complexity of our long-run risks model and the multi-dimensional nature of learning, we find that solving our model with parameter uncertainties as additional risk factors is too computationally prohibitive.<sup>5</sup> Therefore, we adopt the anticipated utility approach as the more convenient alternative.

The rest of this paper is as follows. In Section 1, we introduce our dividend model and evaluate its performance in capturing dividend dynamics. In Section 2, we show that investors' beliefs about dividend model parameters can vary significantly over time as a result of Kreps' learning about dividend dynamics. In Sections 3, we show that learning accounts for a significant fraction of the variation in both long-run and short-run expected stock index returns. In Section 4, we first discuss how an asset pricing model's performance in predicting stock index returns can be used as a criterion to evaluate that model. Then, we demonstrate that, between 1976 and 2015, a model that incorporates Kreps' learning into a long-run risks model predicts 22.4 percent of the variation in annual stock index returns and explain why such a finding provides us insight into investor preferences and the role of learning in describing investors' behavior. In Section 5, we conclude.

## 1 The Dividend Model

In this section, we present a model for dividend growth rates that extends the latent variable model of Cochrane (2008), van Binsbergen and Kojien (2010), and others by incorporating information in corporate payout policy into the model. The inclusion of corporate payout policy in explaining dividend dynamics is inspired by Campbell and

---

<sup>5</sup>Collin-Dufresne, Johannes, and Lochstoer (2016) provide the theoretical foundation for studying uncertainties about model parameters as priced risk factors.

Shiller (1988b), who show that cyclical-adjusted price-to-earnings (CAPE) ratios, defined as the log ratios between real prices and real earnings averaged over the past decade, can predict future growth rates in dividends.

We begin with the latent variable model in Cochrane (2008), van Binsbergen and Koijen (2010), and others. Let  $D_t$  be nominal dividend of the stock index,  $d_t = \log(D_t)$ , and  $\Delta d_{t+1} = d_{t+1} - d_t$  be log dividend growth rate. The model is described as:

$$\begin{aligned} \Delta d_{t+1} - \mu_d &= x_t + \sigma_d \epsilon_{d,t+1} \\ x_{t+1} &= \rho x_t + \sigma_x \epsilon_{x,t+1} \\ \begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \end{pmatrix} &\sim \text{i.i.d. } \mathbb{N} \left( \mathbf{0}, \begin{pmatrix} 1 & \lambda_{dx} \\ \lambda_{dx} & 1 \end{pmatrix} \right), \end{aligned} \quad (1)$$

where time- $t$  is defined in years to control for potential seasonality in dividend payments. Following van Binsbergen and Koijen (2010), we fit our model to the nominal dividend process. As shown in Boudoukh, Michaely, Richardson, and Roberts (2007), equity issuance and repurchase tend to be more sporadic and random compared to cash dividends. For this reason, we focus on modeling the cash dividend process. In (1), expected dividend growth rates are a function of the latent variable  $x_t$ , the unconditional mean  $\mu_d$  of dividend growth rates, and the persistence coefficient  $\rho$  of the latent variable  $x_t$ :

$$E_t [\Delta d_{t+s+1}] = \mu_d + \rho^s x_t, \quad \forall s \geq 0. \quad (2)$$

Before we introduce corporate payout policy into this model, we first recall the dividend model used in Campbell and Shiller (1988b). Define  $p_t$  as log nominal price of the stock index,  $e_t$  as log nominal earnings,  $\pi_t$  as log consumer price index, and, following Campbell and Shiller (1988b), consider the following vector-autoregression for annual nominal dividend growth rates, log price-to-dividend ratios, and CAPE ratios:

$$\begin{aligned} \begin{pmatrix} \Delta d_{t+1} \\ p_{t+1} - d_{t+1} \\ p_{t+1} - \bar{e}_{t+1} \end{pmatrix} &= \begin{pmatrix} b_{10} \\ b_{20} \\ b_{30} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} \Delta d_t \\ p_t - d_t \\ p_t - \bar{e}_t \end{pmatrix} + \begin{pmatrix} \sigma_d \epsilon_{d,t+1} \\ \sigma_{(p-d)} \epsilon_{(p-d),t+1} \\ \sigma_{(p-\bar{e})} \epsilon_{(p-\bar{e}),t+1} \end{pmatrix}, \\ \begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{(p-d),t+1} \\ \epsilon_{(p-\bar{e}),t+1} \end{pmatrix} &\sim \text{i.i.d. } \mathbb{N} \left( \mathbf{0}, \begin{pmatrix} 1 & \lambda_{12} & \lambda_{13} \\ \lambda_{12} & 1 & \lambda_{23} \\ \lambda_{13} & \lambda_{23} & 1 \end{pmatrix} \right). \end{aligned} \quad (3)$$

where, as in Campbell and Shiller (1988b), CAPE ratio is defined as:

$$p_t - \bar{e}_t = p_t - \left( \pi_t + \frac{1}{10} \sum_{s=1}^{10} (e_{t-s+1} - \pi_{t-s+1}) \right). \quad (4)$$

Estimates of  $b_{10}$ ,  $b_{11}$ ,  $b_{12}$ , and  $b_{13}$  from (3), based on data between 1946 and 2015, are reported in the first row of Table 1.<sup>6</sup> We see that both price-to-dividend ratios and CAPE ratios have significant effects on future dividends, but in the opposite direction. That is, increases in price-to-dividend ratios predict increases in future dividend growth rates, but increases in CAPE ratios predict decreases in future dividend growth rates. Further, we note from Table 1 that  $b_{12} + b_{13} = 0$  cannot be statistically rejected. For this reason, we restrict  $b_{13} = -b_{12}$  and re-write (3) as:

$$\Delta d_{t+1} = \beta_0 + \beta_1 \Delta d_t + \beta_2 (\bar{e}_t - d_t) + \sigma_d \epsilon_{d,t+1}, \quad \epsilon_{d,t+1} \sim \text{i.i.d } \mathbb{N}(0, 1). \quad (5)$$

Stock index price  $p_t$  does not appear in (5). Instead, dividend growth rates over the next year are a function of some measure of retention ratios, i.e.  $\bar{e}_t - d_t$ . Estimated coefficients from (5) are in the second row of Table 1. We see that the  $\beta_2$  estimate is significant, suggesting that expected dividend growth rates respond to corporate payout policy. High earnings relative to dividends implies that firms have been retaining earnings in the past and so they are expected to pay more dividends in the future.

---



---

$b_{10}$	$b_{11}$	$b_{12}$	$b_{13}$
-0.004	0.478***	0.154***	-0.167***
(0.061)	(0.111)	(0.058)	(0.067)
$\beta_0$	$\beta_1$	$\beta_2$	
-0.038	0.482***	0.143***	
(0.028)	(0.111)	(0.056)	

---



---

Table 1: **Campbell and Shiller (1988b) Betas for Predicting Dividend Growth Rates:** This table reports coefficients from predicting annual nominal dividend growth rates using (3) and (5), based on data between 1946 and 2015. Newey and West (1987) adjusted standard errors are reported in parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using \*, \*\*, and \*\*\*.

We extend (1) based on this insight that corporate payout policy contains information

---

<sup>6</sup>We report results based on overlapping monthly data. In each month, we fit or predict dividend growth rates and stock index returns over the next 12 months. We report standard errors,  $F$ -statistics,  $p$ -values, and  $Q$ -statistics adjusted to reflect the dependence introduced by overlapping monthly data.



about future dividends. Define  $\Delta e_{t+1} = e_{t+1} - e_t$  as log nominal earnings growth rate and  $q_t = e_t - d_t$  as log earnings-to-dividend ratio, i.e. retention ratio. We write our dividend model as the following system of equations:

$$\begin{aligned}\Delta d_{t+1} - \mu_d &= x_t + \phi(q_t - \mu_q) + \sigma_d \epsilon_{d,t+1}, \\ x_{t+1} &= \rho x_t + \sigma_x \epsilon_{x,t+1}, \\ q_{t+1} - \mu_q &= \theta(q_t - \mu_q) + \sigma_q \epsilon_{q,t+1}, \\ \begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{q,t+1} \end{pmatrix} &\sim \text{i.i.d. } \mathbb{N} \left( \mathbf{0}, \begin{pmatrix} 1 & \lambda_{dx} & \lambda_{dq} \\ \lambda_{dx} & 1 & \lambda_{xq} \\ \lambda_{dq} & \lambda_{xq} & 1 \end{pmatrix} \right).\end{aligned}\tag{6}$$

In our model, dividend growth rates over the next year are a linear combination of three components. First, they consist of the latent variable  $x_t$ , which follows a stationary AR[1] process. Second, they are affected by changes in retention ratios. That is, we expect firms to pay more future dividends if they have retained more earnings. Third, they consist of white noises  $\epsilon_{d,t}$ . For convenience, we model retention ratios as an AR[1] process, and assuming that it is stationary implies that dividend and earnings growth rates have the same unconditional mean  $\mu_d$ . In (6), expected dividend growth rates are:

$$E_t[\Delta d_{t+s+1}] = \mu_d + \rho^s x_t + \phi \theta^s (q_t - \mu_q), \quad \forall s \geq 0.\tag{7}$$

This means that, aside from the latent variable  $x_t$  and retention ratios, expected dividend growth rates are a function of the unconditional mean  $\mu_d$  of dividend growth rates, the unconditional mean  $\mu_q$  and persistence  $\theta$  of retention ratios, the persistence  $\rho$  of the latent variable  $x_t$ , and coefficient  $\phi$  that connects corporate payout policy to dividend dynamics. The earnings process is not modeled explicitly in (6). However, because earnings growth rates are, by definition, a function of dividend growth rates and retention ratios, i.e.:

$$\Delta e_{t+1} = (q_{t+1} - q_t) + \Delta d_{t+1},\tag{8}$$

and because both dividend growth rates and retention ratios are modeled in (6), we can solve for earnings growth rates as:

$$\Delta e_{t+1} = \mu_d + x_t + (\theta + \phi - 1)(q_t - \mu_q) + \sigma_e \epsilon_{e,t+1}, \quad \epsilon_{e,t+1} \sim \mathbb{N}(0, 1),\tag{9}$$

where  $\sigma_e = \sqrt{\sigma_d^2 + \sigma_q^2 + 2\sigma_d\sigma_q\lambda_{dq}}$  and  $\epsilon_{e,t+1} = \frac{\sigma_d\epsilon_{d,t+1} + \sigma_q\epsilon_{q,t+1}}{\sigma_e}$ .

A type of model commonly used to forecast macroeconomic variables is a Markov-

switching model. For example, the Markov-switching model is used to describe consumption dynamics in Johannes, Lochstoer, and Mou (2014). The same model can be applied to dividend growth rates:

$$\begin{aligned}\Delta d_{t+1} &= \mu_d(s_t) + \sigma_d(s_t)\epsilon_{d,t+1}, \quad s_t \in \{1, 2, 3\}, \\ \mathbb{P}(s_{t+1} = i | s_t = j) &= \phi_{ij}, \\ \phi_{ij} \in [0, 1] \quad \forall i, j \in \{1, 2, 3\}, \quad \sum_{i=1}^3 \phi_{ij} &= 1 \quad \forall j \in \{1, 2, 3\}.\end{aligned}\tag{10}$$

That is,  $\mathbb{P}(s_{t+1} = i | s_t = j)$  is the probability that the economy transfers from state  $j \in \{1, 2, 3\}$  to state  $i \in \{1, 2, 3\}$ , and  $\mu_d(s_t)$  and  $\sigma_d(s_t)$  are the mean and standard deviation of dividend growth rates in a particular state.<sup>7</sup> A key feature of this model that is not present in dividend models discussed so far is that it is able to incorporate, albeit in a restricted manner, both regime changes and stochastic volatility. We adopt this model as another baseline and compare it against our dividend model in our subsequent analysis.

## 1.1 Data and Estimation

Due to the lack of reliable historical earnings data on the CRSP value-weighted market index, we use the S&P500 index as the proxy for the market portfolio. That is, throughout this study, data on prices, dividends, and earnings are from the S&P500 index. These data can be found on Prof. Robert Shiller’s website.

We compute the likelihood of our dividend model (i.e. (6) and (9)) using Kalman filters (Hamilton (1994)) and estimate model parameters,

$$\Theta = \{\mu_d, \phi, \sigma_d, \rho, \sigma_x, \mu_q, \theta, \sigma_q, \lambda_{dx}, \lambda_{dq}, \lambda_{xq}\}\tag{11}$$

based on maximum-likelihood. See Appendix A.1 for details. Table 2 reports model parameter estimates based on overlapping annual data between 1946 and 2015. Standard errors are based on bootstrap simulation, described in Appendix A.2. Because  $\lambda_{dx}$  and  $\lambda_{xq}$  are not significantly different from zero, we force these correlation parameters to be zeros in our subsequent analysis for the purpose of parameter reduction.<sup>8</sup> Previous

<sup>7</sup>An alternative approach of incorporating time varying parameters and stochastic volatility is by estimating parameters in a dividend model at each point in time using a rolling-window of historical data. We explore this option in the Online Appendix.

<sup>8</sup>In the Online Appendix, we replicate our main results using a data sample that stretches to prior to

works have suggested a regime shift in dividend dynamics before and after World War II. Fama and French (1988) note that dividends are more smoothed in the post-war period. Chen, Da, and Priestley (2012) argue that the lack of predictability in dividend growth rates by price-to-dividend ratios in the post-war period is attributable to this dividend smoothing behavior. Consistent with our intuition, coefficient  $\phi$  that connects corporate payout policy to dividend dynamics is estimated to be positive and significant. That is, high retention ratios imply high future dividend growth rates. The annual persistence of retention ratios is estimated to be 0.302. The latent variable  $x_t$  is estimated to be more persistent at 0.496. So there is a moderate level of persistence in dividend growth rates between 1946 and 2015 based on estimates from our model.

---



---

$\mu_d$	$\phi$	$\sigma_d$
0.060	0.139	0.025
(0.015)	(0.018)	(0.013)
	$\rho$	$\sigma_x$
	0.496	0.046
	(0.160)	(0.009)
$\mu_q$	$\theta$	$\sigma_q$
0.722	0.302	0.276
(0.047)	(0.116)	(0.027)
$\lambda_{dx}$	$\lambda_{dq}$	$\lambda_{xq}$
0.016	0.858	0.034
(0.133)	(0.032)	(0.128)

---



---

Table 2: **Dividend Model Parameter Estimates:** This table reports estimated parameters from our dividend model, based on data between 1946 and 2015. Bootstrap simulated standard errors are reported in parentice. Simulation is based on 100,000 iterations.

To provide a more intuitive visualization of how various types of shocks to dividend growth rates at a given time affect investors' expectations of dividends going forward, we consider an one unit change to  $\epsilon_{d,t}$ ,  $\epsilon_{x,t}$ , or  $\epsilon_{q,t}$  and show how such a change affects expected dividend growth rates both immediately and up to 10 years into the future, i.e. time- $t$  to time- $(t + 10)$ . We report these impulse response functions in Figure 1. We see that  $\epsilon_{d,t}$  affects dividend growth rates immediately but its effect does not persist into the

---

the Great Depression, i.e. 1930-2015.

future, whereas  $\epsilon_{x,t}$  and  $\epsilon_{q,t}$  affect dividend growth rates with a one-period lag but their effects are persistent over time.

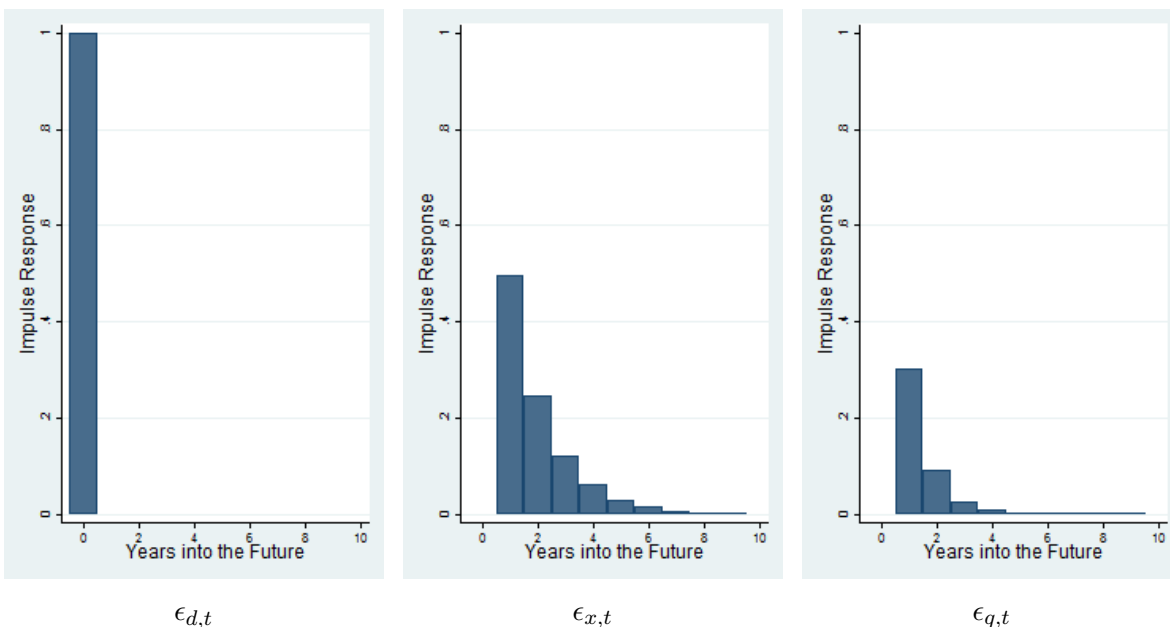


Figure 1: **Impulse Response Functions on Shocks that Affect Expected Dividend Growth Rates.** This figure plots the changes to expected dividend growth rates from immediately to over the next 10 years due to a unit change in shocks to dividend growth rates:  $\epsilon_{d,t}$ ,  $\epsilon_{x,t}$ , and  $\epsilon_{q,t}$ .

In Table 3, we report serial correlations, up to 5 years, for dividend growth rates and dividend growth rate residuals, which we define as the difference between dividend growth rates and our dividend model’s expected rates, along with Ljung and Box (1978)  $Q$ -statistics for testing if dividend growth rates and growth rate residuals are serially correlated. We also report serial correlations and  $Q$ -statistics for dividend growth rate residuals of one of the baseline dividend models described in (1), (3), and (10). We find that our model is reasonably successful at matching serial correlations in dividend growth rates for up to 5 years. That is, our model’s dividend growth rate residuals appear to be serially uncorrelated. In comparison, the baseline models all perform considerably worse along this dimension.

In the first column of Table 4, we report our dividend model’s performance in predicting dividend growth rates. Between 1946 and 2015, our model predicts 42.8 percent of the variation in annual dividend growth rates, which is a significant improvement over the baseline models. Given these statistics are in-sample, we know that at least a part of this improved forecasting performance comes from adding more parameters to existing models and is thus mechanical. Thus, to address the concern that our model overfits the data,

	$\Delta d_{t+1}$	J&L	$\Delta d_{t+1} - E_t[\Delta d_{t+1}]$		
			Baseline 1	Baseline 2	Baseline 3
<u>Serial Correlation (Years)</u>					
1	0.435	-0.029	0.122	0.088	0.343
2	-0.063	-0.122	-0.200	-0.208	-0.160
3	-0.274	-0.019	-0.208	-0.185	-0.284
4	-0.296	0.057	-0.132	-0.091	-0.236
5	-0.209	0.184	-0.063	-0.032	-0.170
 <i>Q</i> -Statistics	 34.72	 4.623	 10.38	 8.221	 26.76
	[0.000]	[0.464]	[0.065]	[0.144]	[0.000]

Table 3: **Serial Correlations in Dividend Growth Rates and Residuals:** This table reports the 1, 2, 3, 4, and 5 years serial correlations for annual nominal dividend growth rates and growth rate residuals of our dividend model (i.e. J&L), the dividend model in van Binsbergen and Koijen (2010) (i.e. Baseline 1), the dividend model in Campbell and Shiller (1988b) (i.e. Baseline 2), or a 3-state Markov-switching model (i.e. Baseline 3), based on data between 1946 and 2015. Also reported are the Ljung and Box (1978) *Q*-statistics for testing if dividend growth rates and growth rate residuals are serially correlated. *p*-values for *Q*-statistics are reported in square parentice.

we also assess our model based on how it forecasts dividend growth rates out-of-sample. That is, instead of estimating model parameters based on the full data sample, we predict dividend growth rates at each point in time using dividend model parameters esstimated based on data available at the time. Forecasting performance of a model  $\mathcal{M}_i$  is then evaluated using out-of-sample *R*-square value as defined in Goyal and Welch (2008):

$$\mathcal{R}^2 = 1 - \frac{\frac{1}{T-T_0+1} \sum_{t=T_0}^{T-1} (\Delta d_{t+1} - E_t[\Delta d_{t+1} | \mathcal{M}_i])^2}{\frac{1}{T-T_0+1} \sum_{t=T_0}^{T-1} (\Delta d_{t+1} - \hat{\mu}_{d,t})^2}, \quad (12)$$

where  $\hat{\mu}_{d,t}$  is the average of dividend growth rates up to time-*t*:

$$\hat{\mu}_{d,t} = \frac{1}{t} \sum_{s=0}^{t-1} \Delta d_{s+1}. \quad (13)$$

We use time-0 to denote the start of the data sample, time- $T_0$  to denote the end of the training period, and time- $T$  to denote the end of the data sample. Due to the relative complexity of our model, we use the first 30 years of our data sample as the training period so that out-of-sample prediction is for the period between 1976 and 2015. Throughout this paper, for predictive analysis, we assume investors have access to earnings information 3 months after fiscal quarter or year end. The choice of 3 months is based on Securities and

Exchange Commission (SEC) rules since 1934 that require public companies to file 10-Q reports no later than 45 days after fiscal quarter end and 10-K reports no later than 90 days after fiscal year end.<sup>9</sup> We assume that information about prices and dividends is known to investors in real time.<sup>10</sup> In the third and fourth columns of Table 4, we report out-of-sample  $R$ -square value for predicting dividend growth rates and the corresponding  $p$ -value from the adjusted-MSPE statistic of Clark and West (2007). Results show that our model predicts 33.4 percent of the variation in annual dividend growth rates between 1976 and 2015 out-of-sample, which is a significant improvement over the 20.8 percent, 26.1 percent, and -1.9 percent from the baseline models.

We proceed to show that the differences in forecasting performance between our model and the baseline models is significant. For two models  $\mathcal{M}_i$  and  $\mathcal{M}_j$ , we define incremental  $R$ -square value of  $\mathcal{M}_i$  over  $\mathcal{M}_j$  as:

$$\mathcal{R}_I^2(\mathcal{M}_i, \mathcal{M}_j) = 1 - \frac{\frac{1}{T-T_0+1} \sum_{t=T_0}^{T-1} (\Delta d_{t+1} - E_t[\Delta d_{t+1} | \mathcal{M}_i])^2}{\frac{1}{T-T_0+1} \sum_{t=T_0}^{T-1} (\Delta d_{t+1} - E_t[\Delta d_{t+1} | \mathcal{M}_j])^2}, \quad (14)$$

and report statistics in Table 4. If incremental  $R$ -square value is significantly positive, it suggests that our dividend model is an improvement over the baseline models in predicting annual dividend growth rates. Taken as a whole, we note that the difference in forecasting performance between our model and the baseline models is significant.

## 1.2 Inflation and Real Rates

In a standard neoclassical asset pricing model, real dividend growth rates, not nominal rates, are of interest to investors in forming their investment decisions. To convert nominal dividend growth rates into real rates, we need to specify a process for inflation. We model inflation as a stationary AR[1] process:<sup>11</sup>

$$\Delta\pi_{t+1} - \mu_\pi = \eta(\Delta\pi_t - \mu_\pi) + \sigma_\pi \epsilon_{\pi,t+1}, \quad \epsilon_{\pi,t+1} \sim \mathbb{N}(0, 1). \quad (15)$$

---

<sup>9</sup>In 2002, these rules were updated to require large firms file 10-Q reports no later than 40 days after fiscal quarter end and 10-K reports no later than 60 days after fiscal year end.

<sup>10</sup>To demonstrate the robustness of these assumptions, in the Online Appendix, we also replicate our main results imposing an additional 3 months lag in our estimation of dividend model parameters.

<sup>11</sup>In the Online Appendix, we explain why AR[1] is the most appropriate ARMA( $p, q$ ) model for inflation. Nevertheless, to show that our results are not driven by our inflation model or learning about the inflation process, we replicate our main results assuming inflation rates to be i.i.d. with a fixed mean and standard deviation.

	<u>In-Sample</u> $R^2$	<u>Out-of-Sample</u> $R^2$ $p$ -value	
J&L	0.428	0.334	0.000
Baseline 1	0.190	0.208	0.004
Baseline 2	0.217	0.261	0.001
Baseline 3	0.133	-0.019	1.000

	<u>Out-of-Sample</u> $\bar{R}_T^2$ $p$ -value	
J&L over Baseline 1	0.159	0.013
J&L over Baseline 2	0.100	0.053
J&L over Baseline 3	0.347	0.000

Table 4: **Dividend Growth Rates and Expected Growth Rates.** The top panel. The first column of this table reports in-sample  $R^2$  value for predicting annual nominal dividend growth rates using our dividend model (i.e. J&L), the dividend model in van Binsbergen and Koijen (2010) (i.e. Baseline 1), the dividend model in Campbell and Shiller (1988b) (i.e. Baseline 2), or a Markov-switching model (i.e. Baseline 3). The second and third columns report the out-of-sample  $R$ -square value for predicting annual dividend growth rates and the corresponding  $p$ -value from the adjusted-MSPE statistic of Clark and West (2007). The bottom panel. This table reports incremental  $R$ -square value of our dividend model over each one of the baseline models for predicting annual dividend growth rates. In-sample (out-of-sample) statistics are based on data between 1946 and 2015 (1976 and 2015).

Table 5 reports parameter estimates of the inflation model based on data between 1946 and 2015. We see a moderate level of persistence in inflation. Because  $\lambda_{d\pi}$ ,  $\lambda_{x\pi}$ , and  $\lambda_{q\pi}$  are not significantly different from zero, we force these correlation parameters to be zeros in our subsequent analysis for the purpose of parameters reduction. Given this inflation model, we can then derive the expression for expected real dividend growth rates based on expected nominal rates and inflation as:

$$E_t[\Delta\tilde{d}_{t+s+1}] = (\mu_d - \mu_\pi) + \rho^s x_t + \phi\theta^s(q_t - \mu_q) - \eta^{s+1}(\Delta\pi_t - \mu_\pi), \forall s \geq 0. \quad (16)$$

where  $\Delta\tilde{d}_t = \Delta d_t - \Delta\pi_t$  denotes real dividend growth rate.<sup>12</sup>

$\mu_\pi$	$\eta$	$\sigma_\pi$	$\lambda_{d\pi}$	$\lambda_{x\pi}$	$\lambda_{q\pi}$
0.035	0.568	0.026	0.265	-0.268	0.232
(0.013)	(0.109)	(0.020)	(0.142)	(0.146)	(0.137)

**Table 5: Inflation Model Parameter Estimates:** This table reports estimated parameters from our inflation model, based on data between 1946 and 2015. Bootstrap simulated standard errors are reported in parentice. Simulation is based on 100,000 iterations.

In Table 6, we report the in-sample and out-of-sample  $R$ -square values for predicting real, rather than nominal, dividend growth rates using either our model or one of the baseline models. We find that our model also outperforms the baseline models in forecasting real dividend growth rates. It predicts 37.3 percent of the variation in real annual dividend growth rates between 1946 and 2015 in-sample and 32.7 percent of the variation in real annual dividend growth rates between 1976 and 2015 out-of-sample.

## 2 Parameter Uncertainty and Learning

The difference between in-sample and out-of-sample prediction is the assumption made on investors' information set. Model parameters reported in Table 2 are estimated using data up to 2015, so they reflect investors' knowledge of dividend dynamics at the end of 2015. That is, if investors were to estimate our dividend model at an earlier date, they would have estimated a set of parameter values different from those reported in Table 2. This is a result of investors' knowledge of dividend dynamics evolving as more data

<sup>12</sup>Throughout this paper, we put  $\sim$  on top of a variable to denote that the variable is defined in real, not nominal, terms.



	<u>In-Sample</u> $R^2$	<u>Out-of-Sample</u> $R^2$ $p$ -value	
J&L	0.373	0.327	0.000
Baseline 1	0.228	0.220	0.003
Baseline 2	0.241	0.239	0.002
Baseline 3	0.188	-0.099	1.000

	<u>Out-of-Sample</u> $R^2$ $p$ -value	
J&L over Baseline 1	0.138	0.021
J&L over Baseline 2	0.115	0.037
J&L over Baseline 3	0.388	0.000

Table 6: **Dividend Growth Rates and Expected Growth Rates (Real Rates)**. The top panel. The first column of this table reports in-sample  $R^2$  value for predicting annual real dividend growth rates using our dividend model (i.e. J&L), the dividend model in van Binsbergen and Kojen (2010) (i.e. Baseline 1), the dividend model in Campbell and Shiller (1988b) (i.e. Baseline 2), or a Markov-switching model (i.e. Baseline 3). The second and third columns report the out-of-sample  $R$ -square value for predicting annual dividend growth rates and the corresponding  $p$ -value from the adjusted-MSPE statistic of Clark and West (2007). The bottom panel. This table reports incremental  $R$ -square value of our dividend model over each one of the baseline models for predicting annual dividend growth rates. In-sample (out-of-sample) statistics are based on data between 1946 and 2015 (1976 and 2015).

become available. We call this learning. That is, we use learning to refer to investors estimating model parameters at each point in time based on data available at the time. In this section, we summarize how learning affects investors' beliefs about the parameters governing the dividend process, assuming that investors behave as if they learn about dividend dynamics using our model. We then show that learning can have significant asset pricing implications.

We report, in Figure 2, dividend model parameters, estimated based on data up to time- $\tau$ , for  $\tau$  between 1976 and 2015. There are several points we take away from Figure 2. First, there is a gradual upward drift in investors' beliefs about the unconditional mean  $\mu_q$  of retention ratios. This suggests that firms have been paying a smaller fraction of earnings as cash dividends in recent decades. Second, there are gradual downward drifts in investors' beliefs about  $\phi$  that connects corporate payout policy to dividend dynamics. This means that dividends have become more smoothed over time. Third, a sharp drop in investors' beliefs about the persistence  $\theta$  of retention ratios towards the end of our data sample is due to the abnormally low earnings reported in late 2008 and early 2009 as a result of the financial crisis and the strong stock market recovery that followed.

Figure 2 shows that the persistence  $\rho$  of the latent variable  $x_t$ , as well as volatility  $\sigma_d$ ,  $\sigma_x$ , and correlation  $\lambda_{dq}$  of shocks to the dividend process, are the parameters hardest to learn and least stable over time. Investors' beliefs about  $\rho$ ,  $\sigma_d$ ,  $\sigma_x$ , and  $\lambda_{dq}$  fluctuate significantly over the sample period. For example, there are three times when investors' beliefs about  $\rho$  sharply drops. The first is at the start of Dot-Com bubble between 1995 and 1998. The second is during the crash of that bubble in late 2002 and early 2003. The third is during the financial crisis in late 2008 and early 2009. Further, there is also a long-run trend that sees a gradual decrease in investors' beliefs about  $\rho$  since early 1980s. That is, if we were to pick a random date between 1976 and 2015 and estimate our model based on data up to that date, on average we would have estimated a  $\rho$  of 0.734.<sup>13</sup> This would be significantly higher than the 0.496 reported in Table 2 estimated using the full data sample.

We can infer, from standard errors reported in Table 2, that learning about dividend dynamics is a slow process. That is, even with 70 years of data, there are still significant uncertainties surrounding the estimates of some model parameters. For example, the 95 percent confidence interval for  $\rho$  is between 0.183 and 0.807. To quantify the speed of

---

<sup>13</sup>To establish a point of reference, Bansal and Yaron (2004) calibrate annualized persistence of expected dividend growth rate to be  $0.979^{12} = 0.775$ .

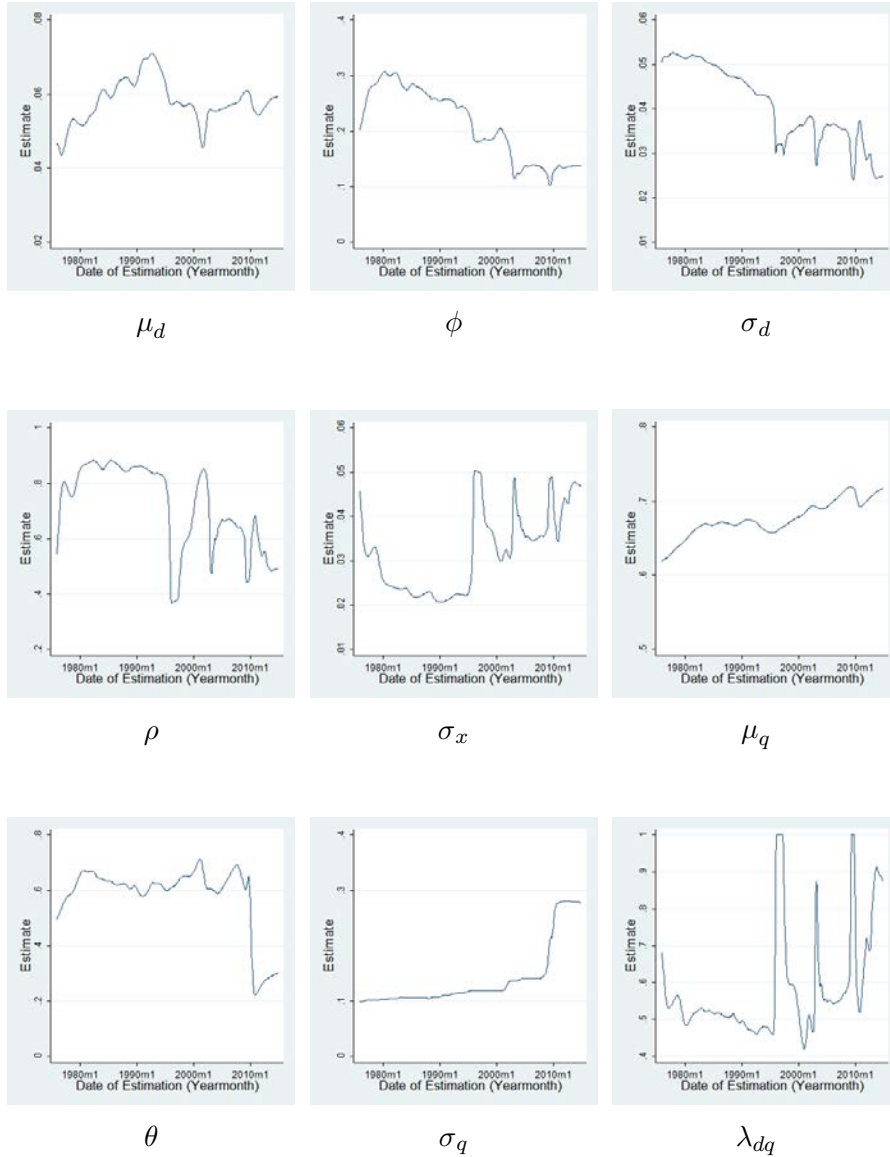


Figure 2: **Evolution of Dividend Model Parameter Estimates Over Time.** This figure plots estimates of the nine dividend model parameters, assuming that these parameters are estimated based on data up to time- $\tau$  for  $\tau$  between 1976 and 2015.

learning, following Johannes, Lochstoer, and Mou (2014), for a parameter in our dividend model, we construct a measure that is one minus the inverse ratio between the bootstrap simulated standard error assuming that the parameter is estimated based on data up to 2015 and the bootstrap simulated standard error assuming that the parameter is estimated based on 10 additional years of data (i.e. if the parameter were estimated in 2025). In other words, this ratio reports how much an estimated parameter’s standard error would have reduced if investors were to have 10 more years of data. So the closer this ratio is to zero, the more difficult it is for investors to learn about that parameter. In Table 7, we report this measure for each of the nine model parameters. Overall, 10 years of additional data would only decrease the standard errors of parameter estimates by between 3 and 8 percent. Further, consistent with results in Figure 2 and in Table 1, we find that it is more difficult to reduce uncertainties surrounding  $\rho$ ,  $\sigma_d$ ,  $\sigma_x$ , and  $\lambda_{dq}$  than any of the other parameters.

---



---

$\mu_d$	$\phi$	$\sigma_d$	$\rho$	$\sigma_x$	$\mu_q$	$\theta$	$\sigma_q$	$\lambda_{dq}$
0.076	0.076	0.051	0.049	0.032	0.073	0.082	0.079	0.050

---



---

Table 7: **Speed of Learning about Dividend Model Parameters:** This table reports the speed of learning for the nine dividend model parameters. Speed of learning is defined as one minus the inverse ratio between the bootstrap simulated standard error assuming that the parameter is estimated based on data between 1946 and 2015 and the bootstrap simulated standard error assuming that the parameter is estimated based on 10 additional years of data (i.e. if the parameter were estimated in 2025). Simulation is based on 100,000 iterations.

## 2.1 Parameter Uncertainty and Investors’ Expectations for the Long-Run

We show that learning about dividend dynamics can have significant asset pricing implications. That is, consider the log linearized present value relationship in Campbell and Shiller (1988):

$$p_t - d_t = \frac{\kappa_0}{1 - \kappa_1} + \sum_{s=0}^{\infty} \kappa_1^s (E_t[\Delta d_{t+s+1}] - E_t[R_{t+s+1}]), \quad (17)$$

where  $\kappa_0$  and  $\kappa_1$  are log-linearizing constants and  $R_{t+1}$  is the stock index’s log return.<sup>14</sup> The expression is a mathematical identity that connects price-to-dividend ratios, expected

---

<sup>14</sup>To solve for  $\kappa_0 = \log(1 + \exp(\overline{p-d})) - \kappa_1(\overline{p-d})$  and  $\kappa_1 = \frac{\exp(\overline{p-d})}{1 + \exp(\overline{p-d})}$ , we set unconditional mean of log price-to-dividend ratios  $\overline{p-d}$  to 3.46. This gives  $\kappa_0 = 0.059$  and  $\kappa_1 = 0.970$ .

dividend growth rates, and discount rates, i.e. expected returns. We define stock yields as discount rates that equate the present value of expected future dividends to the current price of the stock index. That is, rearranging (17), we can write stock yield as:

$$\begin{aligned} sy_t &\equiv (1 - \kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta R_{t+s+1}] \\ &= \kappa_0 - (1 - \kappa_1)(p_t - d_t) + (1 - \kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta d_{t+s+1}]. \end{aligned} \quad (18)$$

We define long-run dividend growth expectations as:

$$\partial_t \equiv (1 - \kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta d_{t+s+1}]. \quad (19)$$

Given that price-to-dividend ratios are observed, there is a one-to-one mapping between long-run dividend growth expectations and stock yields. Further, long-run dividend growth expectations are specific to the dividend model and its parameters. For example, using our dividend model, we can re-write expected long-run dividend growth rates as:

$$\partial_t = (1 - \kappa_1) \sum_{s=0}^{\infty} \kappa_1^s (\mu_d + \rho^s x_t + \phi \theta^s (q_t - \mu_q)). \quad (20)$$

If investors instead use a different dividend model, their expectations of long-run dividend growth rates will also be different. For example, if we assume that dividend growth rates follow a white noise process centered around  $\mu_d$ , we can re-write (19) instead as  $\partial_t = \mu_d$ . Further, because long-run dividend growth expectations are functions of dividend model parameters, it is also affected by whether these parameters are estimated once based on the full data sample, or estimated at each point in time based on data available at the time. The first case corresponds to investors having complete knowledge of the parameters describing the dividend process. The second case corresponds to investors having to learn about dividend dynamics. In Figure 3, we plot our model's long-run dividend growth expectations, either assuming learning or assuming full information. The plot shows that learning can have a considerable effect on investors' long-run dividend growth expectations.

In Figure 4, we plot stock yields, either assuming learning or assuming full information, computed by substituting (20) into (18):

$$sy_t = \kappa_0 - (1 - \kappa_1)(p_t - d_t) + \mu_d + (1 - \kappa_1) \left( \frac{1}{1 - \kappa_1 \rho} x_t + \frac{\phi}{1 - \kappa_1 \theta} (q_t - \mu_q) \right). \quad (21)$$

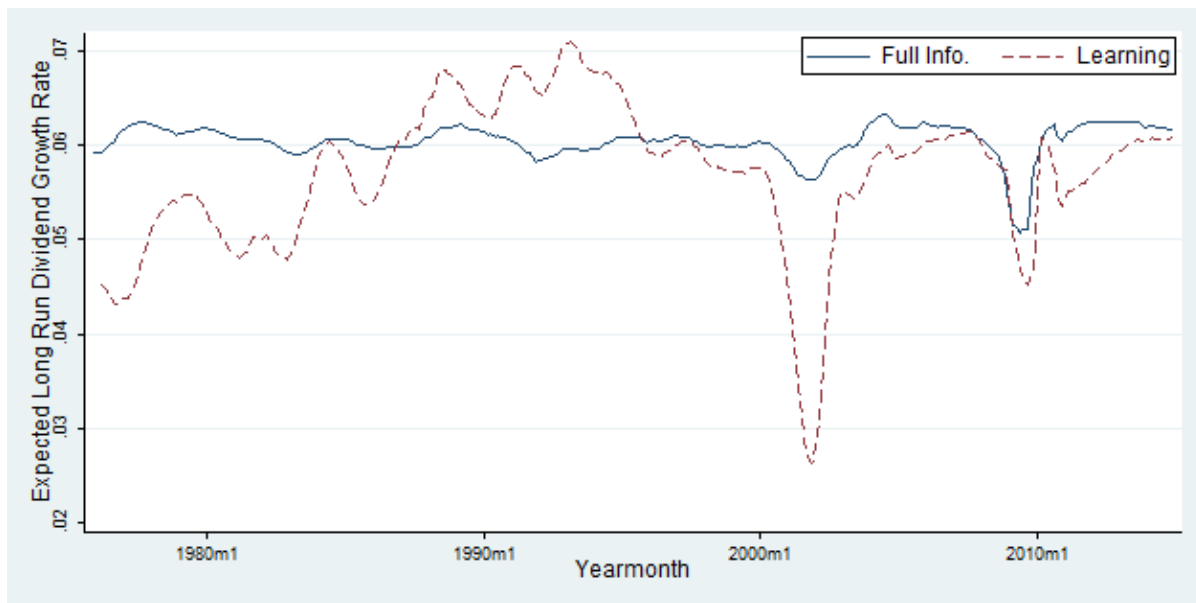


Figure 3: **Expected Long-Run Dividend Growth Rates.** This figure plots long-run dividend growth expectations, computed using our dividend model, for the period between 1976 and 2015. Dividend model parameters are estimated based on data since 1946. Assuming full information, model parameters are estimated once based on the full data sample. Assuming learning, those parameters are estimated at each point in time based on data available at the time.

We also plot price-to-dividend ratios in Figure 4, and scale price-to-dividend ratios to allow for easy comparison to stock yields. We note that, assuming full information, there is almost no noticeable difference between the time series of price-to-dividend ratios and stock yields. This suggests that the variation in long-run dividend growth expectations, assuming that investors do not learn, is minimal relative to the variation in price-to-dividend ratios, so the latter dominates the variation in stock yields, as stock yields are a linear combination of these two components. However, assuming learning, we find significant differences between the time series of price-to-dividend ratios and stock yields.

### 3 Learning about Dividends and the Time Variation in Discount Rates

Results in the previous section show that parameters in our dividend model can be difficult to estimate with precision in finite sample. As a result, we argue that learning about model parameters can have significant asset pricing implications. This claim is based on

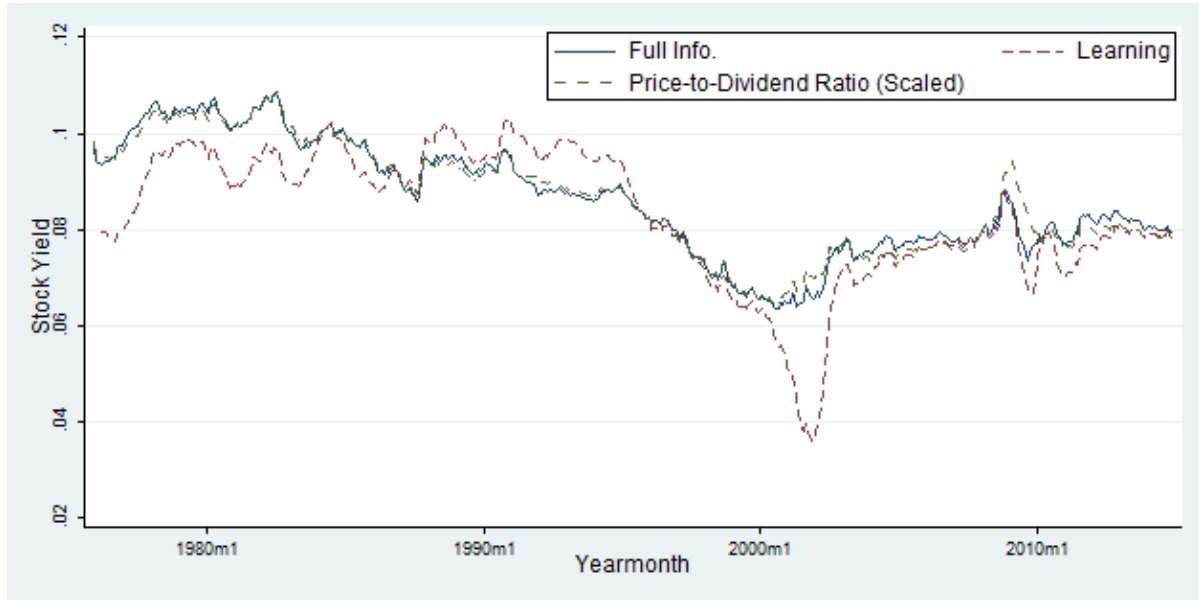


Figure 4: **Stock Yields.** This figure plots stock yields  $sy_t$ , computed using our dividend model, and log price-to-dividend ratios (scaled) for the period between 1976 and 2015. Dividend model parameters are estimated based on data since 1946. Assuming full information, model parameters are estimated once based on the full data sample. Assuming learning, those parameters are estimated at each point in time based on data available at the time.

the assumption that our model captures investors' expectations about future dividends. That is, we assume that investors behave as if they learn about dividend dynamics using our model, not a model that is very different from ours. In this section, we present evidence that supports this assumption.

First, we show that stock yields, assuming learning, predict annual stock index returns.<sup>15</sup> To establish a baseline, note that, if we assume dividend growth rates follow a white noise process centered around  $\mu_d$ , stock yields can be simplified to:

$$sy_t = \kappa_0 - (1 - \kappa_1)(p_t - d_t) + \mu_d. \quad (22)$$

That is, under the white noise assumption, stock yields are just scaled price-to-dividend ratios. So, we regress stock index returns over the next year on price-to-dividend ratios, based on data between 1976 and 2015. We report regression statistics in the first column of Table 8. Standard errors reported are Newey and West (1987) adjusted.<sup>16</sup> Results from

<sup>15</sup>Ideally, we should show that stock yields, assuming learning, better capture the variation in long-run stock index returns. However, due to our limited data sample, we do not have sufficient non-overlapping data points of long-run returns to carry out this analysis. For this reason, we instead rely on annual returns.

<sup>16</sup>Stambaugh (1999) shows that, when variables are highly serially correlated, OLS estimators' finite-

Table 8 show that, between 1976 and 2015, price-to-dividend ratios predict 9.9 percent of the variation in annual stock index returns.

We then regress stock index returns over the next year on stock yields in (21), assuming learning. We report regression statistics in the second column of Table 8. We see that  $R$ -square value from this regression is 14.9 percent. We note that the only difference between this regression and the baseline regression is the assumption on dividend dynamics. That is, here we assume that investors behave as if they learn about dividend dynamics using our model, whereas in the baseline regression we assume that expected dividend growth rates are constant. This means that we can attribute the increase in  $R$ -square value from 9.9 percent to 14.9 percent to our modeling of learning about dividend dynamics. To emphasize the importance of learning, we regress stock index returns over the next year on stock yields in (21), assuming full information. We report regression statistics in the fourth column of Table 8. Results show that stock yields, assuming full information, perform roughly as well as price-to-dividend ratios in predicting annual stock index returns. This is consistent with results in Figure 4, which show that there is almost no noticeable difference between the time series of price-to-dividend ratios and stock yields, assuming full information. To show that the differences in the  $R$ -square values in Table 8 are significant, we run bi-variate regressions of stock index returns over the next year on both stock yields, assuming learning, and either price-to-dividend ratios or stock yields, assuming full information. Statistics are in the fourth and fifth columns of Table 8. Results show that stock yields, assuming learning, strictly dominate both price-to-dividend ratios and stock yields, assuming full information, in predicting annual stock index returns.

It is worth noting that, for learning to be relevant in our context, investors must behave as if they are learning about dividend dynamics using our model. To illustrate this point, we regress stock index returns over the next year on stock yields, assuming instead that investors behave as if they learn about dividend dynamics using one of the three baseline models. Statistics are in the sixth to eighth columns of Table 8. We find that stock yields, assuming learning based on one of the baseline models, perform no better than price-to-dividend ratios in predicting annual stock index returns.

We can decompose stock index returns into the risk free rate component and the risk premium component. To investigate whether the gap in forecasting performance is for predicting risk free rates or risk premium, in the last row of Table 8, we report  $R$ -square sample properties can deviate from the standard regression setting.



		J&L			Baseline			
					1	2	3	
$p_t - d_t$	-0.114** (0.046)			0.017 (0.075)				
$sy_t(\mathcal{L})$		3.955*** (1.101)		4.368** (1.756)	4.554** (1.871)	2.971** (1.254)	3.205** (1.254)	2.026** (1.030)
$sy_t(\mathcal{F})$			3.742** (1.485)		-0.778 (2.533)			
$\underline{R^2}$								
Return	0.099	0.149	0.103	0.149	0.150	0.086	0.101	0.054
Exc. Return	0.092	0.142	0.096	0.142	0.143	0.079	0.094	0.046

Table 8: **Stock Index Returns and Stock Yields:** This table reports the coefficient estimates and  $R$ -square value from regressing stock index returns over the next year on log price-to-dividend ratios and stock yields, computed using our dividend model (i.e. J&L), the dividend model in van Binsbergen and Koijen (2010) (i.e. Baseline 1), the dividend model in Campbell and Shiller (1988b) (i.e. Baseline 2), or a Markov-switching model (i.e. Baseline 3), and assuming investors either learn (i.e.  $\mathcal{L}$ ), or do not learn (i.e.  $\mathcal{F}$ ), about model parameters. Regression is based on data between 1976 and 2015. Dividend model parameters are estimated based on data since 1946. Newey and West (1987) standard errors are reported in parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using \*, \*\*, and \*\*\*.

value for predicting stock index excess returns.<sup>17</sup> Results show that the differences in forecasting performance is mostly for predicting risk premium.

### 3.1 Learning about Dividend Dynamics and the Term-Structure of Discount Rates

Although stock yields are, by definition, optimal forecasts of long-run stock index returns, they do not reveal information about the term-structure of discount rates, i.e. the difference in how investors discount near-term versus distant-term cash flows. Further, if learning about dividend dynamics also affects the term structure of discount rates aside from its effect on long-run discount rates, the effect of learning on the variation in annual

<sup>17</sup>Let  $\hat{R}_t$  be stock index return forecast and  $R_{f,t}$  be the risk free rate. The in-sample  $R$ -square value for predicting stock index returns is  $\frac{\text{var}(R_{t+1} - \hat{R}_{t+1})}{\text{var}(R_{t+1})}$ , where  $\text{var}(\cdot)$  is the sample variance. The in-sample  $R$ -square value for predicting stock index excess returns is  $\frac{\text{var}((R_{t+1} - R_{f,t+1}) - (\hat{R}_{t+1} - R_{f,t+1}))}{\text{var}(R_{t+1} - R_{f,t+1})}$ .

stock index returns cannot be captured by its effect on stock yields alone. To see this, we run bi-variate regressions of stock index returns over the next year on stock yields in (21), assuming learning, and investors beliefs about one of the nine dividend model parameters, estimated at each point of time based on data available at the time. For example, for the parameter  $\rho$ , let  $\hat{\rho}_t$  denote the estimate of  $\rho$  based on data up to time- $t$ , we run the bi-variate regression:

$$R_{t+1} = \alpha + \beta \cdot sy_t(\mathcal{L}) + \gamma \cdot \hat{\rho}_t + \epsilon_{t+1} \quad (23)$$

where  $\mathcal{L}$  stands for our learning. Estimated coefficients from (23) are reported in Table 9. The regression examines whether the effect of learning on the variation in annual stock index returns can be fully accounted for by the effect of learning on stock yields. If learning affects the variation in annual stock index returns only through its effect on long-run discount rates, we should see that the  $\gamma$  estimates are not significantly different from zero. Instead, results show that learning about four of the nine model parameters, namely persistence  $\rho$  of the latent variable  $x_t$  and volatility  $\sigma_d$ ,  $\sigma_x$ , and correlation  $\lambda_{dq}$  of shocks to the dividend process, significantly influences the variation in annual stock index returns, even after controlling for the effect of learning on stock yields. Including investors' beliefs about one of these four parameters as the additional regressant increases  $R$ -square value for predicting the variation in annual stock index returns from 14.9 percent to between 21.6 percent and 25.6 percent.<sup>18</sup> This confirms that, aside from its effect on long-run discount rates, learning about dividend dynamics also affects the term structure of discount rates in a way that is not captured by its effect on stock yields alone.

## 4 Learning about Dividends in a Dynamic Equilibrium Model

Our results thus far show that learning about dividend dynamics affects both long-run discount rates and the term structure of discount rates. This motivates a model of stock index returns that also allow for learning. In this section, we search for a dynamic equilibrium asset pricing model that is able to quantitatively capture the role of learning in determining asset prices and expected returns. That is, a model that, after incorporating learning about dividend dynamics, is able to show strong performance in predicting annual stock index returns that is consistent with the data.

---

<sup>18</sup>Interestingly, we recall that, based on earlier results in Table 2, Table 7, and Figure 2, these four parameters are also the hardest for investors to learn over time.

	$\mu_d$	$\phi$	$\sigma_d$
$\beta$	4.032*** (1.158)	5.606*** (1.270)	6.165*** (0.884)
$\gamma$	-0.446 (2.499)	-0.601 (0.449)	-6.320*** (1.981)
<hr/>			
$\underline{R^2}$			
Return	0.149	0.179	0.216
Exc. Return	0.142	0.173	0.210
<hr/>			
	$\rho$	$\sigma_x$	$\mu_q$
$\beta$	5.700** (0.688)	6.142** (0.728)	3.971*** (1.290)
$\gamma$	-0.401*** (0.082)	6.773*** (1.463)	0.022 (1.107)
<hr/>			
$\underline{R^2}$			
Return	0.256	0.250	0.149
Exc. Return	0.250	0.244	0.142
<hr/>			
	$\theta$	$\sigma_q$	$\lambda_{dq}$
$\beta$	4.058*** (0.981)	4.414*** (1.078)	4.731*** (0.766)
$\gamma$	-0.192 (0.117)	0.413* (0.215)	0.370*** (0.077)
<hr/>			
$\underline{R^2}$			
Return	0.169	0.167	0.249
Exc. Return	0.163	0.160	0.243

**Table 9: Stock Index Returns, Stock Yields, and Investors' Beliefs about Dividend Model Parameters:** This table reports the coefficient estimates and  $R$ -square value from regressing stock index returns over the next year on stock yields, assuming learning, and investors beliefs about one of our dividend model parameters, estimated using our dividend model at each point in time based on only data available at the time. Regression is based on data between 1976 and 2015. Dividend model parameters are estimated based on data since 1946. Newey and West (1987) standard errors are reported in parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using \*, \*\*, and \*\*\*.

For the rest of this section, we first provide the theoretical foundation that how well stock index returns can be predicted using an asset pricing model can be used to assess that model. Then, we incorporate learning about dividend dynamics into a long-run risks model and show that 22.4 percent of the variation in annual stock index returns can be predicted using such a model.

## 4.1 Return Predictability and Assessing Asset Pricing Models

The criterion we propose to assess an asset pricing model is the deviation of that candidate model's expected returns from the expected returns of the true model. The true model here is defined as the asset pricing model that best describes the behavior of the marginal investor who prices that asset, here the stock index, in a frictionless and efficient market. Let  $\mathcal{M}_i$  be a candidate model,  $\mathcal{M}_0$  be the unobserved true asset pricing model,  $R_t$  be log return of the stock index,  $E_t[R_{t+1}|\mathcal{M}_i]$  be the  $\mathcal{M}_i$ -endowed-investors' expectation of stock index returns over the next year, and  $E_t[R_{t+1}|\mathcal{M}_0]$  be expected return under the true model. The following definition defines a better asset pricing model, i.e. the candidate model that is closer to the true model, as the model that minimizes the mean squared difference between its expected returns and the expected returns of the true model.

**Definition 1** *A candidate asset pricing model  $\mathcal{M}_i$  is a better approximation of the true asset pricing model ( $\mathcal{M}_0$ ) than model  $\mathcal{M}_j$  if and only if:*

$$E \left[ (E_t[R_{t+1}|\mathcal{M}_0] - E_t[R_{t+1}|\mathcal{M}_i])^2 \right] < E \left[ (E_t[R_{t+1}|\mathcal{M}_0] - E_t[R_{t+1}|\mathcal{M}_j])^2 \right] .$$

A clear inconvenience of this definition is that the true asset pricing model  $\mathcal{M}_0$  is never observable, and thus  $E_t[R_{t+1}|\mathcal{M}_0]$  is unobservable. To circumvent this issue, we notice that, assuming markets are frictionless and efficient and investors form rational expectations, the error term  $\epsilon_{t+1} = R_{t+1} - E_t[R_{t+1}|\mathcal{M}_0]$  is orthogonal to any information that is time- $t$  measurable. This leads to the following proposition.

**Proposition 1** *A candidate asset pricing model  $\mathcal{M}_i$  is a better approximation of the true asset pricing model ( $\mathcal{M}_0$ ) than model  $\mathcal{M}_j$  if and only if:*

$$1 - \frac{E \left[ (R_{t+1} - E_t[R_{t+1}|\mathcal{M}_i])^2 \right]}{E \left[ (R_{t+1} - E[R_{t+1}])^2 \right]} > 1 - \frac{E \left[ (R_{t+1} - E_t[R_{t+1}|\mathcal{M}_j])^2 \right]}{E \left[ (R_{t+1} - E[R_{t+1}])^2 \right]}$$

Proofs are in Appendix A.3. In other words, if we define out-of-sample  $R$ -square value:

$$\mathcal{R}^2(\mathcal{M}_i) = 1 - \frac{\frac{1}{T-T_0+1} \sum_{t=T_0}^{T-1} (R_{t+1} - E_t[R_{t+1}|\mathcal{M}_i])^2}{\frac{1}{T-T_0+1} \sum_{t=T_0}^{T-1} (R_{t+1} - \hat{\mu}_{r,t})^2}, \quad (24)$$

where  $\hat{\mu}_{r,t} = \frac{1}{t} \sum_{s=0}^{t-1} R_{s+1}$  is the average of stock index returns up to time- $t$ , as the performance of a candidate model  $\mathcal{M}_i$  in predicting annual stock index returns, and assuming we have a sufficiently long data sample, then we can use it to assess how close the candidate model is to the true model.

## 4.2 Return Predictability and Learning about Long-Run Risks

We propose a long-run risks model that combines our dividend model, Epstein and Zin (1989) investor preferences, and persistent consumption growth rates as in Bansal and Yaron (2004) and show that such a model predicts 22.4 percent of the variation in annual stock index returns.

Epstein and Zin (1989) has been one of the most widely used expressions for investor preferences in the literature. Investor preferences are defined recursively as:

$$U_t = \left[ (1 - \delta) \tilde{C}_t^{\frac{1-\alpha}{\zeta}} + \delta (E_t [U_{t+1}^{1-\alpha}])^{\frac{1}{\zeta}} \right]^{\frac{\zeta}{1-\alpha}}, \quad \zeta = \frac{1 - \alpha}{1 - \frac{1}{\psi}}, \quad (25)$$

where  $\tilde{C}_t$  is real consumption,  $\psi$  is the elasticity of intertemporal substitution (EIS), and  $\alpha$  is the coefficient of risk aversion. We note that, the representative agent prefers early resolution of uncertainty if  $\zeta < 0$  and prefers late resolution of uncertainty if  $\zeta > 0$ .<sup>19</sup> Log of the intertemporal marginal rate of substitution (IMRS) is:

$$m_{t+1} = \zeta \log(\delta) - \frac{\zeta}{\psi} \Delta \tilde{c}_{t+1} + (\zeta - 1) \tilde{R}_{t+1}^c, \quad (26)$$

where  $\tilde{c} = \log(\tilde{C})$  and  $\tilde{R}_{t+1}^c$  denotes the real return of the representative agent's wealth portfolio. Following Bansal and Yaron (2004), we assume that consumption and dividend growth rates carry the same persistent latent component  $x_t$  and allow volatility in consumption growth rates to be time varying.<sup>20</sup> That is, we describe real consumption

<sup>19</sup>Or equivalently, if  $\alpha > 1$ , then the representative agent prefers early resolution of uncertainty if  $\psi > 1$  and prefers late resolution of uncertainty if  $\psi < 1$ .

<sup>20</sup>A deviation from Bansal and Yaron (2004) is that the latent variable  $x_t$  is assumed to be homoskedastic in our model.

growth rates using the following system of equations:

$$\begin{aligned}\Delta\tilde{c}_{t+1} - \mu_c &= \frac{1}{\gamma}x_t + \sigma_t\epsilon_{c,t+1} \\ \sigma_{t+1}^2 - \sigma_c^2 &= \omega(\sigma_t^2 - \sigma_c^2) + \sigma_\varsigma\epsilon_{\varsigma,t+1}.\end{aligned}\tag{27}$$

The correlation matrix for shocks to consumption, dividends, and retention ratios can be written as:

$$\begin{pmatrix} \epsilon_{c,t+1} \\ \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{\varsigma,t+1} \\ \epsilon_{q,t+1} \\ \epsilon_{\pi,t+1} \end{pmatrix} \sim \text{i.i.d. } \mathbb{N} \left( \mathbf{0}, \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \lambda_{dq} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \lambda_{dq} & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \right).\tag{28}$$

Because we do not use actual consumption data in this paper, the correlations that involve shocks  $\epsilon_{c,t}$  or  $\epsilon_{\varsigma,t}$  to the consumption process cannot be identified. So, for convenience, we set them to zeros. The remaining  $\lambda_{dq}$  is estimated as a part of the dividend process.

We note that the unconditional mean of consumption growth rates must equal to that of dividend growth rates, i.e.  $\mu_c = \mu_d - \mu_\pi$ , or dividend as a fraction of consumption will either become negligible or explode. For the remaining parameters in (27), apart from those concerning the latent variable  $x_t$ , we adopt the calibration of Bansal and Yaron (2004), converted to quarterly frequency. They are:  $\sigma_c = 0.0078 \cdot 3 = 0.0234$ ,  $\omega = 0.987^3 = 0.962$ ,  $\sigma_\varsigma = 0.000023 \cdot 3^2 = 0.00021$ , and  $\gamma = 3$ . The persistence of the latent variable  $x_t$  is set to  $\rho = 0.979^3 = 0.938$  in Bansal and Yaron (2004). A common criticism of the long-run risk model has always been that it requires a small but highly persistent component in consumption and dividend growth rates that is difficult to find support in the data.<sup>21</sup> This criticism serves as the rationale for why we expect learning to be important. For investor preferences, we choose  $\psi = 1.5$  to be consistent with preferences for the early resolution of uncertainty, and choose  $\alpha = 4$  and  $\delta = 0.995$  so that the unconditional means of risk free rates and risk premium from our model are roughly consistent with data between 1946 and 2015.

We solve our long-run risk model in Appendix A.4. In solving this model, we closely follow the steps in Bansal and Yaron (2004). The model consists of four state variables: latent variables  $x_t$  and  $\sigma_t^2$ , retention ratios, and inflation rates. We can solve for price-to-

---

<sup>21</sup>See Beeler and Campbell (2012), Jagannathan and Marakani (2015).

dividend ratio as a linear function of these four state variables:

$$p_t - d_t = A_{d,0} + A_{d,1}x_t + A_{d,2}\sigma_t^2 + A_{d,3}(q_t - \mu_q) + A_{d,4}(\Delta\pi_t - \mu_\pi). \quad (29)$$

We can solve for expected returns over the next year as:

$$E_t[R_{t+1}] = A_{r,0} + A_{r,1}x_t + A_{r,2}\sigma_t^2 + A_{r,4}(\Delta\pi_t - \mu_\pi), \quad (30)$$

where coefficients  $A_{d,\cdot}$  and  $A_{r,\cdot}$ , derived in Appendix A.4, are functions of the parameters that describe investor preferences, consumption dynamics, and dividend dynamics. We note that, substituting (29) into (30), we can avoid estimating the latent variable  $\sigma_t^2$  directly and instead write expected returns over the next year as a function state variables that can be estimated from dividend dynamics and price-to-dividend ratios:

$$E_t[R_{t+1}] = A_0 + A_1x_t + A_2(p_t - d_t) + A_3(q_t - \mu_q) + A_4(\Delta\pi_t - \mu_\pi),$$

$$A_0 = \frac{A_{r,0}A_{d,2} - A_{r,2}A_{d,0}}{A_{d,2}}, \quad A_1 = \frac{A_{r,1}A_{d,2} - A_{r,2}A_{d,1}}{A_{d,2}} \quad (31)$$

$$A_2 = \frac{A_{r,2}}{A_{d,2}}, \quad A_3 = -\frac{A_{r,2}A_{d,3}}{A_{d,2}}, \quad A_4 = \frac{A_{r,4}A_{d,2} - A_{r,2}A_{d,4}}{A_{d,2}}. \quad (32)$$

We examine how our long-run risks model, assuming learning, i.e. our learning model, performs in predicting stock index returns.<sup>22</sup> Here, learning refers to estimating dividend model parameters at each point in time based on data available at the time and substituting these parameters into (31) to compute expected returns. We measure forecasting performance using (quasi) out-of-sample  $R$ -square value defined as:<sup>23</sup>

$$\mathcal{R}^2(\mathcal{L}) = 1 - \frac{\frac{1}{T-T_0+1} \sum_{t=T_0}^{T-1} (R_{t+1} - E_t[R_{t+1}|\mathcal{L}])^2}{\frac{1}{T-T_0+1} \sum_{t=T_0}^{T-1} (R_{t+1} - \hat{\mu}_{r,t})^2}. \quad (33)$$

We use the first 30 years of the data sample as the training period and compute the out-of-sample  $R$ -square value using data between 1976 and 2015. In the first row of Table 12, we report out-of-sample  $R$ -square value for predicting annual stock index returns using our learning model. We find that, between 1976 and 2015, our long-run risks model predicts 22.4 percent of the variation in annual stock index returns. This level

<sup>22</sup>See Figure 7 for a plot of expected returns against returns realized over the next year between 1976 and 2015.

<sup>23</sup>The term quasi refers to the fact that some of parameters in our long-run risks model cannot be estimated from data and are calibrated instead.

of forecasting performance significantly exceeds what is commonly documented in the existing literature.

To isolate the incremental contribution of learning to the model’s performance in predicting annual stock index returns, we compute expected returns in (31) using dividend model parameters estimated based on the full data sample, i.e. our full information model. We report out-of-sample  $R$ -square value for predicting stock index returns using our full information model in the second row of Table 10. We see that, assuming full information,  $R$ -square value reduces from 22.4 percent to 6.4 percent. This means that learning accounts for more than half of the forecasting performance. To examine the significance of this difference, we report, in the third row of Table 10, incremental  $R$ -square value of our learning model over full information model:

$$\mathcal{R}_I^2(\mathcal{L}, \mathcal{F}) = 1 - \frac{\frac{1}{T-T_0+1} \sum_{t=T_0}^{T-1} (R_{t+1} - E_t[R_{t+1}|\mathcal{L}])^2}{\frac{1}{T-T_0+1} \sum_{t=T_0}^{T-1} (R_{t+1} - E_t[R_{t+1}|\mathcal{F}])^2}. \quad (34)$$

Also, we note that  $\mathcal{R}_I^2(\mathcal{L}, \mathcal{F})$ ,  $\mathcal{R}^2(\mathcal{L})$  and out-of-sample  $R$ -square value  $\mathcal{R}^2(\mathcal{F})$  of our full information model are related through the following equality:

$$\mathcal{R}_I^2(\mathcal{L}, \mathcal{F}) = 1 - \frac{1 - \mathcal{R}^2(\mathcal{L})}{1 - \mathcal{R}^2(\mathcal{F})}. \quad (35)$$

Results from Table 10 show that the incremental gain in forecasting performance that is attributable to modeling investors’ learning about dividend dynamics is significant.

	$R^2$	$p$ -value
Learning Model	0.224	0.002
Full Information Model	0.064	0.118
Incremental	0.171	0.009

Table 10: **Stock Index Returns and the Long-Run Risks Model.** This table reports out-of-sample  $R$ -square value for predicting annual stock index returns over the next year using our long run risks model, assuming investors either learn, or do not learn, about dividend model parameters, and the corresponding  $p$ -value from the adjusted-MSPE statistic of Clark and West (2007). Also reported is incremental out-of-sample  $R$ -square value. Dividends are estimated using data since 1946. Statistics are based on data between 1976 and 2015.



### 4.2.1 Forecasting Performance of Long-Run Risks Model over Time

For additional details on how our long-run risk model’s forecasting performance evolves over time, we follow Goyal and Welch (2008) and define the cumulative sum of squared errors difference (SSED) between predicting annual stock index return using our learning model and using the historical mean of returns as:

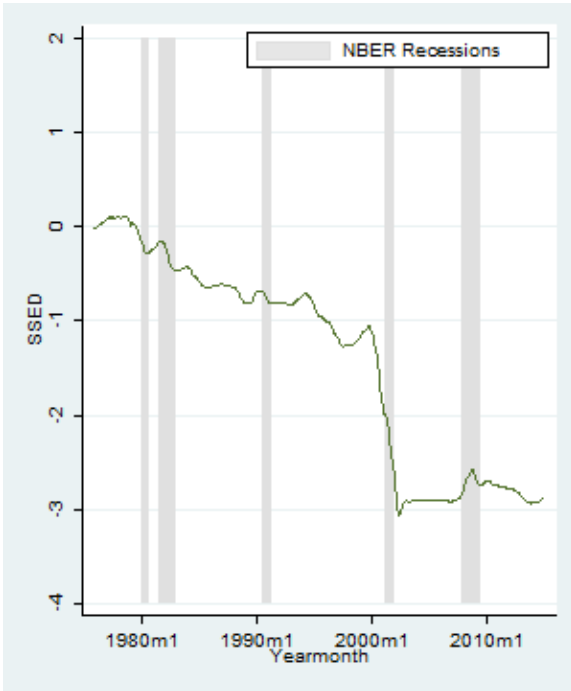
$$\mathcal{D}_t(\mathcal{L}) = \sum_{s=T_0}^{t-1} (R_{s+1} - E_t[R_{s+1}|\mathcal{L}])^2 - \sum_{s=T_0}^{t-1} (R_{s+1} - \hat{\mu}_{r,t})^2. \quad (36)$$

The SSED is plotted on the left side of Figure 5. If the forecasting performance of our learning model is stable and robust over time, we should observe a steady but constant decline in SSED. Instead, if the forecasting performance is especially poor in certain sub-period of the data, we should see a significant drawback in SSED during that sub-period. We note that our model’s forecasting performance is positive through the majority of the data sample. Overall, about one third of the forecasting performance is realized during the first two decades of the data sample, the remaining two-thirds during the Dot-Com crash. In contrast, performance has been relatively flat over the last decade. As a baseline, on the right of Figure 5, we report the SSED between predicting annual stock index returns using our full information model  $\mathcal{F}$  and using the historical mean of returns. We see that, assuming full information, there are clearly periods during which the model performs poorly, such as during the Dot-Com boom in the late 1990s.

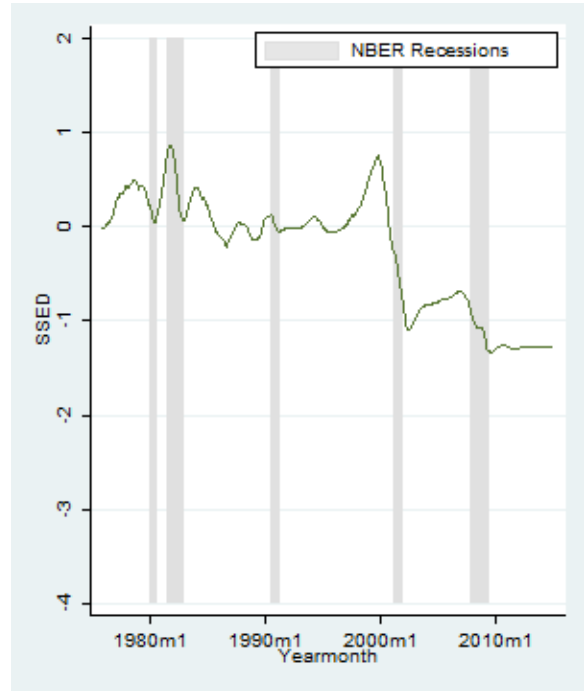
To isolate the incremental contribution of learning to SSED, we plot, in Figure 6, the incremental SSED defined as the difference in SED between our learning model and our full information model:

$$\mathcal{D}_t(\mathcal{L}) - \mathcal{D}_t(\mathcal{F}) = \sum_{s=T_0}^{t-1} (R_{s+1} - E_t[R_{s+1}|\mathcal{L}])^2 - \sum_{s=T_0}^{t-1} (R_{s+1} - E_t[R_{s+1}|\mathcal{F}])^2. \quad (37)$$

We note that the incremental gain in forecasting performance from learning is positive and consistent throughout most of the data sample, except during the 2008-2009 financial crisis. We conjecture one possible explanation for this finding is that the financial crisis represent a rare regime change. As a consequence, investors no longer learn from much of the historical dividend data, realized under a different regime, as they become irrelevant under the new regime, leading to our learning model becoming misspecified during that sub-period.



Learning



Full Info.

Figure 5: **Cumulative Sum of Squared Errors Difference.** This figure plots the cumulative sum of squared errors difference, assuming investors either learn, or do not learn, about dividend model parameters, for the period between 1976 and 2015. Dividend model parameters are estimated based on data since 1946.

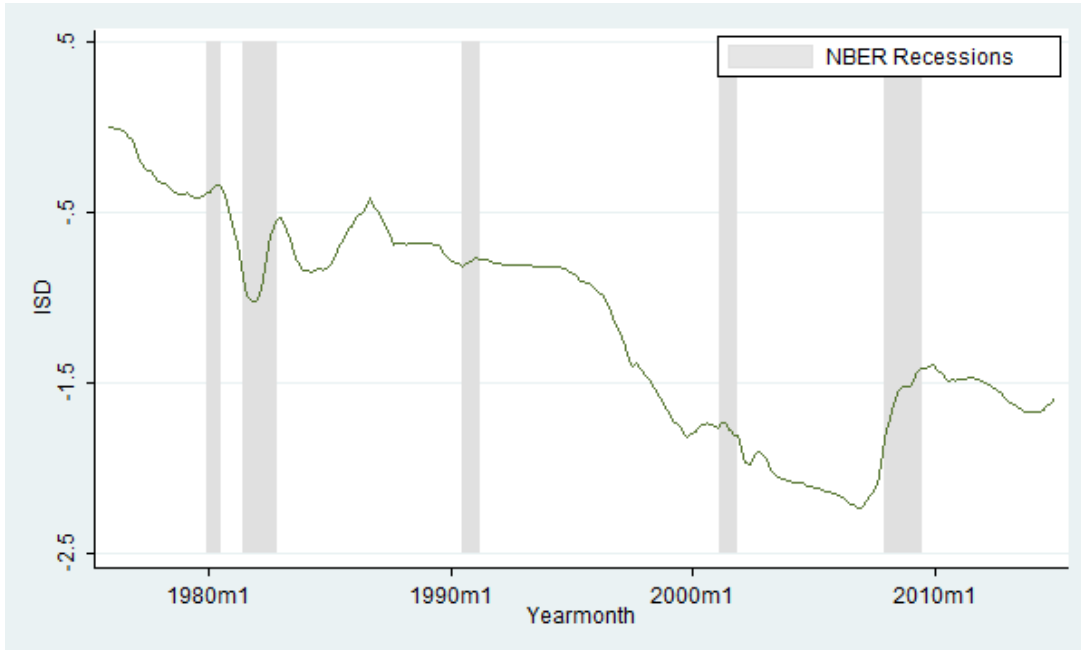


Figure 6: **Incremental Contribution of Learning to Cumulative Sum of Squared Errors Difference.** This figure plots the incremental cumulative sum of squared errors difference, attributable to learning about dividend dynamics, for the period between 1976 and 2015. Dividend model parameters are estimated based on data since 1946.

#### 4.2.2 The Dot-Com Crash

Figure 5 suggests that the Dot-Com crash plays an especially important role in the return predictability results. To emphasize that our learning model’s forecasting performance and the incremental contribution of learning to this performance is not exclusively restricted to the few years surrounding the Dot-Com crash, we recalculate  $R$ -squared value and incremental  $R$ -squared value separately for using our learning model to predict annual stock index returns, but for the Dot-Com crash alone, defined as between March 2000 and October 2002, and for the sub-period even after excluding the Dot-Com crash. Results are reported in Table 11. We find that the forecasting performance of our learning model is significant excluding the Dot-Com crash, and so is the incremental contribution of learning to this predictive performance.

#### 4.2.3 Long-Run Risks Model and the Term-Structure of Discount Rates

Earlier in Table 9, we have shown that the effect of learning on the variation in annual stock index returns is not fully captured by its effect on stock yields. We also know from comparing results in Table 8 and Table 11 that, assuming learning, our long-run

	Dot-Com Crash		Rest of the Sample	
	$R^2$	$p$ -value	$R^2$	$p$ -value
Learning Model	0.725	0.287	0.090	0.074
Full Information Model	0.762	0.251	-0.123	1.000
Incremental	-0.153	1.000	0.189	0.007

Table 11: **Stock Index Returns and the Long-Run Risks Model (Dot-Com Crash)**. This table reports out-of-sample  $R$ -square value for predicting stock index returns over the next year using our long run risks model, assuming investors either learn, or do not learn, about dividend model parameters, and the corresponding  $p$ -value from the adjusted-MSPE statistic of Clark and West (2007). Also reported is incremental out-of-sample  $R$ -square value. Dividends are estimated using data since 1946. Statistics are based on two subsamples of the data: 1) the Dot-Com crash between March 2000 and October 2002 and 2) excluding the Dot-Com crash.

risks model does a significantly better job than stock yields in predicting the variation in annual stock index returns. This outperformance may be because our model is able to capture the full effect of learning on the variation in annual stock index returns. To see this, we run bi-variate regressions of stock index returns over the next year on expected returns of our learning model and investors beliefs about one of the nine dividend model parameters. For example, for parameter  $\rho$ , let  $\hat{\rho}_t$  denote the estimate of  $\rho$  based on data up to time- $t$ , we run the bi-variate regression:

$$R_{t+1} = \alpha + \beta \cdot E_t [R_{t+1} | \mathcal{L}] + \gamma \cdot \hat{\rho}_t + \epsilon_{t+1} \quad (38)$$

where  $E_t [R_{t+1} | \mathcal{L}]$  stands for expected return of our learning model. Estimated coefficients are reported in Table 12. If our learning model is able to fully capture the effect of learning on the variation in annual stock index returns, then in contrast to results in Table 9, we should expect none of the  $\gamma$  estimates to be different from zero. Results show that this is indeed the case.

### 4.3 The Role of Epstein and Zin (1989) Preferences

To show that Epstein and Zin (1989) preferences is critical to our long-run risks model's performance in predicting annual stock index returns, we replace Epstein and Zin (1989) preferences in our model with Constant Relative Risk Aversion (CRRA) preferences:

	$\mu_d$	$\phi$	$\sigma_d$
$\beta$	0.801*** (0.106)	0.802*** (0.101)	0.810*** (0.097)
$\gamma$	0.491 (1.859)	0.284 (0.308)	0.804 (1.717)
$\underline{R^2}$			
Return	0.223	0.235	0.224
Exc. Return	0.216	0.229	0.218
	$\rho$	$\sigma_x$	$\mu_q$
$\beta$	0.834*** (0.100)	0.823** (0.097)	0.783*** (0.108)
$\gamma$	0.047 (0.097)	-0.702 (1.427)	-0.748 (1.040)
$\underline{R^2}$			
Return	0.224	0.224	0.234
Exc. Return	0.218	0.218	0.228
	$\theta$	$\sigma_q$	$\lambda_{dq}$
$\beta$	0.811*** (0.110)	0.820*** (0.095)	0.785*** (0.104)
$\gamma$	0.002 (0.127)	-0.140 (0.170)	0.043 (0.078)
$\underline{R^2}$			
Return	0.222	0.225	0.224
Exc. Return	0.216	0.218	0.217

Table 12: **Stock Index Returns, the Long-Run Risks Model, and Investors' Beliefs about Dividend Model Parameters:** This table reports the coefficient estimates and  $R$ -square value from regressing stock index returns over the next year on expected returns of our learning model and investors beliefs about one of our dividend model parameters, estimated using our dividend model at each point in time based on only data available at the time. Regression is based on data between 1976 and 2015. Dividend model parameters are estimated based on data since 1946. Newey and West (1987) standard errors are reported in parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using \*, \*\*, and \*\*\*.

$$U_t = \sum_{t=0}^{\infty} \delta^t \frac{\tilde{C}_t^{1-\alpha}}{1-\alpha} \quad (39)$$

where we set preference parameters  $\alpha$  and  $\delta$  so that the model’s risk free rates and risk premium can roughly match the data between 1946 and 2015. We then report, in Table 13,  $R$ -square value for predicting annual stock index returns using the CRRA model, either assuming learning or assuming full information. We see that, assuming investors learn about dividend dynamics,  $R$ -square value for predicting stock index returns reduces from 22.4 percent for Epstein and Zin (1989) preferences to 11.1 percent for CRRA preferences, and the incremental contribution of learning to  $R$ -square value reduces from 17.1 percent to 6.6 percent. It is clear from these results that CRRA preferences cannot fully capture the effect of learning about dividend dynamics on the variation in annual stock index returns.

	$R^2$	$p$ -value
Learning Model	0.111	0.037
Full Information Model	0.049	0.174
Incremental	0.066	0.115

Table 13: **Stock Index Returns and the Long-Run Risks Model (CRRA Preferences)**. This table reports out-of-sample  $R$ -square value for predicting stock index returns over the next year using our long run risks model, assuming investors either learn, or do not learn, about dividend parameters, and the corresponding  $p$ -value from the adjusted-MSPE statistic of Clark and West (2007). Also reported is incremental out-of-sample  $R$ -square value. Dividends are estimated using data since 1946. Statistics are based on data between 1976 and 2015.

#### 4.4 Robustness: An Alternative Long-Run Risks Model

To show that our findings are not restricted to one specification of the long-run risks model, we propose an alternative long-run risks model and examine its performance in predicting annual stock index returns. In contrast to our default model, where stochastic volatility of consumption growth rates serve as one of the model’s state variables, we assume consumption volatility to be constant over time in this alternative model. To make up for the lost state variable, we assume instead that the correlations between shocks  $\epsilon_{c,t+1}$  to consumption growth rates and shocks  $\epsilon_{d,t+1}$  and  $\epsilon_{e,t+1}$  to dividend and earnings

growth rates are equal and time varying, i.e.  $\lambda(\epsilon_{c,t+1}, \epsilon_{d,t+1}) = \lambda(\epsilon_{c,t+1}, \epsilon_{e,t+1}) = \lambda_t$ .<sup>24</sup> In other words, consumption growth rates and the variance-covariance matrix of shocks to consumption, dividends, and retention ratios can together be summarized by the following system of equations:

$$\Delta \tilde{c}_{t+1} - \mu_c = \frac{1}{\gamma} x_t + \sigma_c \epsilon_{c,t+1}$$

$$\begin{pmatrix} \epsilon_{c,t+1} \\ \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{\lambda,t+1} \\ \epsilon_{q,t+1} \\ \epsilon_{\pi,t+1} \end{pmatrix} \sim \text{i.i.d. } \mathbb{N} \left( \mathbf{0}, \begin{pmatrix} 1 & \lambda_t & 0 & 0 & \frac{\sqrt{\sigma_d^2 + \sigma_q^2 + 2\sigma_d\sigma_q\lambda_{dq}} - \sigma_d}{\sigma_q} \lambda_t & 0 \\ \lambda_t & 1 & 0 & 0 & \lambda_{dq} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{\sqrt{\sigma_d^2 + \sigma_q^2 + 2\sigma_d\sigma_q\lambda_{dq}} - \sigma_d}{\sigma_q} \lambda_t & \lambda_{dq} & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \right)$$

$$\lambda_{t+1} = \omega \lambda_t + \sigma_\lambda \epsilon_{\lambda,t+1}, \quad (40)$$

where  $\lambda_t$  is the time varying correlation between  $\epsilon_{c,t+1}$  and both  $\epsilon_{d,t+1}$  and  $\epsilon_{e,t+1}$ . Our calibration of parameters in this alternative model is:  $\sigma_c = 0.0234$ ,  $\omega = 0.962$ , and  $\gamma = 3$ . In choosing how we calibrate these parameters, we try to minimize our deviations from the default model. We solve this alternative model in Appendix A.4. In solving this model, we closely follow the steps in Bansal and Yaron (2004). The model consists of four state variables: latent variables  $x_t$  and  $\lambda_t$ , retention ratios, and inflation rates. We can solve for price-to-dividend ratio as a linear function of these four state variables:

$$p_t - d_t = A_{d,0} + A_{d,1}x_t + A_{d,2}\lambda_t + A_{d,3}(q_t - \mu_q) + A_{d,4}(\Delta\pi_t - \mu_\pi). \quad (41)$$

We can solve for expected returns over the next year as:

$$E_t[R_{t+1}] = A_{r,0} + A_{r,1}x_t + A_{r,2}\lambda_t + A_{r,4}(\Delta\pi_t - \mu_\pi), \quad (42)$$

where coefficients  $A_{d,\cdot}$  and  $A_{r,\cdot}$ , derived in Appendix A.4, are functions of the parameters governing investors' preferences, consumption dynamics, and dividend dynamics. We note that, substituting (41) into (42), we can avoid estimating  $\lambda_t$  directly and instead write expected returns over the next year as a function state variables that can be estimated

---

<sup>24</sup>It can be derived that the correlation between  $\epsilon_{c,t+1}$  and  $\epsilon_{q,t+1}$  is then  $\frac{\sqrt{\sigma_d^2 + \sigma_q^2 + 2\sigma_d\sigma_q\lambda_{dq}} - \sigma_d}{\sigma_q} \lambda_t$ .

from dividend dynamics and price-to-dividend ratios:

$$E_t [R_{t+1}] = A_0 + A_1 x_t + A_2 (p_t - d_t) + A_3 (q_t - \mu_q) + A_4 (\Delta \pi_t - \mu_\pi),$$

$$A_0 = \frac{A_{r,0} A_{d,2} - A_{r,2} A_{d,0}}{A_{d,2}}, \quad A_1 = \frac{A_{r,1} A_{d,2} - A_{r,2} A_{d,1}}{A_{d,2}}, \quad (43)$$

$$A_2 = \frac{A_{r,2}}{A_{d,2}}, \quad A_3 = -\frac{A_{r,2} A_{d,3}}{A_{d,2}}, \quad A_4 = \frac{A_{r,4} A_{d,2} - A_{r,2} A_{d,4}}{A_{d,2}}. \quad (44)$$

In Table 14, we report out-of-sample  $R$ -square value for predicting stock index returns using this alternative model, assuming learning. These results are very similar to those reported using the default model. Between 1976 and 2013, this alternative learning model predicts as much as 22.6 percent of the variation in annual stock index returns. Learning accounts for over half of this 22.6 percent.

	$R^2$	$p$ -value
Learning Model	0.226	0.002
Full Information Model	0.065	0.117
Incremental	0.171	0.009

Table 14: **Stock Index Returns and the Long-Run Risks Model (Alternative Specification).** This table reports out-of-sample  $R$ -square value for predicting stock index returns over the next year using our long run risks model, assuming investors either learn, or do not learn, about dividend model parameters, and the corresponding  $p$ -value from the adjusted-MSPE statistic of Clark and West (2007). Also reported is incremental out-of-sample  $R$ -square value. Dividends are estimated using data since 1946. Statistics are based on data between 1976 and 2015.

We proceed to discuss the differences between the default and the alternative long-run risks models. In Figure 7, we plot the time series of expected returns of the two models, as well as stock index returns realized over the next year, between 1976 and 2015. As Figure 7 shows, expected returns of the two models are virtually identical across the sample period. Instead, the two models differ in how expected returns are decomposed into risk premium and risk free rates. To show this, we decompose expected returns of each of the two models into model implied risk free rates and model implied risk premium. The derivation of risk free rates and risk premium as a function of state variables and model parameters are detailed in Appendix A.4



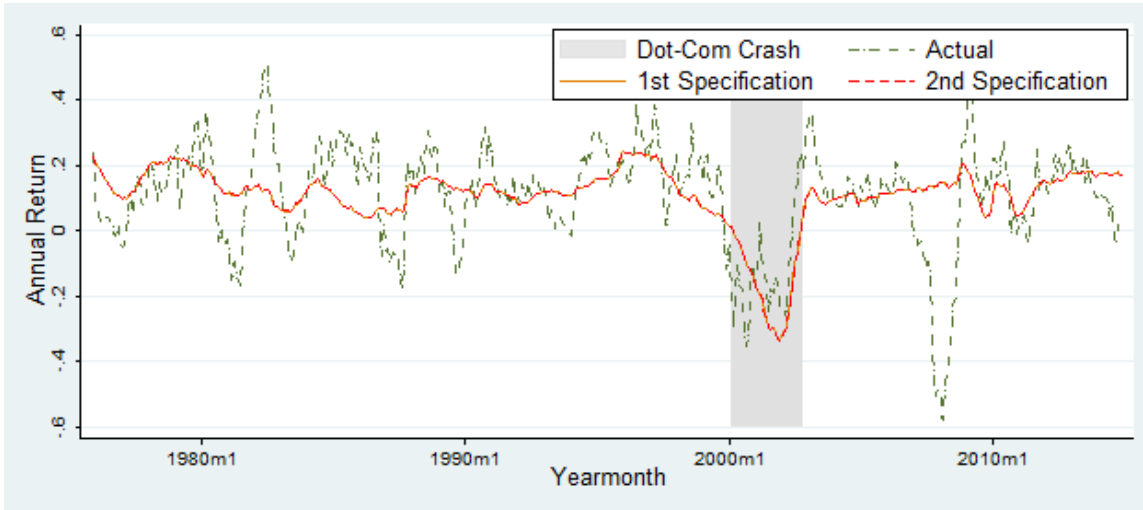


Figure 7: **Long-Run Risks Model Implied Expected Returns.** This figure plots expected returns of the stock index implied by each of the two long run risks models specified, as well as the actual stock index returns realized over the next year, for the period between 1976 and 2015. Dividends are estimated using data since 1946.

In Figure 8, we plot risk free rates and risk premium of each of the two models, as well as actual risk free rates and excess returns realized over the next year, for the period between 1976 and 2015. Interestingly, Figure 8 shows that the two models have completely different implications on the decomposition of expected returns into risk free rates and risk premium. That is, according to consumption dynamics in our default model, almost all of the variation in expected returns is attributable to the variation in risk free rates, whereas the risk premium hardly changes over time. To the contrary, according to consumption dynamics in the alternative model, almost all of the variation in expected returns is attributable to the variation in risk premium, whereas risk free rates hardly change over time. Clearly, we know from the data that risk free rates are relatively constant over time. Thus, we can infer that, consumption dynamics in the alternative model is the more realistic one of the two. In other words, because different models of the consumption process can have different implications for the decomposition of expected returns into risk premium and risk free rates, we can use this decomposition to shed light on the true consumption dynamics. However, because modeling consumption is not the focus of this paper, we leave this to potential future research.

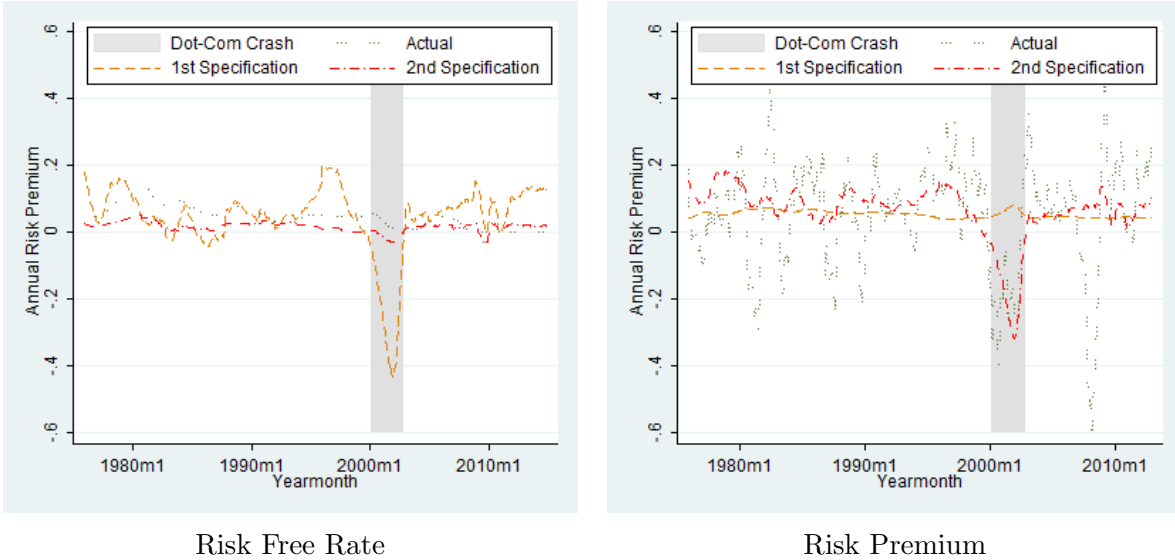


Figure 8: **Long-Run Risks Model Implied Risk Free Rates and Risk Premium.** This figure plots risk free rate and risk premium derived from the two specifications of our long run risk model, as well as the actual risk free rate and excess returns over the next year, for the period between 1976 and 2015. Dividend model parameters are estimated based on data since 1946.

## 5 Conclusion

In this paper, we develop a time series model for stock index dividend growth rates that is inspired by both the latent variable model of Cochrane (2008), van Binsbergen and Koijen (2010), and others and the vector-autoregressive model of Campbell and Shiller (1988b). The model shows strong performance in predicting dividend growth rates. We find that some parameters in our dividend model are difficult to estimate with precision in finite sample. As a consequence, learning about dividend model parameters significantly changes investors beliefs about future dividends and the nature of the long run risks in the economy.

We show how to evaluate the economic and statistical significance of learning about parameters in the dividend process in determining asset prices and returns. We argue that a better asset pricing model should forecast returns better. We find that a long run risks model that incorporates learning about dividend dynamics is surprisingly successful in forecasting stock index returns. While the long run risks model, assuming learning, explains 22.4 percent of the variation in annual stock index returns, shutting down learning reduces the  $R$ -square value to 6.4 percent. This drop in  $R$ -square value is statistically significant. Our findings highlight the joint importance of investors aversion to long run risks and investors learning about these risks for understanding asset prices.

## References

- ALBUQUERQUE, R., M. S. EICHENBAUM, AND S. REBELO (2015): “Valuation Risk and Asset Pricing,” *Working Paper*.
- ANG, A., AND G. BEKAERT (2007): “Stock Return Predictability: Is It There?,” *Review of Financial Studies*, 20, 651–707.
- BAKER, M., AND J. WURGLER (2000): “The Equity Share in New Issues and Aggregate Stock Returns,” *Journal of Finance*, 55, 2219–2257.
- BANSAL, R., AND A. YARON (2004): “Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles,” *Journal of Finance*, 59, 1481–1509.
- BEELER, J., AND J. Y. CAMPBELL (2012): “The Long Run Risks Model and Aggregate Asset Prices: An Empirical Assessment,” *Critical Finance Review*, 1, 141–182.
- BOUDOUKH, J., R. MICHAELY, M. RICHARDSON, AND M. R. ROBERTS (2007): “On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing,” *The Journal of Finance*, 62, 877–915.
- CAMPBELL, J. Y., AND R. J. SHILLER (1988): “The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors,” *Review of financial studies*, 1, 195–228.
- CAMPBELL, J. Y., AND R. J. SHILLER (1988b): “The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors,” *Review of Financial Studies*, 1, 195–227.
- CAMPBELL, J. Y., AND T. VUOLTEENAHO (2004): “Bad Beta, Good Beta,” *Journal of Finance*, 55, 2219–2257.
- CHEN, L., Z. DA, AND R. PRIESTLEY (2012): “Dividend Smoothing and Predictability,” *Management Science*, 58, 1834–1853.
- CLARK, T. E., AND K. D. WEST (2007): “Approximately Normal Tests for Equal Predictive Accuracy in Nested Models,” *Journal of econometrics*, 138, 291–311.
- COCHRANE, J. H. (2008): “The Dog That Did Not Bark: A Defense of Return Predictability,” *Review of Financial Studies*, 21, 1533–1575.

- COGLEY, T., AND T. J. SARGENT (2008): “Anticipated Utility and Rational Expectations as Approximations of Bayesian Decision Making,” *International Economic Review*, 49, 185–221.
- COLLIN-DUFRESNE, P., M. JOHANNES, AND L. A. LOCHSTOER (2016): “Parameter Learning in General Equilibrium: The Asset Pricing Implications,” *The American Economic Review*, 106, 664–698.
- CROCE, M. M., M. LETTAU, AND S. C. LUDVIGSON (2014): “Investor Information, Long-Run Risk, and the Term Structure of Equity,” *Working Paper*.
- DA, Z., R. JAGANNATHAN, AND J. SHEN (2014): “Growth Expectations, Dividend Yields, and Future Stock Returns,” *Working Paper*.
- EPSTEIN, L. G., AND S. E. ZIN (1989): “Substitution, Risk Aversion, and the Intertemporal Behavior of Consumption and Asset Returns: A Theoretical Framework,” *Econometrica*, 57, 937–969.
- FAMA, E. F., AND K. R. FRENCH (1988): “Dividend Yield and Expected Stock Returns,” *Journal of Financial Economics*, 22, 3–25.
- FAMA, E. F., AND K. R. FRENCH (1993): “Common Risk Factors in the Returns on Stocks and Bonds,” *Journal of Financial Economics*, 33, 3–56.
- GOYAL, A., AND I. WELCH (2008): “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction,” *Review of Financial Studies*, 21, 1455–1508.
- HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press.
- HANSEN, L. P. (2007): “Beliefs, Doubts and Learning: Valuing Macroeconomic Risk,” *The American Economic Review*, 97, 1–30.
- JAGANNATHAN, R., AND S. MARAKANI (2015): “Price-Dividend Ratio Factor Proxies for Long-Run Risks,” *Review of Asset Pricing Studies*, 5, 1–47.
- JOHANNES, M. S., L. A. LOCHSTOER, AND Y. MOU (2014): “Learning about Consumption Dynamics,” *Journal of Finance*, Forthcoming.
- KELLY, B., AND S. PRUITT (2013): “Market Expectations in the Cross-Section of Present Values,” *Journal of Finance*, 68, 1721–1756.

- KREPS, D. M. (1998): “Anticipated Utility and Dynamic Choice,” *Econometric Society Monographs*, 29, 242–274.
- LAMONT, O. (1998): “Earnings and Expected Returns,” *Journal of Finance*, 53, 1563–1587.
- LETTAU, M., AND S. LUDVIGSON (2001): “Consumption, Aggregate Wealth, and Expected Stock Returns,” *Journal of Finance*, 56, 815–849.
- LETTAU, M., AND S. C. LUDVIGSON (2005): “Expected Returns and Expected Dividend Growth,” *Journal of Financial Economics*, 76, 583–626.
- LETTAU, M., AND S. VAN NIEUWERBURGH (2008): “Reconciling the Return Predictability Evidence,” *Review of Financial Studies*, 21, 1607–1652.
- LEWELLEN, J., AND J. SHANKEN (2002): “Learning, Asset-Pricing Tests, and Market Efficiency,” *The Journal of Finance*, 57, 1113–1145.
- LI, Y., D. T. NG, AND B. SWAMINATHAN (2013): “Predicting Market Returns Using Aggregate Implied Cost of Capital,” *Journal of Financial Economics*, 110, 419–436.
- LJUNG, G. M., AND G. E. BOX (1978): “On a Measure of Lack of Fit in Time Series Models,” *Biometrika*, 65(2), 297–303.
- NEWBY, W. K., AND K. D. WEST (1987): “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703–708.
- PIAZZESI, M., AND M. SCHNEIDER (2010): “Interest Rate Risk in Credit Markets,” *The American Economic Review*, 100, 579–584.
- POLK, C., S. THOMPSON, AND T. VUOLTEENAHO (2006): “Cross-Sectional Forecasts of the Equity Premium,” *Journal of Finance*, 55, 2219–2257.
- STAMBAUGH, R. F. (1999): “Predictive Regressions,” *Journal of Financial Economics*, 54, 375–421.
- TIMMERMANN, A. G. (1993): “How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices,” *The Quarterly Journal of Economics*, pp. 1135–1145.

VAN BINSBERGEN, J., W. HUESKES, R. KOIJEN, AND E. VRUGT (2013): “Equity Yields,” *Journal of Financial Economics*, 110, 503–519.

VAN BINSBERGEN, J. H., AND R. S. KOIJEN (2010): “Predictive Regression: A Present-Value Approach,” *Journal of Finance*, 65, 1439–1471.

# A Appendix

## A.1 Estimation of Parameters in Our Dividend Model

We estimate parameters of the following system of equations that jointly describe the dividend, earnings, and inflation processes (see (6), (9), and (15)).

$$\begin{aligned}
\Delta d_{t+1} - \mu_d &= x_t + \phi(q_t - \mu_q) + \sigma_d \epsilon_{d,t+1}, \\
\Delta e_{t+1} - \mu_e &= x_t + (\theta + \phi - 1)(q_t - \mu_q) + \sigma_d \epsilon_{d,t+1} + \sigma_q \epsilon_{q,t+1}, \\
x_{t+1} &= \rho x_t + \sigma_x \epsilon_{x,t+1}, \\
q_{t+1} - \mu_q &= \theta(q_t - \mu_q) + \sigma_q \epsilon_{q,t+1}, \\
\Delta \pi_{t+1} - \mu_\pi &= \theta(\Delta \pi_t - \mu_\pi) + \sigma_\pi \epsilon_{\pi,t+1}, \\
\begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{q,t+1} \\ \epsilon_{\pi,t+1} \end{pmatrix} &\sim \text{i.i.d. } \mathbb{N} \left( \mathbf{0}, \begin{pmatrix} 1 & 0 & \lambda_{dq} & 0 \\ 0 & 1 & 0 & 0 \\ \lambda_{dq} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right).
\end{aligned} \tag{45}$$

For parameters in the fourth equation of (45), we run an autoregression on retention ratios:

$$q_{t+1} - \alpha_1 = b_1(q_t - \alpha_1) + \nu_{q,t+1}, \quad \nu_{q,t+1} \sim \mathbb{N}(0, \varsigma_q). \tag{46}$$

For parameters in the fifth equation of (45), we run an autoregression on inflation rates:

$$\Delta \pi_{t+1} - \alpha_2 = b_2(\Delta \pi_t - \alpha_2) + \nu_{\pi,t+1}, \quad \nu_{\pi,t+1} \sim \mathbb{N}(0, \varsigma_q). \tag{47}$$

For the remaining parameters in the first to third equations and the correlation matrix of (45), we note that dividend growth rates are affected by contemporaneous earnings and vice versa. So we estimate the cointegrated process of dividends and earnings through the following system of equations:

$$\begin{aligned}
\Delta d_{t+1} - \alpha_3 &= y_{t+1} + b_3(\Delta e_{t+1} - \alpha_3) + b_4(q_t - \hat{\alpha}_1) + \nu_{d,t+1} \\
y_{t+1} &= b_5 y_t + \nu_{y,t+1} \\
\begin{pmatrix} v_{d,t+1} \\ v_{y,t+1} \end{pmatrix} &\sim \text{i.i.d. } \mathbb{N} \left( \mathbf{0}, \begin{pmatrix} \varsigma_d & 0 \\ 0 & \varsigma_y \end{pmatrix} \right).
\end{aligned} \tag{48}$$

To apply the Kalman filter, let  $\hat{y}_{t|s}$  denote the time- $s$  expectation of the latent variable  $y_t$  and  $P_{t|s}$  denote the variance of  $y_t$  conditioning on information in time- $s$ . Set initial

conditions  $\hat{y}_{0|0} = 0$  and  $P_{0|0} = \frac{\sigma_y^2}{1-b_5^2}$ . We can then iterate the following system of equations:

$$\begin{aligned}\hat{y}_{t+1|t} &= b_5 \cdot \hat{y}_{t|t}, & P_{t+1|t} &= b_5^2 \cdot P_{t|t} + \varsigma_y^2, \\ \epsilon_{t+1} &= \Delta d_{t+1} - \alpha_3 - \hat{y}_{t+1|t} - b_3(\Delta e_{t+1} - \alpha_3) - b_4(q_t - \alpha_1), \\ \hat{y}_{t+1|t+1} &= \hat{y}_{t+1|t} + \frac{P_{t+1|t}}{P_{t+1|t} + \varsigma_d^2} \cdot \epsilon_{t+1}, & P_{t+1|t+1} &= P_{t+1|t} - \frac{P_{t+1|t}^2}{P_{t+1|t} + \varsigma_d^2}.\end{aligned}\quad (49)$$

To estimate parameters in (48) using data up to time- $\tau$ , define the log likelihood function:

$$\log(\mathcal{L}) = - \sum_{t=0}^{\tau-1} \left( \log(P_{t+1|t} + \varsigma_d^2) + \frac{\epsilon_{t+1}^2}{P_{t+1|t} + \varsigma_d^2} \right). \quad (50)$$

In our implementation of Kalman filter that, because we use overlapping monthly data, we obtain twelve log likelihoods, one for the 12 month periods that begin in January, one for the 12 month periods that begin in February, etc. We choose model parameters by maximizing the sum of the twelve log likelihood.

Based on estimates from (46), (47), and (48), we compute residual estimates:

$$\begin{aligned}\hat{\nu}_{d,t+1} &= \Delta d_{t+1} - \hat{\alpha}_3 - \hat{y}_{t+1|t} - \hat{b}_3(\Delta e_{t+1} - \hat{\alpha}_1) - \hat{b}_4(q_t - \hat{\alpha}_2), \\ \hat{\nu}_{y,t+1} &= \hat{y}_{t+1|t} - \hat{b}_5 \hat{y}_{t|t-1}, \\ \hat{\nu}_{q,t+1} &= (q_{t+1} - \hat{\alpha}_1) - \hat{b}_1(q_t - \hat{\alpha}_1), \\ \hat{\nu}_{\pi,t+1} &= (\Delta \pi_{t+1} - \hat{\alpha}_2) - \hat{b}_2(\Delta \pi_t - \hat{\alpha}_2).\end{aligned}\quad (51)$$

and use these estimated residuals to estimate covariances among  $\nu_{d,t+1}$ ,  $\nu_{x,t+1}$ ,  $\nu_{q,t+1}$ , and  $\nu_{\pi,t+1}$ . Last, we note that (47) is our inflation model and we can map (46) and (48) to our dividend model as follows. We can map (46) to the fourth equation of our dividend model. Also, by substituting  $\Delta e_{t+1} = q_{t+1} - q_t + \Delta d_{t+1}$  and (46) into (48), we can re-arrange (48) into a format consistent with the first and second equations of our dividend model:

$$\begin{aligned}\Delta d_{t+1} - \alpha_3 &= x_t + \frac{b_3(b_1 - 1) + b_4}{1 - b_3} (q_t - \hat{\alpha}_1) + \frac{b_3}{1 - b_3} \nu_{q,t+1} + \frac{1}{1 - b_3} \nu_{d,t+1} \\ x_{t+1} &= b_5 x_t + \nu_{x,t+1} \\ \begin{pmatrix} \nu_{d,t+1} \\ \nu_{x,t} \end{pmatrix} &\sim \text{i.i.d. } \mathbb{N} \left( \mathbf{0}, \begin{pmatrix} \varsigma_d & 0 \\ 0 & \frac{1}{1-b_3} \varsigma_y \end{pmatrix} \right).\end{aligned}\quad (52)$$

where  $\frac{y_t}{1-b_3}$ . So we can convert estimates in (46), (48), (49), and covariance estimates to



estimates of parameters in our dividend model:

$$\begin{aligned}
\hat{\mu}_d &= \hat{\alpha}_3, \quad \hat{\mu}_q = \hat{\alpha}_1, \quad \hat{\phi} = \frac{\hat{b}_3(\hat{b}_1 - 1) + \hat{b}_2}{1 - \hat{b}_3}, \quad \hat{\rho} = \hat{b}_5, \quad \hat{\theta} = \hat{b}_1, \quad \hat{\eta} = \hat{b}_2, \\
\hat{\sigma}_x &= \frac{1}{1 - \hat{b}_3} \hat{\zeta}_y, \quad \hat{\sigma}_d = \sqrt{\left(\frac{\hat{b}_3}{1 - \hat{b}_3}\right)^2 \hat{\zeta}_q^2 + \left(\frac{1}{1 - \hat{b}_3}\right)^2 \hat{\zeta}_d^2 + \frac{2\hat{b}_3}{(1 - \hat{b}_3)^2} \text{cov}(\hat{\nu}_d, \hat{\nu}_q)}, \quad \hat{\sigma}_q = \hat{\zeta}_q, \quad \hat{\sigma}_\pi = \hat{\zeta}_\pi, \\
\hat{\lambda}_{d\pi} &= \frac{1}{\hat{\sigma}_d \hat{\sigma}_\pi} \left( \frac{\hat{b}_3}{1 - \hat{b}_3} \text{cov}(\hat{\nu}_{q,t+1}, \hat{\nu}_{\pi,t+1}) + \frac{1}{1 - \hat{b}_3} \text{cov}(\hat{\nu}_{d,t+1}, \hat{\nu}_{\pi,t+1}) \right), \\
\hat{\lambda}_{dx} &= \frac{1}{\hat{\sigma}_x \hat{\sigma}_d} \left( \frac{\hat{b}_3}{(1 - \hat{b}_3)^2} \text{cov}(\hat{\nu}_{y,t}, \hat{\nu}_{q,t+1}) + \left(\frac{\hat{b}_3}{1 - \hat{b}_3}\right)^2 \text{cov}(\hat{\nu}_{y,t}, \hat{\nu}_{d,t+1}) \right), \\
\hat{\lambda}_{x\pi} &= \frac{1}{\hat{\sigma}_x \hat{\sigma}_\pi} \frac{1}{1 - \hat{b}_1} \text{cov}(\hat{\nu}_{y,t}, \hat{\nu}_{\pi,t+1}), \quad \hat{\lambda}_{q\pi} = \frac{1}{\hat{\sigma}_q \hat{\sigma}_\pi} \text{cov}(\hat{\nu}_{q,t+1}, \hat{\nu}_{\pi,t+1}), \\
\hat{\lambda}_{dq} &= \frac{1}{\hat{\sigma}_d \hat{\sigma}_q} \left( \frac{\hat{b}_3}{1 - \hat{b}_3} \hat{\zeta}_q^2 + \frac{1}{1 - \hat{b}_3} \text{cov}(\hat{\nu}_{d,t+1}, \hat{\nu}_{q,t+1}) \right), \quad \hat{\lambda}_{qx} = \frac{1}{\hat{\sigma}_q \hat{\sigma}_y} \frac{\hat{b}_3}{1 - \hat{b}_3} \text{cov}(\hat{\nu}_{q,t+1}, \hat{\nu}_{y,t}). \quad (53)
\end{aligned}$$

## A.2 Bootstrap Simulation

Each simulation is based on 100,000 iterations. First, we simulate innovations to dividend growth rates and retention ratios:

$$\begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{q,t+1} \end{pmatrix} \sim \text{i.i.d. } \mathbb{N} \left( \mathbf{0}, \begin{pmatrix} 1 & 0 & \lambda_{dq} \\ 0 & 1 & 0 \\ \lambda_{dq} & 0 & 1 \end{pmatrix} \right). \quad (54)$$

Dividend model parameters used for simulations are those reported in Table 2, which are estimated based on the full data sample between 1946 and 2015. From these innovations, we can simulate the latent variable  $x_t$  and retention ratios iteratively as:

$$\begin{aligned}
x_{t+1} &= \rho x_t + \sigma_x \epsilon_{x,t+1}, \\
q_{t+1} - \mu_q &= \theta (q_t - \mu_q) + \sigma_q \epsilon_{q,t+1}. \quad (55)
\end{aligned}$$

Given the simulated time series of the latent variable  $x_t$  and retention ratios, we can simulate dividend growth rates iteratively as:

$$\begin{aligned}
\Delta d_{t+1} - \mu_d &= x_t + \phi (q_t - \mu_q) + \sigma_d \epsilon_{d,t+1}, \\
\Delta e_{t+1} &= q_{t+1} - q_t + \Delta d_{t+1}. \quad (56)
\end{aligned}$$

### A.3 Proof of Proposition 1

Let  $\mathcal{M}_0$  be the true asset pricing model and let  $\mathcal{M}_i$  and  $\mathcal{M}_j$  be two candidate models. Define  $\epsilon_{t+1} = R_{t+1} - E_t[R_{t+1}|\mathcal{M}_0]$ . We can write:

$$\begin{aligned}
& E \left[ (E_t[R_{t+1}|\mathcal{M}_0] - E_t[R_{t+1}|\mathcal{M}_i])^2 \right] \\
&= E \left[ (R_{t+1} - E_t[R_{t+1}|\mathcal{M}_i])^2 \right] + E \left[ \epsilon_{t+1}^2 \right] - 2 \cdot E \left[ (R_{t+1} - E_t[R_{t+1}|\mathcal{M}_i]) \epsilon_{t+1} \right] \\
&= E \left[ (R_{t+1} - E_t[R_{t+1}|\mathcal{M}_i])^2 \right] + E \left[ \epsilon_{t+1}^2 \right] + 2 \cdot E \left[ (E_t[R_{t+1}|\mathcal{M}_i] \epsilon_{t+1}) - 2 \cdot E \left[ R_{t+1} \epsilon_{t+1} \right] \right] \\
&= E \left[ (R_{t+1} - E_t[R_{t+1}|\mathcal{M}_i])^2 \right] + E \left[ \epsilon_{t+1}^2 \right] - 2 \cdot E \left[ R_{t+1} \epsilon_{t+1} \right].
\end{aligned}$$

Last equality assumes frictionless and efficient market and investors having rational expectations. As a result, marginal investor's investment decisions are based on all information available and so  $\epsilon_{t+1}$  is orthogonal to any variable that is time- $t$  measurable.  $E \left[ \epsilon_{t+1}^2 \right]$  and  $E \left[ R_{t+1} \epsilon_{t+1} \right]$  are independent of the model  $\mathcal{M}_i$  and so:

$$\begin{aligned}
& E \left[ (E_t[R_{t+1}|\mathcal{M}_0] - E_t[R_{t+1}|\mathcal{M}_i])^2 \right] < E \left[ (E_t[R_{t+1}|\mathcal{M}_0] - E_t[R_{t+1}|\mathcal{M}_j])^2 \right] \\
&\Leftrightarrow E \left[ (R_{t+1} - E_t[R_{t+1}|\mathcal{M}_i])^2 \right] < E \left[ (R_{t+1} - E_t[R_{t+1}|\mathcal{M}_j])^2 \right] \\
&\Leftrightarrow 1 - \frac{E \left[ (R_{t+1} - E_t[R_{t+1}|\mathcal{M}_i])^2 \right]}{E \left[ (R_{t+1} - E \left[ R_{t+1} \right])^2 \right]} > 1 - \frac{E \left[ (R_{t+1} - E_t[R_{t+1}|\mathcal{M}_j])^2 \right]}{E \left[ (R_{t+1} - E \left[ R_{t+1} \right])^2 \right]}.
\end{aligned}$$

### A.4 Derivation of Price-Dividend Ratios and Expected Returns in Long-Run Risks Models

#### A.4.1 The Default Model

We derive price-to-dividend ratios and expected returns implied by our long-run risk model, which features dividend dynamics in (3), consumption dynamics in (27) and (28), and investors preferences in (25). Our derivation closely follows the steps in Bansal and Yaron (2004). The log stochastic discount factor is given as:

$$m_{t+1} = \zeta \log(\delta) - \frac{\zeta}{\psi} \Delta \tilde{c}_{t+1} + (\zeta - 1) \tilde{R}_{t+1}^c. \quad (57)$$

where  $\zeta = \frac{1-\alpha}{1-\frac{\alpha}{\psi}}$ ,  $\alpha$  is risk aversion, and  $\psi$  is the elasticity of intertemporal substitution. Let  $z_{c,t}$  be the log wealth-to-consumption ratio, by first order Taylor series approximation,

log real return of the representative agent's wealth portfolio can be written as:

$$\tilde{R}_{t+1}^c = g_0 + g_1 z_{c,t+1} - z_{c,t} + \Delta \tilde{c}_{t+1}. \quad (58)$$

The log-linearizing constants are:

$$g_0 = \log(1 + \exp(\bar{z}_c)) - g_1(\bar{z}_c) \text{ and } g_1 = \frac{\exp(\bar{z}_c)}{1 + \exp(\bar{z}_c)}.$$

Assume that log wealth-to-consumption ratio is of the form:

$$z_{c,t} = A_{c,0} + A_{c,1}x_t + A_{c,2}\sigma_t^2. \quad (59)$$

Let  $\mu_c = \mu_d - \mu_\pi$ . Then we can write:

$$\begin{aligned} E_t \left[ m_{t+1} + \tilde{R}_{t+1}^c \right] &= \zeta \log(\delta) + \left( \zeta - \frac{\zeta}{\psi} \right) \left( \mu_c + \frac{1}{\gamma} x_t \right) + \zeta g_0 + \zeta (g_1 - 1) A_{c,0} + \zeta (g_1 \rho - 1) A_{c,1} x_t \\ &\quad + \zeta (g_1 \omega - 1) A_{c,2} \sigma_t^2 + \zeta g_1 (1 - \omega) A_{c,2} \sigma_c^2, \\ \text{var}_t \left( m_{t+1} + \tilde{R}_{t+1}^c \right) &= \zeta^2 \left( 1 - \frac{1}{\psi} \right)^2 \sigma_t^2 + \zeta^2 (g_1 A_{c,1} \sigma_x)^2 + \zeta^2 (g_1 A_{c,2} \sigma_c)^2. \end{aligned} \quad (60)$$

Based on the condition  $E_t[\exp(m_{t+1} + \tilde{R}_{t+1}^c)] = 1$ , we can solve for  $A_{c,0}$ ,  $A_{c,1}$ , and  $A_{c,2}$  as:

$$\begin{aligned} A_{c,0} &= \frac{\log(\delta) + (1 - \frac{1}{\psi})(\mu_d - \mu_\pi) + g_0 + g_1 A_{c,2}(1 - \omega)\sigma_c^2 + \frac{1}{2}\zeta g_1^2 (A_{c,1}^2 \sigma_x^2 + A_{c,2}^2 \sigma_c^2)}{1 - g_1}, \\ A_{c,1} &= \frac{\left(1 - \frac{1}{\psi}\right) \frac{1}{\gamma}}{1 - g_1 \rho}, \quad A_{c,2} = \frac{\zeta \left(1 - \frac{1}{\psi}\right)^2}{2(1 - g_1 \omega)}. \end{aligned} \quad (61)$$

Next, let  $z_{d,t}$  be log price-to-dividend ratio of the stock index and  $\tilde{R}_{t+1}$  be log stock index real return. Then, by first order Taylor series approximation, we can write:

$$\tilde{R}_{t+1} = \kappa_0 + \kappa_1 z_{d,t+1} - z_{d,t} + \Delta \tilde{d}_{t+1}. \quad (62)$$

where  $\Delta \tilde{d}_{t+1}$  is real dividend growth rate.

Assume that log price-to-dividend ratio is of the form:

$$z_{d,t} = A_{d,0} + A_{d,1}x_t + A_{d,2}\sigma_t^2 + A_{d,3}(q_t - \mu_q) + A_{d,4}(\Delta\pi_t - \mu_\pi). \quad (63)$$

Then note that:

$$\begin{aligned}
E_t \left[ m_{t+1} + \tilde{R}_{t+1} \right] &= \zeta \log(\delta) + (\zeta - 1)(g_1 - 1)A_{c,0} + (\zeta - 1)(g_1\rho - 1)A_{c,1}x_t + (\zeta - 1)(g_1\omega - 1)A_{c,2}\sigma_t^2 \\
&\quad + (\zeta - 1)g_1(1 - \omega)A_{d,2}\sigma_c^2 + \left( \zeta - \frac{\zeta}{\psi} - 1 \right) \left( \mu_c + \frac{1}{\gamma}x_t \right) + (\zeta - 1)g_0 + \kappa_0 + (\kappa_1 - 1)A_{d,0} \\
&\quad + (\kappa_1\rho - 1)A_{d,1}x_t + (\kappa_1\omega - 1)A_{d,2}\sigma_t^2 + \kappa_1(1 - \omega)A_{d,2}\sigma_c^2 + (\kappa_1\theta - 1)A_{d,3}(q_t - \mu_q) \\
&\quad + (\kappa_1\eta - 1)A_{d,4}(\Delta\pi_t - \mu_\pi) + \mu_c + x_t + \phi(q_t - \mu_q) - \eta(\Delta\pi_t - \mu_\pi), \\
\text{var}_t \left( m_{t+1} + \tilde{R}_{t+1} \right) &= \left( \zeta - 1 - \frac{\zeta}{\psi} \right)^2 \sigma_t^2 + \sigma_d^2 + ((\zeta - 1)g_1A_{c,1} + \kappa_1A_{d,1})^2 \sigma_x^2 \\
&\quad + ((\zeta - 1)g_1A_{c,2} + \kappa_1A_{d,2})^2 \sigma_c^2 + (\kappa_1A_{d,3})^2 \sigma_q^2 \\
&\quad + (\kappa_1A_{d,4} - 1)^2 \sigma_\pi^2 + 2(\kappa_1A_{d,3})\sigma_d\sigma_q\lambda_{dq}. \tag{64}
\end{aligned}$$

Based on the condition  $E_t[\exp(m_{t+1} + \tilde{R}_{t+1})] = 1$ , we can solve for  $A_{d,0}$ ,  $A_{d,1}$ ,  $A_{d,2}$ , and  $A_{d,3}$  as:

$$\begin{aligned}
A_{d,0} &= \frac{\left( \begin{aligned} &\zeta \log(\delta) + (\zeta - 1)g_0 + (\zeta - 1)(g_1 - 1)A_{c,0} + ((\zeta - 1)g_1A_{c,2} + \kappa_1A_{d,2})(1 - \omega)\sigma_c^2 \\ &\quad + \left( \zeta - \frac{\zeta}{\psi} - 1 \right) \mu_c + \kappa_0 + \mu_c \\ &\quad + \frac{1}{2}((\zeta - 1)g_1A_{c,1} + \kappa_1A_{d,1})^2 \sigma_x^2 + \frac{1}{2}((\zeta - 1)g_1A_{c,2} + \kappa_1A_{d,2})^2 \sigma_c^2 + \frac{1}{2}\sigma_d^2 \\ &\quad + \frac{1}{2}(\kappa_1A_{d,3})^2 \sigma_q^2 + (\kappa_1A_{d,4})\sigma_d\sigma_q\lambda_{dq} \end{aligned} \right)}{1 - \kappa_1}, \\
A_{d,1} &= \frac{\left( \zeta - 1 - \frac{\zeta}{\psi} \right) \frac{1}{\gamma} + (\zeta - 1)(g_1\rho - 1)A_{c,1} + 1}{1 - \kappa_1\rho}, \\
A_{d,2} &= \frac{(\zeta - 1)(g_1\omega - 1)A_{c,2} + \frac{1}{2} \left( \zeta - 1 - \frac{\zeta}{\psi} \right)^2}{1 - \kappa_1\omega}, \quad A_{d,3} = \frac{\phi}{1 - \kappa_1\theta}, \quad A_{d,4} = \frac{-\eta}{1 - \kappa_1\eta} \tag{65}
\end{aligned}$$

Substituting the expression for  $z_{d,t}$  into  $\tilde{R}_{t+1} = \kappa_0 + \kappa_1 z_{d,t+1} - z_{d,t} + \Delta \tilde{d}_{t+1}$  leads:

$$E_t[\tilde{R}_{t+1}] = A_{r,0} + A_{r,1}x_t + A_{r,2}\sigma_t^2 + A_{r,3}(q_t - \mu_q) + A_{r,4}(\Delta\pi_t - \mu_\pi), \tag{66}$$

where:

$$\begin{aligned}
A_{r,0} &= \kappa_0 - (1 - \kappa_1)A_{d,0} + (\mu_c + \mu_\pi) + \kappa_1(1 - \omega)A_{d,2}\sigma_c^2, \quad A_{r,1} = 1 - (1 - \kappa_1\rho)A_{d,1}, \\
A_{r,2} &= -(1 - \kappa_1\omega)A_{d,2}, \quad A_{r,3} = \phi - (1 - \kappa_1\theta)A_{d,3}, \quad A_{r,4} = -\eta - (1 - \kappa_1\eta)A_{d,4}. \tag{67}
\end{aligned}$$

Expected real return over the next  $\tau$  period is:

$$\begin{aligned} \sum_{s=0}^{\tau-1} \tilde{R}_{t+s+1} = & \tau A_{r,0} + \left( \sum_{s=0}^{\tau-1} A_{r,1} \rho^s \right) x_t + \left( \sum_{s=0}^{\tau-1} A_{r,2} \omega^s \right) \sigma_t^2 + \left( \sum_{s=0}^{\tau-1} A_{r,2} (1 - \omega^s) \right) \sigma_c^2 \\ & + \left( \sum_{s=0}^{\tau-1} A_{r,3} \theta^s \right) (q_t - \mu_q) + \left( \sum_{s=0}^{\tau-1} A_{r,4} \eta^s \right) (\Delta \pi_t - \mu_\pi). \end{aligned} \quad (68)$$

For nominal returns, add expected inflation. Finally, the real risk free rate can be written as:

$$R_{f,t+1} = A_{f,0} + A_{f,1} x_t + A_{f,2} \sigma_t^2, \quad (69)$$

where:

$$\begin{aligned} A_{f,0} = & -\zeta \log(\delta) - (\zeta - 1 - \frac{\zeta}{\psi}) \mu_c - (\zeta - 1) (g_0 + (g_1 - 1) A_{c,0} + g_1 (1 - \omega) A_{c,2} \sigma_c^2) \\ & - \frac{1}{2} ((\zeta - 1) g_1 A_{c,1})^2 \sigma_x^2 - \frac{1}{2} ((\zeta - 1) g_1 A_{c,2} \sigma_c)^2. \\ A_{f,1} = & -(\zeta - 1 - \frac{\zeta}{\psi}) \frac{1}{\gamma} + (\zeta - 1) (1 - g_1 \rho) A_{c,1}, \quad A_{f,2} = (\zeta - 1) (1 - g_1 \omega) A_{c,2} - \frac{1}{2} (\zeta - 1 - \frac{\zeta}{\psi})^2. \end{aligned} \quad (70)$$

#### A.4.2 The Alternative Model

We derive price-to-dividend ratios and expected returns implied by our long-run risk model, which features dividend dynamics in (3), consumption dynamics in (40), and investors preferences in (25). The difference between the alternative model from the default model is that, while the default model assumes consumption growth rates to have stochastic volatility, the alternative model assumes consumption volatility to be constant. Instead, the alternative model assumes that the correlation between shocks to consumption and shocks to dividend and earnings growth rates are time varying. Nevertheless, our derivation closely follows the steps in Bansal and Yaron (2004). The log stochastic discount factor is given as:

$$m_{t+1} = \zeta \log(\delta) - \frac{\zeta}{\psi} \Delta \tilde{c}_{t+1} + (\zeta - 1) \tilde{R}_{t+1}^c. \quad (71)$$

Let  $z_{c,t}$  be the log wealth-to-consumption ratio, by first order Taylor series approximation, log real return of the representative agent's wealth portfolio can be written as:

$$\tilde{R}_{t+1}^c = g_0 + g_1 z_{c,t+1} - z_{c,t} + \Delta \tilde{c}_{t+1}. \quad (72)$$

The log-linearizing constants are:

$$g_0 = \log(1 + \exp(\bar{z}_c)) - g_1(\bar{z}_c) \text{ and } g_1 = \frac{\exp(\bar{z}_c)}{1 + \exp(\bar{z}_c)}.$$

Assume that log wealth-to-consumption ratio is of the form:

$$z_{c,t} = A_{c,0} + A_{c,1}x_t. \quad (73)$$

Let  $\mu_c = \mu_d - \mu_\pi$ . Then we can write:

$$\begin{aligned} E_t \left[ m_{t+1} + \tilde{R}_{t+1}^c \right] &= \zeta \log(\delta) + \left( \zeta - \frac{\zeta}{\psi} \right) \left( \mu_c + \frac{1}{\gamma} x_t \right) + \zeta g_0 + \zeta (g_1 - 1) A_{c,0} + \zeta (g_1 \rho - 1) A_{c,1} x_t, \\ \text{var}_t \left( m_{t+1} + \tilde{R}_{t+1}^c \right) &= \zeta^2 \left( 1 - \frac{1}{\psi} \right)^2 \sigma_c^2 + \zeta^2 (g_1 A_{c,1})^2 \sigma_x^2. \end{aligned} \quad (74)$$

Based on  $E_t[\exp(m_{t+1} + \tilde{R}_{t+1}^c)] = 1$ , we can solve for coefficients  $A_{c,0}$ ,  $A_{c,1}$ , and  $A_{c,2}$  as:

$$\begin{aligned} A_{c,0} &= \frac{\log(\delta) + (1 - \frac{1}{\psi})\mu_c + g_0 + \frac{1}{2}\zeta(1 - \frac{1}{\psi})^2\sigma_c^2 + \frac{1}{2}\zeta(g_1 A_{c,1})^2\sigma_x^2}{1 - g_1}, \\ A_{c,1} &= \frac{\left(1 - \frac{1}{\psi}\right) \frac{1}{\gamma}}{1 - g_1 \rho}. \end{aligned} \quad (75)$$

Next, let  $z_{d,t}$  be log price-to-dividend ratio of the stock index and  $\tilde{R}_{t+1}$  be log real stock index return. Then, by first order Taylor series approximation, we can write:

$$\tilde{R}_{t+1} = \kappa_0 + \kappa_1 z_{d,t+1} - z_{d,t} + \Delta \tilde{d}_{t+1}. \quad (76)$$

where  $\Delta \tilde{d}_{t+1}$  is real dividend growth rate.

Assume that log price-to-dividend ratio is of the form:

$$z_{d,t} = A_{d,0} + A_{d,1}x_t + A_{d,2}\lambda_t + A_{d,3}(q_t - \mu_q) + A_{d,4}(\Delta\pi_t - \mu_\pi). \quad (77)$$

Then note that:

$$\begin{aligned}
E_t [m_{t+1} + \tilde{R}_{t+1}] &= \zeta \log(\delta) + (\zeta - 1)(g_1 - 1)A_{c,0} + (\zeta - 1)(g_1\rho - 1)A_{c,1}x_t \\
&\quad + \left(\zeta - \frac{\zeta}{\psi} - 1\right)(\mu_c + \gamma x_t) + (\zeta - 1)g_0 + \kappa_0 + (\kappa_1 - 1)A_{d,0} + (\kappa_1\rho - 1)A_{d,1}x_t \\
&\quad + (\kappa_1\omega - 1)A_{d,2}\lambda_t + (\kappa_1\theta - 1)A_{d,3}(q_t - \mu_q) + (\kappa_1\eta - 1)A_{d,4}(\Delta\pi_t - \mu_\pi) \\
&\quad + \mu_c + x_t + \phi(q_t - \mu_q) - \eta(\Delta\pi_t - \mu_\pi). \\
\text{var}_t (m_{t+1} + \tilde{R}_{t+1}) &= \left(\zeta - 1 - \frac{\zeta}{\psi}\right)^2 \sigma_c^2 + \sigma_d^2 + ((\zeta - 1)g_1A_{c,1} + \kappa_1A_{d,1})^2 \sigma_x^2 \\
&\quad + (\kappa_1A_{d,2})^2 \sigma_\lambda^2 + (\kappa_1A_{d,3})^2 \sigma_q^2 + (\kappa_1A_{d,4} - 1)^2 \sigma_\pi^2 \\
&\quad + 2\left(\zeta - 1 - \frac{\zeta}{\psi}\right) \sigma_c \sigma_d \lambda_t + 2(\kappa_1A_{d,3}) \sigma_d \sigma_q \lambda_{dq} \\
&\quad + 2\left(\zeta - 1 - \frac{\zeta}{\psi}\right) (\kappa_1A_{d,3}) \left(\sqrt{\sigma_d^2 + \sigma_q^2 + 2\sigma_d \sigma_q \lambda_{dq}} - \sigma_d\right) \sigma_c \lambda_t. \tag{78}
\end{aligned}$$

Based on  $E_t[\exp(m_{t+1} + \tilde{R}_{t+1})] = 1$ , we can solve for  $A_{d,0}$ ,  $A_{d,1}$ ,  $A_{d,2}$ ,  $A_{d,3}$ , and  $A_{d,4}$  as:

$$\begin{aligned}
A_{d,0} &= \frac{\left( \begin{aligned} &\zeta \log(\delta) + (\zeta - 1)g_0 + (\zeta - 1)(g_1 - 1)A_{c,0} \\ &+ \left(\zeta - \frac{\zeta}{\psi} - 1\right) \mu_c + \kappa_0 + \mu_c + \frac{1}{2}\sigma_d^2 + \frac{1}{2}((\zeta - 1)g_1A_{c,1} + \kappa_1A_{d,1})^2 \sigma_x^2 \\ &+ \frac{1}{2}(\kappa_1A_{d,2})^2 \sigma_\lambda^2 + \frac{1}{2}(\kappa_1A_{d,3})^2 \sigma_q^2 + \frac{1}{2}(\kappa_1A_{d,4} - 1)^2 \sigma_\pi^2 \\ &+ \frac{1}{2}\left(\zeta - \zeta - \frac{\zeta}{\psi}\right)^2 \sigma_c^2 + (\kappa_1A_{d,3})\sigma_d \sigma_q \lambda_{dq} \end{aligned} \right)}{1 - \kappa_1}, \\
A_{d,1} &= \frac{\left(\zeta - 1 - \frac{\zeta}{\psi}\right) \frac{1}{\gamma} + (\zeta - 1)(g_1\rho - 1)A_{c,1} + 1}{1 - \kappa_1\rho}, \\
A_{d,2} &= \frac{\left(\zeta - 1 - \frac{\zeta}{\psi}\right) \left((\kappa_1A_{d,3}) \left(\sqrt{\sigma_d^2 + \sigma_q^2 + 2\sigma_d \sigma_q \lambda_{dq}} - \sigma_d\right) + \sigma_d\right) \sigma_c}{1 - \kappa_1\omega}, \\
A_{d,3} &= \frac{\phi}{1 - \kappa_1\theta}, \quad A_{d,4} = \frac{-\eta}{1 - \kappa_1\eta}. \tag{79}
\end{aligned}$$

Substituting the expression for  $z_{d,t}$  into  $\tilde{R}_{t+1} = \kappa_0 + \kappa_1 z_{d,t+1} - z_{d,t} + \Delta\tilde{d}_{t+1}$  leads:

$$E_t[\tilde{R}_{t+1}] = A_{r,0} + A_{r,1}x_t + A_{r,2}\lambda_t + A_{r,3}(q_t - \mu_q) + A_{r,4}(\Delta\pi_t - \mu_\pi), \tag{80}$$

where:

$$\begin{aligned}
A_{r,0} &= \kappa_0 - (1 - \kappa_1)A_{d,0} + \mu_d, \quad A_{r,1} = 1 - (1 - \kappa_1\rho)A_{d,1}, \\
A_{r,2} &= -(1 - \kappa_1\omega)A_{d,2}, \quad A_{r,3} = \phi - (1 - \kappa_1\theta)A_{d,3}, \quad A_{r,4} = -\eta - (1 - \kappa_1\eta)A_{d,4}. \tag{81}
\end{aligned}$$

Expected real return over the next  $\tau$  period is:

$$\begin{aligned} \sum_{s=0}^{\tau-1} \tilde{R}_{t+s+1} = & \tau A_{r,0} + \left( \sum_{s=0}^{\tau-1} A_{r,1} \rho^s \right) x_t + \left( \sum_{s=0}^{\tau-1} A_{r,2} \omega^s \right) \lambda_t + \left( \sum_{s=0}^{\tau-1} A_{r,3} \theta^s \right) (q_t - \mu_q) \\ & + \left( \sum_{s=0}^{\tau-1} A_{r,4} \eta^s \right) (\Delta \pi_t - \mu_\pi) \end{aligned} \quad (82)$$

For nominal returns, add expected inflation. Finally, real risk free rates can be written as:

$$\tilde{R}_{f,t+1} = A_{f,0} + A_{f,1} x_t, \quad (83)$$

where:

$$\begin{aligned} A_{f,0} = & -\zeta \log(\delta) - \left( \zeta - 1 - \frac{\zeta}{\psi} \right) \mu_c - (\zeta - 1) (g_0 + (g_1 - 1) A_{c,0}) - \frac{1}{2} \left( \zeta - 1 - \frac{\zeta}{\psi} \right)^2 \sigma_c^2 \\ & - \frac{1}{2} \left( (\zeta - 1) g_1 A_{c,1} \right)^2 \sigma_x^2. \\ A_{f,1} = & - \left( \zeta - 1 - \frac{\zeta}{\psi} \right) \frac{1}{\gamma} + (\zeta - 1) (1 - g_1 \rho) A_{c,1}. \end{aligned} \quad (84)$$