Dividend Dynamics, Learning, and Expected Stock Index Returns

Ravi Jagannathan
Northwestern University and NBER

Binying Liu
Northwestern University

September 30, 2015

Abstract

We develop a model for dividend dynamics and allow investors to learn about model parameters over time. The model predicts 31.3% of the variation in annual dividend growth rates during 1976-2013. We show that when investors' beliefs about the persistence of dividend growth rates increase, dividend-to-price ratios increase, and short-horizon stock returns decrease after controlling for dividend-to-price ratios. These findings support investors' preferences for early resolution of uncertainty. We embed learning about dividend dynamics into an equilibrium asset pricing model. The model predicts 22.9% of the variation in annual stock returns. Learning accounts for over forty-percent of the 22.9%.

Disclosure Statement - Ravi Jagannathan

I have nothing to disclose. I received no payments from interested parties and received no outside research funding.

Disclosure Statement - Binying Liu

I have nothing to disclose. I received no payments from interested parties and received no outside research funding.

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 2013, the S&P500 earned 62 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 risk in equities that require such a large premium and why the risk free rate is so low. A common feature in most of these models is that the 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 the equity premium.[1] 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, they have full knowledge of the parameters describing the economy. For example, these models assume that investors 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.[2] In this paper, we provide empirical evidence that investors learn and that changes in investors' beliefs about the parameters describing the economy is reflected in stock index prices and returns. Further, we show that the way stock index prices and returns covary with investors' beliefs provides us insight into investors' preferences.

The focus of this paper is on learning about dividend dynamics. To study how learning

[1] 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 Koijen (2010), Kelly and Pruitt (2013), van Binsbergen, Hueskes, Koijen, and Vrugt (2013), Li, Ng, and Swaminathan (2013), and Da, Jagannathan, and Shen (2014).

[2] See, among others, Ju and Miao (2012), Croce, Lettau, and Ludvigson (2014), Johannes, Lochstoer, and Mou (2015), and Johannes, Lochstoer, and Collin-Dufresne (2015).

about dividend dynamics affect stock index prices and expected returns, we need a realistic dividend model that is able to capture how investors form expectations about future dividends. Inspired by Campbell and Shiller (1988b), we propose a model for dividend growth rates that incorporates information in aggregate corporate earnings into the latent variable model of van Binsbergen and Koijen (2010). Our model successfully captures serial correlations in annual dividend growth rates up to 5 years. Further, our model explains 55.1 percent of the variation in annual dividend growth rates between 1946 and 2013 in-sample and predicts 31.3 percent of the variation in annual dividend growth rates between 1976 and 2013 out-of-sample. We reject the Null hypothesis that expected dividend growth rates are constant at the 99 percent confidence level.

We document that uncertainties about parameters in our dividend model, especially the parameter governing the persistence of the latent variable, are high and resolve slowly. That is, these uncertainties remain substantial even at the end of our 68 years data sample, suggesting that learning about dividend dynamics is difficult. Further, when our 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 over time. We 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 the stock index using our model. We define stock yield as the discount rate that equates the present value of expected future dividends to the current price of the stock index. From the log-linearized present value relationship of Campbell and Shiller (1988a), we write stock yields as functions of price-to-dividend ratios and long run dividend growth expectations, computed assuming that investors learn about dividend dynamics using our model. We show that, between 1976 and 2013, these stock yields explain 15.2 percent of the variation in stock index returns over the next year. In comparison, stock yields, computed assuming that expected dividend growth rates are constant, explain only 10.2 percent of the same variation. We can attribute this improvement in forecasting performance from 10.2 percent to 15.2 percent to our modeling of learning about dividend dynamics.

We argue that how stock index prices and returns respond to changes in investors' beliefs about dividend dynamics can also provide us insight into investors' preferences, and more specifically, their preferences for the timing of resolution of uncertainty. That is, depending on whether investors prefer early or late resolution of uncertainty, changes in investors' beliefs about the persistence of dividend growth rates have different effects on discount rates. We show that, when investors' beliefs about the persistence of dividend growth rates increase, price-to-dividend ratios decrease, stock yields increase, and stock index returns over the short-horizon decrease after controlling for either price-to-dividend ratios or stock yields. We argue that these findings lend support to investors' preferences for early resolution of uncertainty.

We embed our dividend model into an dynamic equilibrium asset pricing model that features Epstein and Zin (1989) preferences, which capture preferences for the timing of resolution of uncertainty, and consumption dynamics from the long-run risk model of Bansal and Yaron (2004). We refer to this model as our long-run risk model. We find that, between 1976 and 2013, expected returns derived from our long-run risk model, assuming that investors learn about the parameters governing the dividend process, predict 22.9 percent of the variation in annual stock index returns. Learning accounts for forty-percent of the 22.9 percent. We decompose the variation in price-to-dividend ratios and find that over forty-percent of the variation is due to investors' learning about dividend dynamics, and the rest is due to discount rate variation.

We follow Cogley and Sargent (2008), Piazzesi and Schneider (2010), and Johannes, Lochstoer, and Mou (2015), 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 asset pricing 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.[3] Therefore, we adopt the anticipated utility approach as the more convenient alternative.

The rest of this paper is organized as follows. In Section 1, we introduce our dividend model and evaluates its performance in capturing dividend dynamics. In Section 2, we

[3]Johannes, Lochstoer, and Collin-Dufresne (2015) provide the theoretical foundation for studying uncertainties about model parameters as priced risk factors.

discuss how learning about dividend dynamics affects expectations of future dividends. In Section 3, we show that learning about dividend dynamics is reflected in the prices and returns of the stock index. In Section 4, we argue that the way discount rates covary with investors' beliefs about the persistence of dividend growth rates supports investors' preferences for early resolution of uncertainty. In Section 5, we embed our dividend model into an equilibrium asset pricing model to quantify how much learning about dividend dynamics contributes to the variations in price-to-dividend ratios and future stock index returns. In Section 6, we conclude.

1 The Dividend Model

In this section, we present a model for dividend growth rates that extends the latent variable model of van Binsbergen and Koijen (2010) by incorporating information in aggregate corporate earnings. The inclusion of earnings information in explaining dividend dynamics is inspired by Campbell and 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.

Let d_t be log dividend and $\Delta d_{t+1} = d_{t+1} - d_t$ be its growth rate. The latent variable model of van Binsbergen and Koijen (2010) is described by the following system of equations:

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

Following van Binsbergen and Koijen (2010), our focus is on modeling the nominal dividend process.[4] To show that our findings are robust, we also provide results on modeling the real dividend process in the Appendix. Time-t is defined in years to control for potential seasonality in dividend payments. In this model, expected dividend growth rates follow a stationary AR[1] process and are functions of the latent variable x_t, the

[4]As shown in Boudoukh, Michaely, Richardson, and Roberts (2007), equity issuance and repurchase tend to be sporadic and random compared to cash dividends. For this reason, we focus on modeling the cash dividend process and treat equity issuance and repurchase as unpredictable.

unconditional mean μ_d of dividend growth rates, and the persistence coefficient ρ of x_t:

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

To introduce earnings information into this model, first define p_t as log price of the stock index, e_t as log earnings, π_t as log consumer price index, and, following Campbell and Shiller (1988b), run the following vector-autoregression for annual dividend growth rates, log price-to-dividend ratios, and CAPE ratios:

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

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

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

We report estimates of b_{10}, b_{11}, b_{12}, and b_{13} from (3), based on data between 1946 and 2013, in the first four columns of Table 1.[5] Consistent with Campbell and Shiller (1988b), we find 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. Interestingly, we see from 1 that $b_{12} + b_{13} = 0$ cannot be statistically rejected. For this reason, we restrict $b_{13} = -b_{12}$ and re-estimate annual dividend growth rates as:

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

We note that the stock index price does not appear in (5). We report estimated coefficients

[5] Throughout this paper, we report results based on overlapping monthly data. That is, 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.

from (5) in the last three columns of Table 1. Results show that the β_2 estimate is highly statistically significant, suggesting that expected dividend growth rates respond to the log ratios between historical earnings and dividends. Intuitively, high earnings relative to dividends imply that firms have been retaining earnings in the past and so are expected to pay more dividends in the future.

b_{10}	b_{11}	b_{12}	b_{13}	β_0	β_1	β_2
-0.058	0.442***	0.103**	-0.096**	-0.033	0.434***	0.098**
(0.066)	(0.118)	(0.042)	(0.041)	(0.029)	(0.117)	(0.041)

Table 1: **Campbell and Shiller (1988b) Betas for Predicting Dividend Growth Rates**: This table reports coefficients from estimating dividend growth rates using (3) and (5), based on data between 1946 and 2013. 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 earnings contain information about future dividends. Let $\Delta e_{t+1} = e_{t+1} - e_t$ be log earnings growth rate and $q_t = e_t - d_t$ be log earnings-to-dividend ratio, our dividend model can be described by the following system of equations:

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

In our model, dividend growth rates are linear combinations of four components. First, as in van Binsbergen and Koijen (2010), they consist of the latent variable x_t, which follows a stationary AR[1] process. Second, they are affected by fluctuations in contemporaneous earnings growth rates. That is, we expect firms to pay more dividends if their earnings over the same period are high. Third, they are affected by changes in past earnings-to-dividend ratios. That is, we expect firms to pay more dividends if they retained more earnings in the past. Fourth, they consist of white noises $\epsilon_{d,t+1}$. For convenience, we model earnings-to-dividend ratios as an AR[1] process, and assuming that it is stationary implies that dividend and earnings growth rates have the same unconditional mean μ_d. We note that, substituting the third equation into the first equation of (6), we can re-write the first equation of (6) as:

$$\Delta d_{t+1} = \frac{1}{1-\phi}x_t + \frac{\varphi-(1-\theta)\phi}{1-\phi}(q_t-\mu_q) + \frac{\phi}{1-\phi}\sigma_q\epsilon_{q,t+1} + \frac{1}{1-\phi}\sigma_d\epsilon_{d,t+1}. \quad (7)$$

So we can solve for expected dividend growth rates in our model as:

$$E_t[\Delta d_{t+s+1}] = \mu_d + \frac{\rho^s}{1-\phi}x_t + \frac{\theta^s(\varphi-(1-\theta)\phi)}{1-\phi}(q_t-\mu_q),\ \forall s \geq 0. \quad (8)$$

Thus, aside from the two state variables, expected dividend growth rates are functions of the unconditional means μ_d and μ_q of dividend growth rates and earnings-to-dividend ratios, the persistence ρ and θ of the latent variable x_t and earnings-to-dividend ratios, and coefficients ϕ and φ that connect earnings information to dividend dynamics. We note that earnings dynamics is not modeled explicitly in (6). However, we can solve for earnings growth rates from the processes of dividend growth rates and earnings-to-dividend ratios:

$$\Delta e_{t+1} = \mu_d + \frac{1}{1-\phi}x_t + \frac{\varphi+\theta-1}{1-\phi}(q_t-\mu_q) + \frac{1}{1-\phi}\sigma_d\epsilon_{d,t+1} + \frac{1}{1-\phi}\sigma_q\epsilon_{q,t+1}. \quad (9)$$

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 using Kalman filters (Hamilton (1994)) and estimate model parameters,

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

based on maximum-likelihood. See the Appendix for details. Table 2 reports model parameter estimates based on data between 1946 and 2013. Standard errors are based on bootstrap simulation. Previous 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. For this reason, we restrict our data sample to the post-war era. Consistent with our intuition, both ϕ and φ that connect earnings information to dividend dynamics are estimated to be positive and highly statistically significant. That is, high contemporaneous earnings growth rates imply high dividend growth rates, and high past earnings-to-dividend ratios imply high dividend growth rates. The annual persistence of earnings-to-dividend ratios is estimated to be 0.281. The latent variable x_t is estimated to be more persistent at 0.528. In summary, there is a moderate level of persistence in dividend growth rates between 1946 and 2013 based on estimates from our model.

μ_d	ϕ	φ	σ_d	ρ	σ_x
0.059	0.079	0.184	0.017	0.528	0.041
(0.015)	(0.018)	(0.028)	(0.013)	(0.160)	(0.009)
μ_q	θ	σ_q	λ_{dx}	λ_{dq}	λ_{xq}
0.713	0.281	0.280	-0.032	-0.157	0.024
(0.047)	(0.116)	(0.027)	(0.131)	(0.028)	(0.124)

Table 2: **Dividend Model Parameters**: This table reports estimated parameters from our dividend model, based on data between 1946 and 2013. Bootstrap simulated standard errors are reported in parentice. Simulation is based on 100,000 iterations.

In Table 3, we report serial correlations, up to 5 years, for annual dividend growth rates and dividend growth rate residuals, which we define as the difference between dividend growth rates and expected growth rates implied by our dividend model. We also report serial correlations for dividend growth rate residuals implied by either of the dividend models described in (1) and (3), which we refer to as the baseline models. We then provide the Ljung and Box (1978) Q-statistics for testing if dividend growth rates and growth rate residuals are serially correlated. We find that our dividend model is reasonably successful at matching serial correlations in annual dividend growth rates for up to 5 years. That is, our model's dividend growth rate residuals appear to be serially uncorrelated. In comparison, for the baseline models we find that their growth rate residuals are serially correlated at the 95 percent confidence level.

In the first column of Table 4, we report the goodness-of-fit for describing dividend growth rates using our model, based on data between 1946 and 2013. We find that our model explains 55.0 percent of the variation in annual dividend growth rates. To account for the fact that at least part of this fit comes from adding more parameters to existing models and is thus mechanical, we also report the Bayesian information criterion (BIC), which penalizes a model based on the number of free parameters in that model.[6] We report BIC statistics in the second column of Table 4. Results confirm that our model outperforms the baseline models in explaining the variation in dividend growth rates.

Another way to address the concern that our model overfits the data is to assess the model based on how it forecasts dividend growth rates out-of-sample. That is, instead of fitting the model based on the full data sample, we predict dividend growth rates at each point in time based on data available at the time. Forecasting performance is then evaluated using the out-of-sample R-square value as defined in Goyal and Welch (2008):

$$R^2 = 1 - \frac{\sum_{t=t_0}^{T-1} \left(\Delta d_{t+1} - E_t[\Delta d_{t+1}]\right)^2}{\sum_{t=t_0}^{T-1} \left(\Delta d_{t+1} - \hat{\mu}_d(t)\right)^2}, \tag{10}$$

where $\overline{\Delta d}_t$ is the average of dividend growth rates up to time-t:

$$\hat{\mu}_d(t) = \frac{1}{t} \sum_{s=0}^{t-1} \Delta d_{s+1}, \tag{11}$$

and we use time-t_0 to denote the end of the training period. 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 2013. 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.[7] To show that our findings are robust to this assumption, we repeat the main results of this paper in the Appendix, assuming that earnings information is known to investors with a lag of 6, 9 and 12 months. We assume that information about prices and dividends is known to investors in real time.[8] In the third and fourth columns of Table 4, we 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). Results show that our model predicts 31.3 percent of the variation in annual dividend growth rates, which is a significant improvement over the R-square values

[6] $BIC = \log(\hat{var}(\Delta d_{t+1} - E_t[\Delta d_{t+1}])) + m\frac{\log(T)}{T}$, where m is a model's number of parameters, excluding those in the variance-covariance matrix, and T is the number of observations.

[7] 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.

[8] Our results are also robust to assuming that dividend information is known with a 3 months lag.

of 18.5 percent and 13.5 percent from the baseline models. Interestingly, we note that imposing the restriction that $b_{12} + b_{13} = 0$ in (3) significantly improves the out-of-sample forecasting performance of the Campbell and Shiller (1988b) model, lending additional support for our decision to impose this restriction.

		$\Delta d_{t+1} - E_t[\Delta d_{t+1}]$		
	Δd_{t+1}	J&L	vB&K	C&S
Serial Correlation (Years)				
1	0.418	-0.027	0.123	0.156
2	-0.107	-0.128	-0.212	-0.197
3	-0.318	-0.036	-0.249	-0.224
4	-0.280	0.066	-0.153	-0.048
5	-0.139	0.198	-0.031	-0.240
Q-Statistics	32.49	5.263	12.36	12.96
	[0.000]	[0.385]	[0.030]	[0.024]

Table 3: **Serial Correlations in Dividend Growth Rates and Expected Growth Rates**: This table reports the 1, 2, 3, 4, and 5 years serial correlations for dividend growth rates and growth rate residuals implied by our dividend model (i.e. J&L), the dividend model in van Binsbergen and Koijen (2010) (i.e. vB&K), or the dividend model in Campbell and Shiller (1988b) (i.e. C&S), based on data between 1946 and 2013. Also reported are the Ljung-Box (1973) 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.

Although results in this section show that our model is successful in capturing the variation in dividend growth rates both in-sample and out-of-sample, we recognize that it inevitably simplifies the true process governing dividend dynamics. One can add additional lags of earnings-to-dividend ratios to the model.[9] Also, one can extend our model by allowing model parameters, such as the persistence ρ of the latent variable x_t or the standard deviation σ_x of shocks to x_t, to be time varying. However, the disadvantage of incorporating such extensions is that a more complicated model is also more difficult to estimate with precision in finite sample. For example, one way to assess whether accounting for the possibilities of time varying model parameters improves our model's out-of-sample forecasting performance is to estimate model parameters based on a rolling window, rather than an expanding window, of past data, so that observations from the distant past are not used to estimate model parameters. We provide this analysis in the

[9] For example, Campbell and Shiller (1988b) assume dividend growth rates are affected by earnings information with up to 10 years of lag.

Appendix. In summary, we find that our model's forecasting performance is not improved by estimating model parameters based on a rolling window of past data.

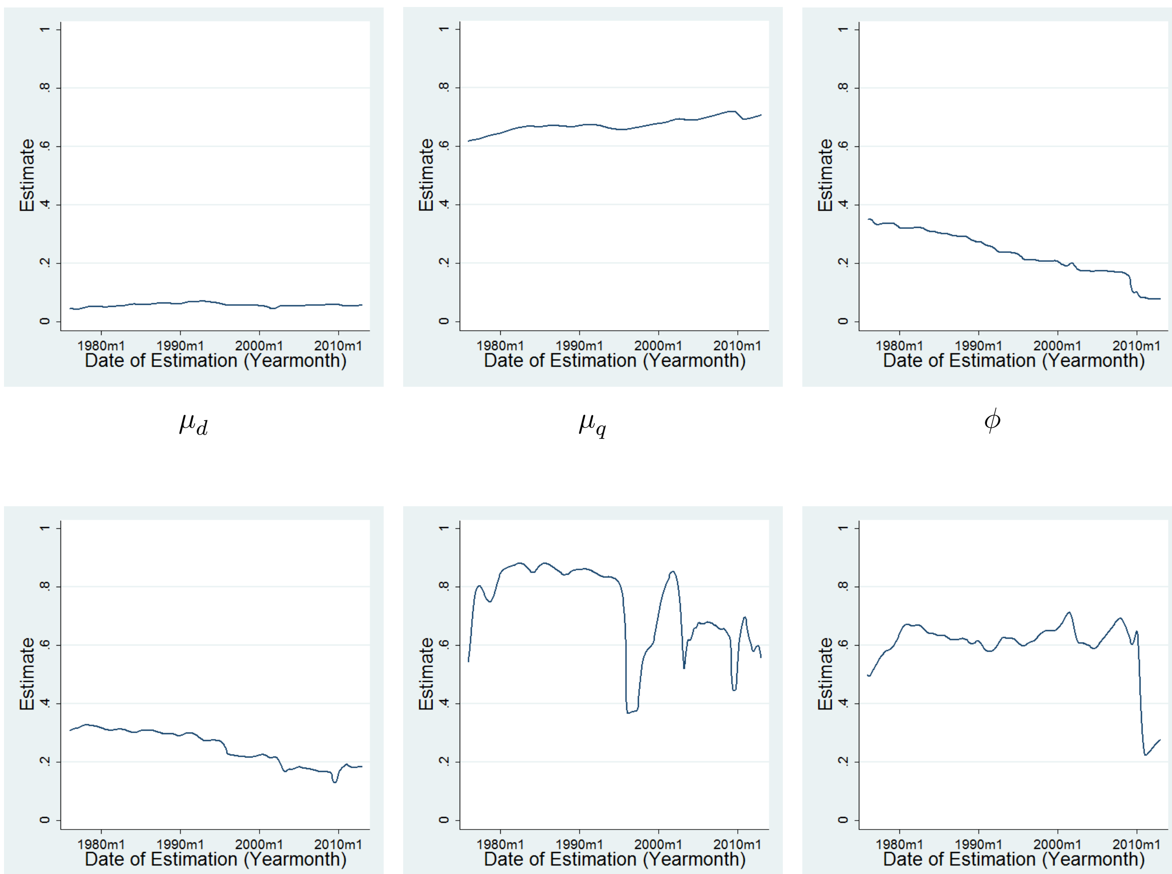
	In-Sample		Out-of-Sample	
	Goodness-of-Fit	BIC	R^2	p-value
J&L	0.551	-5.863	0.313	0.000
vB&K	0.176	-5.509	0.185	0.008
C&S	0.250	-5.683	0.135	0.025
C&S (Restricted)	0.248	-5.680	0.245	0.002

Table 4: **Dividend Growth Rates and Expected Growth Rates**. The first column of this table reports goodness-of-fit for describing dividend growth rates using our dividend model (i.e. J&L), the dividend model in van Binsbergen and Koijen (2010) (i.e. vB&K), the dividend model in Campbell and Shiller (1988b) (i.e. C&S), or its restricted version where we set $b_{12}+b_{13}=0$. The second column reports the Bayesian information criterion. The third and fourth columns report the 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). In-sample (out-of-sample) statistics are based on data between 1946 and 2013 (1976 and 2013).

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 2013, so they reflect investors' knowledge of dividend dynamics at the end of 2013. That is, if investors were to estimate our model in 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 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 learn about dividend dynamics using our model. We then show that learning can have significant asset pricing implications.

In Figure 1, we report estimates of the six model parameters in (8) that affect expected dividend growth rates, assuming that our model is estimated based on data up to time-τ, for τ between 1976 and 2013. There are several points we take away from Figure 1. First, there is a gradual upward drift in investors' beliefs about the unconditional mean μ_q of earnings-to-dividend 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 ϕ and φ that connect earnings information to dividend dynamics. This means that dividends have become more smoothed over time. Third, a sharp drop in investors' beliefs about the persistence θ of earnings-to-dividend 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.



μ_d μ_q ϕ

φ ρ θ

Figure 1: **Evolution of Model Parameters.** This figure plots estimates of the six dividend model parameters that affect expected dividend growth rates, assuming that these parameters are estimated based on data up to time-τ for τ between 1976 and 2013.

It is clear from Figure 1 that the persistence ρ of the latent variable x_t is the parameter hardest to learn and least stable over time. This observation is consistent with results reported in Table 2, which show that, of all model parameters, ρ is estimated with the highest standard error (i.e. 0.160). Investors' beliefs about ρ fluctuate significantly over the sample period, especially around three periods during which beliefs about ρ sharply drop. 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 term trend that sees a gradual decrease in investors' beliefs about ρ since early 1980s. For example, if we were to pick a random date between 1976 and 2013 and estimate our model based on data up to that date, on average we would have estimated a ρ of 0.734.[10] This would be significantly higher than the 0.528 reported in Table 2 that is 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 68 years of data, there are still significant uncertainties surrounding the estimates of some model parameters. For example, the 95 percent confidence interval for the persistence ρ of the latent variable x_t is between 0.214 and 0.842. The same confidence interval for the persistence θ of earnings-to-dividend ratios is between 0.054 and 0.508. To quantify the speed of learning, following Johannes, Lochstoer, and Mou (2015), for each of the six parameters that affect expected dividend growth rates, we construct a measure that is the inverse ratio between the bootstrap simulated standard error assuming that the parameter is estimated based on data up to 2013 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 2023). See the Appendix for details on bootstrap simulation. In other words, this ratio reports how much an estimated parameter's standard error would reduce if investors were to have 10 more years of data. So the closer this ratio is to 1, the more difficult it is for investors to learn about the parameter. In Table 5, we report this ratio for each of the six model parameters. Overall, 10 years of additional data would only decrease the standard errors of parameter estimates by between 5 and 8 percent. Further, consistent with results from Figure 1 and those reported in Table 2, we find that it is more difficult to learn about ρ than about any of the other five model parameters.

We show that learning about dividend dynamics can have significant asset pricing implications. Consider the log linearized present value relationship in Campbell and

[10] To establish a point of reference, Bansal and Yaron (2004) calibrate annualized persistence of expected dividend growth rate to be $0.975^{12} = 0.738$.

μ_d	μ_q	ϕ	φ	ρ	θ
0.924	0.924	0.926	0.928	0.951	0.920

Table 5: **Speed of Learning about Model Parameters**: This table reports the speed of learning for the six model parameters that affect expected dividend growth rates. Speed of learning is defined as the inverse ratio between the bootstrap simulated standard error assuming that the parameter is estimated based on data up to 2013 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 2023). Simulation is based on 100,000 iterations.

Shiller (1988a):

$$p_t - d_t = \frac{\kappa_0}{1-\kappa_1} + \sum_{s=0}^{\infty} \kappa_1^s \left(E_t[\Delta d_{t+s+1}] - E_t[r_{t+s+1}] \right), \tag{12}$$

where κ_0 and κ_1 are log-linearizing constants and r_{t+1} is the stock index's log return.[11] The expression is a mathematical identity that connects price-to-dividend ratios, expected dividend growth rates, and discount rates (i.e. expected returns). We define stock yield as the discount rate that equates the present value of expected future dividends to the current price of the stock index. That is, rearranging (12), we can write stock yield as:

$$\begin{aligned} sy_t &\equiv (1-\kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta r_{t+s+1}] \\ &= \kappa_0 - (1-\kappa_1)(p_t - d_t) + (1-\kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta d_{t+s+1}]. \end{aligned} \tag{13}$$

Define long run dividend growth expectation as:

$$\partial_t \equiv (1-\kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta d_{t+s+1}]. \tag{14}$$

Given that price-to-dividend ratios are observed, there is a one-to-one mapping between long run dividend growth expectations and stock yields. We note that long run dividend growth expectations are specific to the dividend model and its parameters. For example,

[11] 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 to match the data between 1946 and 2013. This gives $\kappa_0 = 0.059$ and $\kappa_1 = 0.970$.

using our dividend model, we can re-write (14) as:

$$\begin{aligned}\partial_t &= (1-\kappa_1)\sum_{s=0}^{\infty}\kappa_1^s\left(\mu_d + \frac{\rho^s}{1-\phi}x_t + \frac{\theta^s(\varphi-(1-\theta)\phi)}{1-\phi}(q_t-\mu_q)\right)\\ &= \mu_d + \frac{1-\kappa_1}{1-\phi}\left(\frac{1}{1-\kappa_1\rho}x_t + \frac{\varphi-(1-\theta)\phi}{1-\kappa_1\theta}(q_t-\mu_q)\right). \qquad (15)\end{aligned}$$

If a different dividend model is used, long run dividend growth expectations will also be different. For example, if we assume that dividend growth rates follow a white noise process centered around μ_d, we can re-write (14) 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 model 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 full knowledge of the parameters describing the dividend process, whereas the second case corresponds to investors having to learn about dividend dynamics. In Figure 2, we plot long run dividend growth expectations, computed using our model and assuming that investors either have to learn, or do not learn, about model parameters. We find that learning has a considerable effect on investors' long run dividend growth expectations, assuming that investors learn about dividend dynamics using our model.

In Figure 3, we plot stock yields, computed by substituting (15) into (13):

$$sy_t = \kappa_0 - (1-\kappa_1)(p_t-d_t) + \mu_d + \frac{1-\kappa_1}{1-\phi}\left(\frac{1}{1-\kappa_1\rho}x_t + \frac{\varphi-(1-\theta)\phi}{1-\kappa_1\theta}(q_t-\mu_q)\right). \qquad (16)$$

Dividend model parameters are either estimated once based on the full data sample or estimated at each point in time based on data available at the time. We also plot price-to-dividend ratios in Figure 3, and scale price-to-dividend ratios to allow for easy comparison to stock yields. We find that there is almost no noticeable difference between the time series of price-to-dividend ratios and stock yields, computed assuming that investors do not learn. 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. However, assuming that investors have to learn, we find significant differences between the time series of price-to-dividend ratios and stock yields.

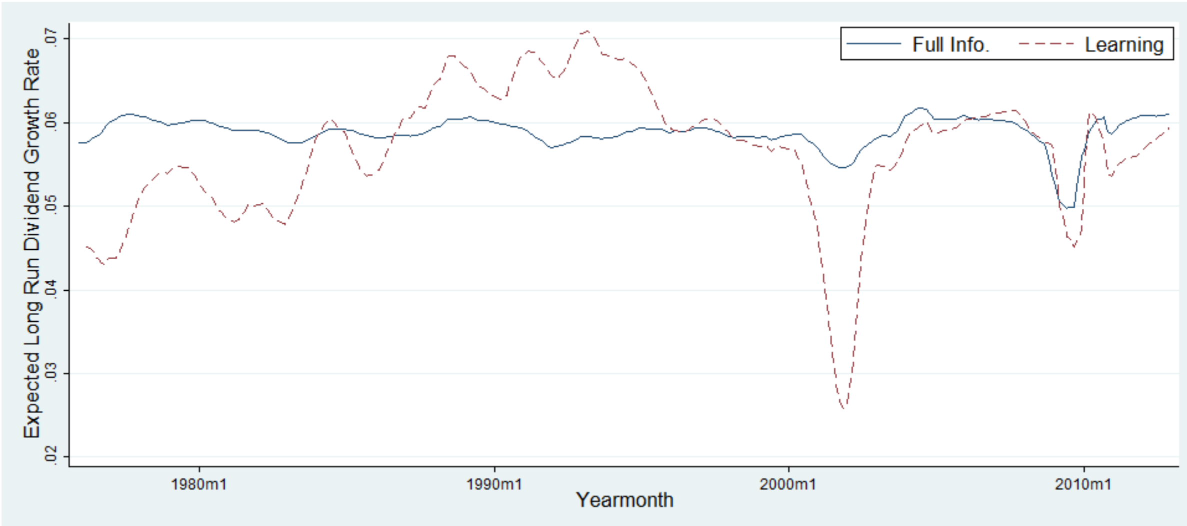


Figure 2: **Expected Long Run Dividend Growth Rates**. This figure plots long run dividend growth expectations, computed using our dividend model, for the period between 1976 and 2013. Dividend model parameters are estimated based on data since 1946. Under full information, model parameters are estimated once based on the full data sample. Under learning, those parameters are estimated at each point in time based on data available at the time.

3 Learning about Dividend Dynamics and Investor Behavior

In the previous section, we 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 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. In this section, we present evidence that supports this assumption. We show that stock yields, computed assuming that investors learn about dividend dynamics using our model (see (16)), predict future stock index returns. To establish a baseline, note that, if we assume dividend growth rates follow a white noise process centered around μ_d, stock yield can be simplified to:

$$sy_t = \kappa_0 - (1 - \kappa_1)(p_t - d_t) + \mu_d. \tag{17}$$

That is, under the white noise assumption, stock yields are just scaled price-to-dividend ratios. We regress stock index returns over the next year on price-to-dividend ratios, based on data between 1976 and 2013. We report regression statistics in the first column

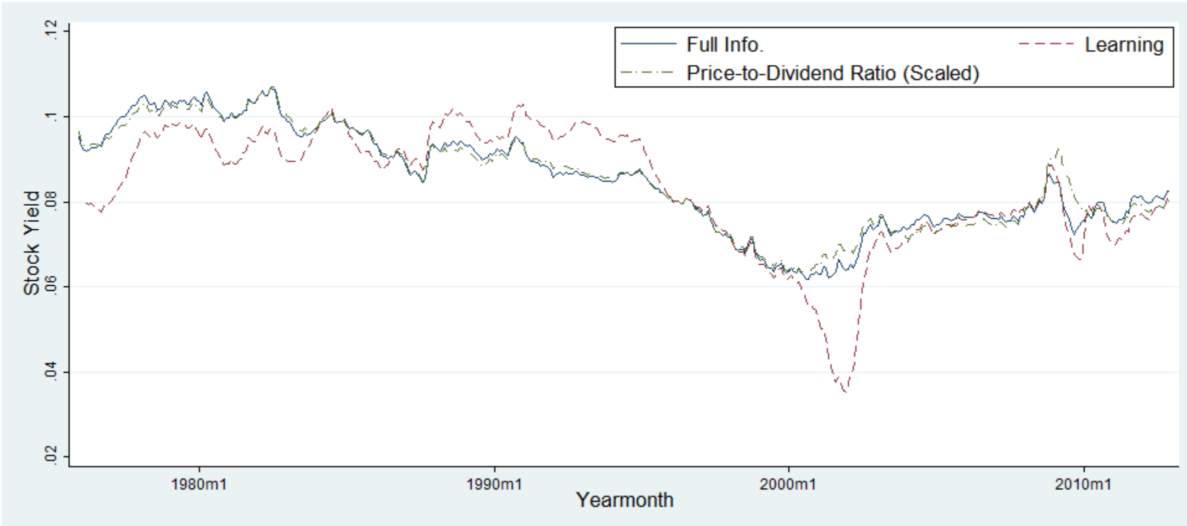


Figure 3: **Stock Yields.** This figure plots stock yields sy_t, computed using our dividend model, and log price-to-dividend ratios (scaled) for the period between 1976 and 2013. Dividend model parameters are estimated based on data since 1946. Under full information, model parameters are estimated once based on the full data sample. Under learning, those parameters are estimated at each point in time based on data available at the time.

of Table 6. Standard errors reported are Newey and West (1987) adjusted.[12] Results from Table 6 show that, between 1976 and 2013, price-to-dividend ratios explain 10.2 percent of the variation in stock index returns over the next year.

We then regress stock index returns over the next year on stock yields in (16), computed assuming that investors estimate model parameters at each point in time based on data available at the time. We report regression statistics in the second column of Table 6. The R-square value from this regression is 15.2 percent. We note that the only difference between this regression and the baseline regression is the assumption on dividend dynamics. That is, we assume that investors learn about dividend dynamics using our model in this regression, whereas in the baseline regression we assume that expected dividend growth rates are constant. This means that we can attribute the increase in the R-square value from 10.2 percent to 15.2 percent to our modeling of learning about dividend dynamics. We also run a bivariate regression of stock index returns over the next year on both price-to-dividend ratios and stock yields, computed assuming that investors learn about dividend dynamics using our model, and report regression statistics in the third column of Table 6. Results show that stock yields, computed assuming that investors learn about dividend dynamics using our model, strictly

[12]Stambaugh (1999) shows that, when variables are highly serially correlated, OLS estimators' finite-sample properties can deviate from the standard regression setting.

dominate price-to-dividend ratios in explaining future stock index returns.

To emphasize the importance of learning, we regress stock index returns over the next year on stock yields in (16), computed assuming that investors do not learn. That is, instead of estimating model parameters at each point in time based on data available at the time, we estimate those parameters once based on the full data sample. So at every point in time, the same parameter estimates are used to compute stock yields. We report regression statistics in the fourth column of Table 6. Results show that stock yields, computed using our model but assuming that investors do not learn, perform roughly as well as price-to-dividend ratios in predicting future stock index returns. This is consistent with results from Figure 3, which show that there is almost no noticeable difference between the time series of price-to-dividend ratios and stock yields, computed using our model but assuming that investors do not learn.

	J&L				vB&K	C&S
$p_t - d_t$	-0.116**		0.016			
	(0.054)		(0.089)			
sy_t		3.964***	4.355*		3.000**	2.741**
(Learning)		(1.133)	(2.199)		(1.390)	(1.056)
sy_t				3.753**		
(Full Info.)				(1.674)		
R^2 (Return)	0.102	0.152	0.152	0.105	0.088	0.106
R^2 (Excess Return)	0.090	0.140	0.141	0.093	0.075	0.094

Table 6: **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. vB&K), or the dividend model in Campbell and Shiller (1988b) (i.e. C&S), and assuming investors have to learn (i.e. Learning), or do not learn (i.e. Full Info.), about model parameters. Regression is based on data between 1976 and 2013. 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 $***$.

It is also worth emphasizing that, for learning to be relevant, the dividend model itself must be used by investors. To illustrate this point, we regress stock index returns over the next year on stock yields, computed assuming that investors learn about dividend dynamics using either of the baseline models. We report regression statistics in the fifth and sixth columns of Table 6. We find that stock yields, computed assuming that investors learn using either of the baseline models, also perform roughly as well as price-to-dividend ratios in explaining future dividend growth rates.

We note that stock index returns combine the risk free rate and risk premium. To investigate whether the gain in return predictability is for predicting the risk free rate or the risk premium, in the last row of Table 6, we report the R-square value for predicting stock index excess returns.[13] Results show that the gap in forecasting performance between stock yields, computed assuming that investors learn about dividend dynamics using our model, and price-to-dividend ratios is entirely for predicting the risk premium and is not for predicting the risk free rate.

Recall that there are six model parameters that affect expected dividend growth rates. These parameters are the unconditional means μ_d and μ_q of dividend growth rates and earnings-to-dividend ratios, the persistence ρ and θ of the latent variable x_t and earnings-to-dividend ratios, and coefficients ϕ and φ that connect earnings information to dividend dynamics. We analyze learning about which of the six parameters is most important for asset pricing. We divide the six parameters into one subset that includes the persistence ρ of the latent variable x_t and another subset that includes the other five parameters. We then shut down learning for one subset of parameters while still allowing investors to learn about the remaining parameters in our model. That is, parameters not subject to learning are fixed at their full sample estimated values whereas other parameters are estimated at each point in time based on data available at the time. We call this partial learning. We regress stock index returns over the next year on stock yields, computed assuming partial learning. We report regression statistics in Table 7. Results show that allowing investors to learn about some, but not all, of the six model parameters reduces the performance of the resulting stock yields in explaining future stock index returns. If we shut down learning about ρ, the R-square value is reduced from 15.2 percent to 11.5 percent, but still higher than the 10.2 percent under full information. If we shut down learning about the other five parameters, the R-square value is reduced to 12.5 percent. This shows that investors' learning is multi-dimensional, and not restricted to a specific parameter or few parameters. Nevertheless, we find that shutting down learning about ρ adversely affect

[13]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 $\frac{v\hat{a}r(r_{t+1}-\hat{r}_{t+1})}{v\hat{a}r(r_{t+1})}$, where $v\hat{a}r(\cdot)$ is the sample variance. The in-sample R-square value for predicting stock index excess returns is $\frac{v\hat{a}r((r_{t+1}-r_{f,t+1})-(\hat{r}_{t+1}-r_{f,t+1}))}{v\hat{a}r(r_{t+1}-r_{f,t+1})}$.

return predictability more than shutting down learning about the other five parameters combined. This suggests that learning about ρ has the strongest implications for asset pricing. This is also consistent with results from the previous sections that show ρ is the parameter hardest to learn and investors beliefs about ρ fluctuates the most over time.

	Shutting Down Learning about Φ	
	$\Phi = \{\mu_d, \mu_q, \phi, \varphi, \theta\}$	$\Phi = \{\rho\}$
sy_t (Learning)	3.806^{***} (1.483)	3.854^{***} (1.502)
R^2 (Return)	0.125	0.115
R^2 (Excess Return)	0.113	0.103

Table 7: **Stock Index Returns and Stock Yields (Partial Learning)**: 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 and assuming investor learn about some model parameters but not others. Regression is based on data between 1976 and 2013. 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 $***$.

4 Evidence on Preference for Early Resolution of Uncertainty

A large part of modern asset pricing is built on the assumption, first formalized by Kreps and Porteus (1978), Epstein and Zin (1989), and Weil (1989), and then elaborated by Bansal and Yaron (2004) and others, that investors prefer early resolution of uncertainty. Under this assumption, long run expected growth risk requires additional compensation over short run expected growth risk. In this section, we provide evidence on investors' preference for early resolution of uncertainty. Because results from Table 6 suggest that investors learn about dividend dynamics based on data available at the time, investors' beliefs about the persistence of dividend growth rates vary over time as more data become available. We can examine how discount rates covary with investors' beliefs about the persistence of dividend growth rates and, from this relationship, infer whether investors have a preference for early or late resolution of uncertainty.

In our model, the persistence of dividend growth rates is jointly determined by the persistence ρ of the latent variable x_t and the persistence θ of earnings-to-dividend ratios. To derive a unified measure of persistence, note that, assuming that investors learn about dividend dynamics using our model, one standard deviation shocks to both the latent variable x_t and earnings-to-dividend ratios increase long run dividend growth expectations in (15) by:[14]

$$\begin{aligned}&\partial_t|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\partial_t|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)\\&=(1-\kappa_1)\sum_{s=0}^{\infty}\frac{\kappa_1^s}{1-\phi}\left(\rho^s\sigma_x+\theta^s(\varphi-(1-\theta)\phi)\sigma_q\right)\\&=\frac{(1-\kappa_1)\sigma_x}{1-\phi}\frac{1}{1-\kappa_1\rho}+\frac{(1-\kappa_1)(\varphi-(1-\theta)\phi)\sigma_q}{1-\phi}\frac{1}{1-\kappa_1\theta}.\end{aligned}\tag{18}$$

The same shocks' effect on dividend growth rates over the next year is:

$$\begin{aligned}&\Delta d_{t+1}|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\Delta d_{t+1}|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)\\=&\frac{1}{1-\phi}\left(\sigma_x+(\varphi-(1-\theta)\phi)\sigma_q\right).\end{aligned}\tag{19}$$

The ratio between the short run and the long run effects on dividend growth rates of one standard deviation shocks to both x_t and earnings-to-dividend ratios is:

$$\begin{aligned}&\frac{\Delta d_{t+1}|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\Delta d_{t+1}|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)}{\partial_{t+1}|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\partial_{t+1}|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)}\\&=\frac{\sigma_x+(\varphi-(1-\theta)\phi)\sigma_q}{\frac{(1-\kappa_1)\sigma_x}{1-\kappa_1\rho}+\frac{(1-\kappa_1)(\varphi-(1-\theta)\phi)\sigma_q}{1-\kappa_1\theta}}.\end{aligned}\tag{20}$$

This ratio is increasing in the speed at which shocks to dividend growth rates die out over time. That is, if shocks to dividends have stronger effects on long run dividend growth expectations, this ratio is lower. Thus, we define the persistence of dividend growth rates (i.e. ω) as minus this ratio, scaled so that ω is between -1 and 1:[15]

$$\begin{aligned}\omega&=\frac{1}{\kappa_1}-\frac{(1-\kappa_1)}{\kappa_1}\cdot\frac{\Delta d_{t+1}|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\Delta d_{t+1}|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)}{\partial_{t+1}|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\partial_{t+1}|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)}\\&=\frac{1-\frac{\sigma_x+(\varphi-(1-\theta)\phi)\sigma_q}{\frac{\sigma_x}{1-\kappa_1\rho}+\frac{(\varphi-(1-\theta)\phi)\sigma_q}{1-\kappa_1\theta}}}{\kappa_1}.\end{aligned}\tag{21}$$

[14] We denote $\partial_t|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)\equiv(1-\kappa_1)\sum_{s=0}^{\infty}\kappa_1 E_t[\Delta d_{t+s+1}|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)]$.

[15] We only use investors' beliefs about ω in regressions, so the scaling does not affect the statistical significance of any of our results.

To model investors' learning about the persistence ω of dividend growth rates, we estimate our dividend model at each point in time based on data available at the time. That is, denote $\omega(t)$ as ω, computed using model parameters estimated based on data up to time-t. We use $\omega(t)$ as a proxy for investors' time-t belief about ω. In Figure 4, we plot $\omega(t)$ between 1976 and 2013. We note that investors' beliefs about ω fluctuates significantly over time.

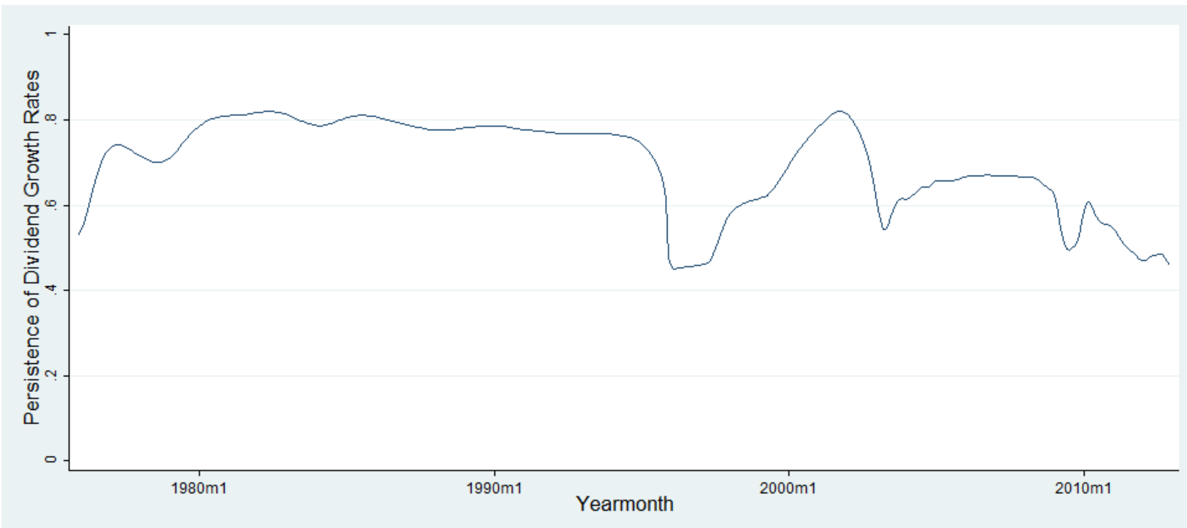


Figure 4: **Investors' Beliefs about the Persistence of Dividend Growth Rates.** This figure plots investors' beliefs about the persistence of dividend growth rates for the period between 1976 and 2013. Dividend model parameters are estimated based on data since 1946.

4.1 A Thought Experiment

We argue that, if investors prefer early resolution of uncertainty, we expect future dividends to be more heavily discounted when investors believe dividend growth rates to be more persistent. On the other hand, if investors prefer late resolution of uncertainty, we expect discount rates to be lower when investors believe dividend growth rates to be more persistent.

To fix ideas, we consider the simplest equilibrium asset pricing model that features 1) investors' preferences for early or late resolution of uncertainty, and 2) persistent dividend growth rates. In this thought experiment, we assume there is a representative agent who has Epstein and Zin (1989) preferences, defined recursively as:

$$U_t = \left[(1-\delta)\tilde{C}_t^{\frac{1-\alpha}{\zeta}} + \delta\left(E_t\left[U_{t+1}^{1-\alpha}\right]\right)^{\frac{1}{\zeta}}\right]^{\frac{\zeta}{1-\alpha}}, \ \zeta = \frac{1-\alpha}{1-\frac{1}{\psi}}, \tag{22}$$

where $\tilde{C}_t$ is real consumption, ψ is the elasticity of intertemporal substitution (EIS), and α 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 $\zeta > 0$.[16] Log of the intertemporal marginal rate of substitution (IMRS) is:

$$m_{t+1} = -\zeta \log(\delta) - \frac{\zeta}{\psi}\Delta\tilde{c}_{t+1} + (\zeta - 1)\,\tilde{s}_{t+1}, \tag{23}$$

where $\tilde{s}_{t+1}$ denotes real return of the representative agent's wealth portfolio.

Only for the purpose of this thought experiment, suppose dividend growth rates follow the AR[1] process:[17]

$$\Delta d_{t+1} - \mu_d = \omega\left(\Delta d_t - \mu_d\right) + \sigma_d \epsilon_{d,t+1}, \ \epsilon_{d,t+1} \sim \text{i.i.d } \mathbb{N}(0,1), \tag{24}$$

and suppose that dividend is the representative agent's only source of consumption. Further, to keep the setup as simple as possible, we assume away inflation. That is, let $\tilde{c}_t = \log(\tilde{C}_t)$ be log real consumption and $\Delta\tilde{c}_{t+1} = \tilde{c}_{t+1} - \tilde{c}_t$ be its growth rate, we can write:

$$\Delta\tilde{c}_{t+1} = \Delta d_{t+1}. \tag{25}$$

We assume the representative agent price the stock index using the Kreps (1998) anticipated utility. Assuming anticipated utility implies that investors maximize utility at each point in time assuming the current model parameter estimates are the true parameters, but then revise estimates as new data arrive. This means that investors do not account for the fact that estimates will continue to be revised in the future in their decisions. In other words, parameter uncertainty itself is not a priced risk factor in this framework. In Kreps's view, anticipated utility captures how investors compute utility when it is too computationally prohibitive to account for the fact that model parameter estimates will be revised in the future. Currently, this is often used in the macroeconomics

[16] Or equivalently, if $\alpha > 1$, then the representative agent prefers early resolution of uncertainty if $\psi > 1$ and prefers late resolution of uncertainty if $\psi < 1$.

[17] The AR[1] assumption is adopted because it is the simplest way to model persistent dividend growth rates.

and asset pricing literature for dealing with parameter uncertainty in a dynamic setup.

Given the representative agent's preferences in (22), consumption dynamics in (25), and dividend dynamics in (24), we solve for equilibrium price-to-dividend ratios and expected returns in the Appendix. In solving this model, we closely follow the steps in Bansal and Yaron (2004). We note that, given the setup of this thought experiment, the unconditional mean μ_r of stock index returns is:

$$\begin{aligned} \mu_r = E\left[r_{t+1}\right] &= -\log(\delta) + \frac{\mu_d}{\psi} - \frac{\zeta \kappa_1^2 A^2 \sigma_d^2}{2}. \\ A &= \frac{1-\frac{1}{\psi}}{1-\kappa_1 \omega}. \end{aligned} \tag{26}$$

where $E\left[\cdot\right]$ denotes unconditional expectation and κ_0 and κ_1 are log-linearizing constants. We note that $-\zeta A^2$ is increasing in beliefs about ω if $\zeta < 0$, and, although changes in μ_r affect κ_0 and κ_1 and feedback into μ_r, these effects on μ_r through the log-linearizing constants and vice versa are relatively small and not of first order importance. Therefore, if the representative agent has preferences for early resolution of uncertainty, $\zeta < 0$, and μ_r is increasing in beliefs about ω. On the other hand, if the representative agent prefers late resolution of uncertainty, μ_r is decreasing in beliefs about ω. Intuitively, this is because, if investors prefer early resolution of uncertainty, persistent shocks to dividend growth rates carry a positive risk premium, and the more persistent the shocks the higher that premium. On the other hand, if investors prefer late resolution of uncertainty, the premium carried by persistent shocks to dividend growth rates is negative.

4.2 Empirical Results

We derive two ways to evaluate the relationship between investors' beliefs about the persistence ω of dividend growth rates and the unconditional mean μ_r of stock index returns. First, we note that as μ_r increases, price-to-dividend ratios on average decrease and stock yields on average increase. To see why, we note from (12) and (13) that:

$$\begin{aligned} E\left[p_t - d_t\right] &= \frac{\kappa_0 + \mu_d - \mu_r}{1-\kappa_1}, \\ E\left[sy_t\right] &= \mu_r. \end{aligned} \tag{27}$$

Thus, if investors prefer early resolution of uncertainty, we expect price-to-dividend ratios to be lower and stock yields to be higher when investors believe dividend growth rates to be more persistent.[18] On the other hand, if investors prefer late resolution of uncertainty, we expect the exact opposite effects. To test this, we regress price-to-dividend ratios and stock yields, computed assuming that investors learn about dividend dynamics using our model, on investors' beliefs about ω and report regression statistics in the first and second columns of Table 8. Consistent with the assumption that investors prefer early resolution of uncertainty, we find higher investors' beliefs about ω are associated with lower price-to-dividend ratios and higher stock yields, and vice versa. Between 1976 and 2013, investors' beliefs about ω explain 25.3 percent of the variation in price-to-dividend ratios and 15.7 percent of the variation in stock yields. It is possible that the covariance between investors' beliefs about ω and either price-to-dividend ratios or stock yields is driven by the covariance between investors' beliefs about ω and long run dividend growth expectations. To rule out this possibility, we regress long run dividend growth expectations, computed assuming that investors learn about dividend dynamics using our model, on investors' beliefs about ω, and report regression statistics in the third column of Table 8. We find that investors' beliefs about ω do not explain long run dividend growth expectations and affect price-to-dividend ratios and stock yields only through the discount rate channel.

	$p_t - d_t$	sy_t (Learning)	∂_t
$\omega(t)$	-2.112*** (0.647)	0.060** (0.025)	-0.003 (0.015)
R^2	0.253	0.157	0.001

Table 8: **Stock Index Prices and Investors' Beliefs about the Persistence of Dividend Growth Rates**. This table reports the coefficient estimates and the R-square value from regressing log price-to-dividend ratios, stock yields, or long run dividend growth expectations, computed assuming that investors learn about dividend dynamics using our dividend model, on investors' beliefs about the persistence ω of dividend growth rates. Regression is based on overlapping annual data between 1976 and 2013. 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 $***$.

Next, we note that investors' preferences for the timing of resolution of uncertainty also have a direct effect on the term structure of expected returns. If investors prefer

[18] Although changes in μ_r affect both κ_0 and κ_1, these effects are relatively small and not of first order importance.

early resolution of uncertainty, then after controlling for price-to-dividend ratios or stock yields, we expect stock index returns over the short-horizon to be lower when investors believe dividend growth rates to be more persistent, and vice versa. That is, if we observe the same price-to-dividend ratios on two separate dates, but know that investors' beliefs about the persistence ω of dividend growth rates are different on these two dates, then we expect stock index returns over the short-horizon to be lower for the date on which investors believe dividend growth rates to be more persistent. To see why, suppose that expected returns follow a stationary AR[1] process as in van Binsbergen and Koijen (2010):

$$E_{t+1}[r_{t+2}] - \mu_r = \gamma \left(E_t[r_{t+1}] - \mu_r\right) + \sigma_r \epsilon_{r,t+1}. \tag{28}$$

Substituting (28) into the log-linearizing present value relationship of Campbell and Shiller (1988a), we can write expected returns over the short-horizon as:

$$E_t[r_{t+1}] = -(1-\kappa_1\gamma)(p_t - d_t) + \frac{1-\kappa_1\gamma}{1-\kappa_1}\partial_t + \frac{(1-\kappa_1\gamma)\kappa_0 - \kappa_1(1-\gamma)\mu_r}{1-\kappa_1} \tag{29}$$

Or equivalently, we can write:

$$E_t[r_{t+1}] = \frac{1-\kappa_1\gamma}{1-\kappa_1} sy_t - \frac{\kappa_1(1-\gamma)\mu_r}{1-\kappa_1}. \tag{30}$$

Although changes in μ_r affect both κ_0 and κ_1, these effects are relatively small and not of first order importance. Thus, we note from (29) and (30) that, after controlling for price-to-dividend ratios or stock yields, expected stock index returns over the short-horizon are decreasing in μ_r, and thus decreasing in investors' beliefs about ω. Intuitively, this is because, when μ_r increases, in order to justify the same price-to-dividend ratio or stock yield, expected returns over the short-horizon must decrease sufficiently to compensate for the effect of an increase in μ_r on expected returns over the long-horizon. To confirm this relationship, we run bivariate regressions of stock index returns over the next year on investors' beliefs about ω and either price-to-dividend ratios or stock yields, computed assuming that investors learn about dividend dynamics using our model. We report regression statistics in Table 9. Results confirm that, after controlling for either price-to-dividend ratios or stock yields, higher investors' beliefs about ω predict lower stock index returns over the short-horizon, and vice versa. We find that, between 1976 and 2013, price-to-dividend ratios (stock yields) and investors' beliefs about ω together explain as much as 23.5 (26.5) percent of the variation in stock index returns over the next year. Further, comparing results reported in Table 9 to those reported in Table 6, we find that, not only is investors' beliefs about ω an effective predictor of future stock index returns, including them as an additional regressor also strengthens the individual predictive performance of price-to-dividend ratios and stock yields.

$\omega(t)$	-0.644**	-0.562**
	[0.184]	[0.129]
$p_t - d_t$	-0.193***	
	[0.052]	
sy_t (Learning)		5.448***
		[0.897]
R^2 (Return)	0.235	0.265
R^2 (Excess Return)	0.224	0.255

Table 9: **Stock Index Returns, Stock Yields, and Investors' Beliefs about the Persistence of Dividend Growth Rates**. This table reports the coefficient estimates and R-square value from regressing stock index returns over the next year on investors' beliefs about the persistence ω of dividend growth rate, log price-to-dividend ratios, and stock yields, computed assuming that investors learn about dividend dynamics using our dividend model. Regression is based on data between 1976 and 2013. 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 $***$.

5 Learning about Dividend Dynamics in an Equilibrium Asset Pricing Model

Results from the previous sections show that investors' learning about dividend dynamics is reflected in stock index prices and returns. In this section, we embed learning about dividend dynamics into a realistic equilibrium asset pricing model to quantitatively capture these features of the data.

5.1 Preferences and Consumption Dynamics

Aside from proposing a dividend model, building an equilibrium asset pricing model requires us to specify investors' preferences and consumption dynamics. Because results in previous sections show that investors prefer early resolution of uncertainty, a natural choice is to combine our dividend model with Epstein and Zin (1989) preferences in (22). Our choice of preference parameters reflect common standards set by the existing literature. That is, we set the coefficients of risk aversion α to 5, the intertemporal elasticity of substitution (EIS) ψ to 1.5, and the annual time discount factor δ to 0.98. The choice of an EIS that is greater than one reflect preferences for early resolution of uncertainty.

We also needs to specify a consumption process. To stress that our results do not rely on a specific model of consumption, we consider two different specifications of consumption dynamics.

5.1.1 First Specification

We adopt the consumption model in Bansal and Yaron (2004) as our first specification of consumption dynamics. That is, we assume that the expected growth rates in real consumption follow an AR[1] process and allow volatility in consumption growth rates to be time varying. In other words, we describe real consumption growth rates using the following system of equations:

$$\begin{aligned}\Delta\tilde{c}_{t+1} - \mu_c &= \frac{1}{\gamma(1-\phi)} x_t + \sigma_t \epsilon_{c,t+1} \\ \sigma^2_{t+1} - \sigma^2_c &= \varrho\left(\sigma^2_t - \sigma^2_c\right) + \sigma_\varsigma \epsilon_{\varsigma,t+1}. \end{aligned} \tag{31}$$

The correlation matrix for shocks to dividend and real consumption dynamics can be written as:

$$\begin{pmatrix} \epsilon_{c,t+1} \\ \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{\varsigma,t+1} \\ \epsilon_{q,t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N}\left(\mathbf{0}, \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \lambda_{dx} & 0 & \lambda_{dq} \\ 0 & \lambda_{dx} & 1 & 0 & \lambda_{xq} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & \lambda_{dq} & \lambda_{xq} & 0 & 1 \end{pmatrix}\right). \tag{32}$$

Because we do not use actual consumption data in this paper, the correlations that involve shocks $\epsilon_{c,t+1}$ or $\epsilon_{\varsigma,t+1}$ to the real consumption process cannot be identified. So, for convenience, we set them to zeros. The rest of the correlation matrix can be estimated from dividend and earnings data.

We note that the unconditional mean of real consumption growth rates must equal to the unconditional mean of dividend growth rates minus inflation rates, or else dividend as a fraction of consumption will either become negligible or explode. To convert between nominal and real rates, we set expected inflation rates to a constant $\mu_\pi = 0.036$, and so $\mu_c = \mu_d - \mu_\pi$.[19] We assume that the latent variable x_t in real consumption growth rates is the same as the latent variable in dividend growth rates. We recall that dividend growth rates in our model have the functional form:

$$\Delta d_{t+1} = \frac{1}{1-\phi} x_t + \frac{\varphi - (1-\theta)\phi}{1-\phi} \left(q_t - \mu_q \right) + \frac{\phi}{1-\phi} \sigma_q \epsilon_{d,t+1} + \frac{1}{1-\phi} \sigma_d \epsilon_{d,t+1}. \tag{33}$$

So γ is the dividend leverage parameter. We set it to 5. The primary effect of this parameter is on the unconditional mean of the equity premium. In Bansal and Yaron (2004), the persistence ρ of the latent variable x_t is set to 0.975 at the monthly frequency. 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.[20] This criticism serves as the rationale for why we expect learning to be important. To calibrate the dynamics of consumption volatility, we follow Bansal and Yaron (2004), who set σ_c to 0.0078, ϱ to 0.987, and σ_ς to $0.23 \cdot 10^{-5}$ at the monthly frequency. We convert these to their annual equivalents. We note that our long run risk model differs from the setup in Bansal and Yaron (2004) in that we shut down heteroskedastic volatility in the dividend process. We do this in order to incorporate our dividend model, which is estimated under homoskedasticity.

We solve this specification of our long-run risk model in the Appendix. In solving this model, we closely follow the steps in Bansal and Yaron (2004). The model consists of three state variables: 1) the latent variable x_t, 2) the latent variable σ_t^2, and 3) earnings-to-dividend ratios. We can solve for price-to-dividend ratio in this model as a linear function

[19] The choice of 0.036 is to match the average inflation rate between 1946 and 2013. We also modeled inflation as an AR[1] process, and our conclusions do not change. These results can be made available upon request.

[20] See Beeler and Campbell (2012), Marakani (2009).

of the three state variables:

$$p_t - d_t = A_{d,0} + A_{d,1} x_t + A_{d,2} \sigma_t^2 + A_{d,3} \left(q_t - \mu_q \right). \tag{34}$$

We can solve for expected return over the short-horizon as:

$$E_t[r_{t+1}] = A_{r,0} + A_{r,1} x_t + A_{r,2} \sigma_t^2, \tag{35}$$

where coefficients $A_{d,\cdot}$ and $A_{r,\cdot}$, derived in the Appendix, are functions of the parameters governing investors' preferences, consumption dynamics, and dividend dynamics. We note that, substituting (34) into (35), we can avoid estimating time varying consumption volatility directly and instead write expected return over the short horizon as a function state variables that can be estimated from dividend dynamics and price-to-dividend ratio:

$$E_t \left[r_{t+1} \right] = \frac{A_{r,0} A_{d,2} - A_{r,2} A_{d,0}}{A_{d,2}} + \frac{A_{r,0} A_{d,2} - A_{r,2} A_{d,0}}{A_{d,2}} x_t - \frac{A_{r,2} A_{d,3}}{A_{d,2}} \left(q_t - \mu_q \right) + \frac{A_{r,2}}{A_{d,2}} \left(p_t - d_t \right). \tag{36}$$

5.1.2 Second Specification

In our second specification of consumption dynamics, we assume consumption volatility to be constant over time. We note that in our first specification, time varying volatility in real consumption growth rates serves as an additional state variable, so to make up for the lost state variable, we assume the correlation between shocks $\epsilon_{c,t+1}$ to real consumption growth rates and shocks $\epsilon_{q,t+1}$ to earnings-to-dividend ratios to be time varying.[21] In other words, this specification allows for the conditional covariance of real consumption and dividend growth rates to vary over time. We note that this conditional covariance can be negative in some states of the world, i.e. dividends can be hedges against shocks to consumption. This gives us our second specification of consumption dynamics, which can be described by the following system of equations:

$$\begin{aligned} \Delta \tilde{c}_{t+1} - \mu_c &= \frac{1}{\gamma(1-\phi)} x_t + \sigma_c \epsilon_{c,t+1} \\ \lambda_{t+1} - \mu_\lambda &= \varrho \left(\lambda_t - \mu_\lambda \right) + \sigma_\lambda \epsilon_{\lambda,t+1}, \end{aligned} \tag{37}$$

[21] We note that this additional state variable that cannot be estimated from dividend dynamics alone is necessary for the model to fit both the time series of dividends and price-to-dividend ratios.

where λ_t is the correlation between $\epsilon_{c,t+1}$ in (37) and $\epsilon_{q,t+1}$ in (6) and serves as the additional state variable in this specification of consumption dynamics. Thus, the correlation matrix for shocks to dividend and real consumption dynamics can be written as:

$$\begin{pmatrix} \epsilon_{c,t+1} \\ \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{\lambda,t+1} \\ \epsilon_{q,t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N} \left(\mathbf{0}, \begin{pmatrix} 1 & 0 & 0 & 0 & \lambda_t \\ 0 & 1 & \lambda_{dx} & 0 & \lambda_{dq} \\ 0 & \lambda_{dx} & 1 & 0 & \lambda_{xq} \\ 0 & 0 & 0 & 1 & 0 \\ \lambda_t & \lambda_{dq} & \lambda_{xq} & 0 & 1 \end{pmatrix} \right). \tag{38}$$

We set the consumption dynamics parameters σ_c and ϱ to be the same as those in the first specification. We assume the unconditional mean μ_λ of the latent variable λ_t to be 0 and the standard deviation of shocks to λ_t to be 0.033 at the monthly frequency.[22] Because this specification of consumption dynamics has not been adopted in the existing literature, our choice of σ_λ can appear arbitrary. However, we note that our results are not sensitive to setting σ_λ to 0.033.[23] We solve this specification of our long run risk model in the Appendix. In solving this model, we closely follow the steps in Bansal and Yaron (2004). The model consists of three state variables: 1) the latent variable x_t, 2) the latent variable λ_t, and 3) earnings-to-dividend ratios. We can solve for price-to-dividend ratio in this model as a linear function of the three state variables:

$$p_t - d_t = A_{d,0} + A_{d,1} x_t + A_{d,2} \lambda_t + A_{d,3} \left(q_t - \mu_q \right). \tag{39}$$

We can solve for expected return over the short-horizon as:

$$E_t[r_{t+1}] = A_{r,0} + A_{r,1} x_t + A_{r,2} \lambda_t, \tag{40}$$

where coefficients $A_{d,\cdot}$ and $A_{r,\cdot}$, derived in the Appendix, are functions of the parameters governing investors' preferences, consumption dynamics, and dividend dynamics. We note that, substituting (39) into (40), we can avoid estimating time varying correlation between shocks to real consumption growth rates and shocks to earnings-to-dividend ratios directly and instead write expected return over the short horizon as a function state variables that

[22]Because, following Bansal and Yaron (2004), calibrations of σ_c and ϱ are reported in monthly frequency, we report our parameter choices for μ_λ and σ_λ in monthly frequency as well for ease of comparison. In solving our model, we convert them to their annual equivalents.

[23]We also tried setting σ_λ to other values between 0.01 and 0.10 and find our results to be relatively unchanged.

can be estimated from dividend dynamics and price-to-dividend ratio:

$$E_t\left[r_{t+1}\right] = \frac{A_{r,0}A_{d,2} - A_{r,2}A_{d,0}}{A_{d,2}} + \frac{A_{r,0}A_{d,2} - A_{r,2}A_{d,0}}{A_{d,2}} x_t - \frac{A_{r,2}A_{d,3}}{A_{d,2}}\left(q_t - \mu_q\right) + \frac{A_{r,2}}{A_{d,2}}\left(p_t - d_t\right). \tag{41}$$

5.2 Empirical Results

5.2.1 Time Variation in Stock Index Returns

We examine how the two specifications of our long-run risk model performs in predicting stock index returns. We measure forecasting performance using the quasi out-of-sample R-square value as defined in Goyal and Welch (2008):

$$R^2 = 1 - \frac{\sum_{t=t_0}^{T-1}\left(r_{t+1} - E_t[r_{t+1}]\right)^2}{\sum_{t=t_0}^{T-1}\left(r_{t+1} - \hat{\mu}_r(t)\right)^2}. \tag{42}$$

where $\hat{\mu}_r(t)$ is the average of stock index returns up to time-t:

$$\hat{\mu}_r(t) = \frac{1}{t}\sum_{s=0}^{t-1} r_{s+1}, \tag{43}$$

and we use time-t_0 to denote the end of the training period. We use the first 30 years of the data sample as the training period and compute the quasi out-of-sample R-square value using data between 1976 and 2013. We use the term quasi to refer to the fact that, although parameters in our dividend model are estimated at each point in time based on data available at the time, parameters governing preferences and consumption dynamics are fixed and can be forward looking.[24] In Table 10, we report the quasi out-of-sample R-square value for predicting annual stock index returns using expected returns, computed assuming investors learn about dividend dynamics using our long-run risk model. That is, we estimate dividend model parameters at each point in time based on data available at the time and substitute these parameters into either (36) or (41) to compute model implied expected returns. We find that, between 1976 and 2013, our long-run risk model predicts up to 22.9 percent of the variation in annual stock index returns.

To see why the expected returns derived from our long run risk model is able to capture the variation in future stock index returns, we regress model implied expected returns,

[24] For example, under the first specification, these parameters include α, ψ, and δ describing investors' preferences and γ, σ_c, ϱ, σ_ς describing the consumption process.

First Specification		Second Specification	
R^2	p-value	R^2	p-value
0.224	0.003	0.229	0.003

Table 10: **Stock Index Returns and Expected Returns.** This table reports the out-of-sample R-square value for predicting stock index returns using expected returns derived from our long-run risk model, assuming that investors learn about dividend dynamics. Also reported is the corresponding p-value from the adjusted-MSPE statistics of Clark and West (2007). Statistics are based on data between 1976 and 2013. Dividend model parameters are estimated based on data since 1946.

computed assuming investors learn about dividend dynamics, on price-to-dividend ratios, long run dividend growth expectations, and investors' beliefs about the persistence ω of dividend growth rates, and report regression statistics in Table 11. Results show that price-to-dividend ratios, long run dividend growth expectations, and investors' beliefs about ω together contribute up to 93.1 percent of the variation in model implied expected returns. That is, expected returns derived from our long run risk model, computed assuming investors learn about dividend dynamics, captures three key relationships that drive to our long run risk model's forecasting performance. First, increases in price-to-dividend ratios are associated with decreases in future stock index returns. Second, increases in long run dividend growth expectations are associated with increases in future stock index returns. Third, increases in investors' beliefs about ω are associated with decreases in future stock index returns.

To isolate the contribution of learning to our long run risk model's forecasting performance, we regress stock index returns over the next year on model implied expected returns, computed assuming investors either have to learn, or do not learn, about dividend dynamics. That is, assuming investors learn, we estimate dividend model parameters at each point in time based on data available at the time and substitute these parameters into either (36) or (41) to compute model implied expected returns. On the other hand, assuming investors do not learn, we estimate model parameters once using the full data sample and apply these same parameter estimates to compute model implied expected returns on different dates. We report regression statistics in Table 12. We note that accounting for learning about dividend dynamics is responsible for over forty-percent of the predictability in stock index returns.

To examine the robustness of our long-run risk model's forecasting performance, we follow Goyal and Welch (2008) and define the cumulative sum of squared errors difference

	First Specification	Second Specification
$p_t - d_t$	-0.159***	-0.159***
	(0.013)	(0.013)
∂_t	5.971***	6.056***
(Learning)	(0.694)	(0.700)
$\omega(t)$	-0.423***	-0.425***
	(0.060)	(0.062)
R^2	0.931	0.929

Table 11: **Decomposition of Model Implied Expected Returns**. This table reports the coefficient estimates and R-square value from regressing model implied expected returns on log price-to-dividend ratios, long run dividend growth expectations, and investors' beliefs about the persistence ω of dividend growth rates. Regression is based on data between 1976 and 2013. 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 $***$.

(SSED) as:

$$SSED_t = \sum_{s=t_0}^{t-1} \Big((r_{s+1} - E_t[r_{s+1}])^2 - (r_{s+1} - \hat{\mu}_r(t))^2 \Big) . \tag{44}$$

We plot SSED in Figure 5. We note that if the forecasting performance of our long-run risk model is stable and robust, 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. Figure 5 shows that our model's forecasting performance is relatively stable and robust between 1976 and 2013, except for right before the Dot-Com crash and during the recent financial crisis, when our long run risk model slightly underperform.Henkel, Martin, and Nardari (2011) and Golez and Koudijs (2014) argue that almost all of price-to-dividend ratios' performance in forecasting future stock index returns is realized during recessions. However, we note that, once we account for learning about dividend dynamics, our model's forecasting performance is roughly as strong during National Bureau of Economic Research (NBER) expansions as during recessions.

	First Specification		Second Specification	
	Learning	Full Info.	Learning	Full Info.
Est.	0.863***	0.694**	0.897***	0.695**
	(0.137)	(0.294)	(0.137)	(0.295)
R^2	0.214	0.116	0.218	0.116

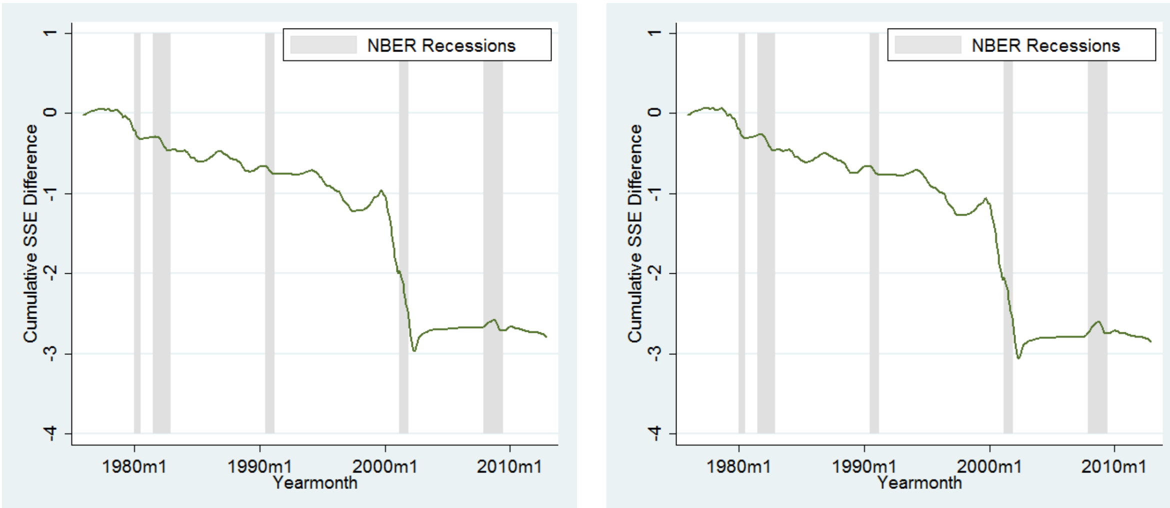
Table 12: **Stock Index Returns and Expected Returns (Learning vs Full Info.)**. This table reports the coefficient estimates and R-square value from regressing stock index returns over the next year on model implied expected returns, computed assuming that investors either have to learn, or do not learn, about parameters in the dividend model. Statistics are based on data between 1975 and 2013. 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 $***$.

5.2.2 Time Variation in Price-to-Dividend Ratios

It is a common wisdom in the asset pricing literature that the variation in price-to-dividend ratios is primarily driven by the variation in discount rates, and not cash flow expectations. For example, Cochrane (2011) states that "the variance of dividend yields or price-dividend ratios corresponds entirely to discount-rate variation". To analyze this statement in light of our findings, we perform a decomposition of price-to-dividend ratios. That is, assuming that investors learn about dividend dynamics using the first specification of our long-run risk model, we can label (34) as:

$$p_t - d_t = \underbrace{A_{d,0}(t)}_{1)} + \underbrace{A_{d,1}(t) \cdot x_t}_{2)} + \underbrace{A_{d,2}(t) \cdot \sigma_t^2}_{3)} + \underbrace{A_{d,3}(t) \cdot \big(q_t - \mu_q(t)\big)}_{4)}, \tag{45}$$

where $A_{d,\cdot}(t)$ denotes $A_{d,\cdot}$ computed using model parameters estimated based on data avaliable at time-t. This means that the variation in price-to-dividend ratios is attributable to 1) the variation in investors' beliefs about parameters in the dividend model, 2) the variation in the latent variable x_t, 3) the variation in the latent variable σ_t^2, or 4) the variation in earnings-to-dividend ratios. Thus, we can decompose the sample variance



First Specification   Second Specification

Figure 5: **Stock Index Returns and Model Implied Expected Returns (Cumulative SSE Difference).** This figure plots the cumulative sum of squared errors difference for the period between 1976 and 2013. Dividend model parameters are estimated based on data since 1946.

of price-to-dividend ratios into:

$$\begin{aligned} \hat{var}\left(p_t - d_t\right) = & \underbrace{\hat{cov}\left(p_t - d_t, A_{d,0}(t)\right)}_{1)} + \underbrace{\hat{cov}\left(p_t - d_t, A_{d,1}(t) \cdot x_t\right)}_{2)} \\ & + \underbrace{\hat{cov}\left(p_t - d_t, A_{d,2}(t) \cdot \sigma_t^2\right)}_{3)} + \underbrace{\hat{cov}\left(p_t - d_t, A_{d,3}(t) \cdot \left(q_t - \mu_q(t)\right)\right)}_{4)}. \end{aligned} \tag{46}$$

where $\hat{var}(\cdot)$ denotes sample variance and $\hat{cov}(\cdot)$ denotes sample covariance.

We perform the same decomposition assuming investors price the stock index using the second specification of our long run risk model. We report the decomposition results in Table 13. Results show that about half of the variation in price-to-dividend ratios is driven by the variation in the latent variable σ_t^2 in the first specification or λ_t in the second specification, which affects discount rates but not cash flow expectations. The other half of the variation in price-to-dividend ratios can be attributed to learning about dividend dynamics, primarily changes in investors' beliefs about the parameters governing the dividend process. These results challenge the common wisdom in the literature that price-to-dividend ratios do not incorporate information about future dividend expectations.

	Learning	x_t	σ_t^2 or λ_t	q_t
First Specification	0.443	0.002	0.533	0.022
Second Specification	0.416	0.008	0.491	0.085

Table 13: **Decomposing the Variation in Price-to-Dividend Ratios**. This table reports the fraction of the variation in price-to-dividend ratios that in attributable to the variation in investors beliefs about dividend model parameters, the variation in the latent variable x_t, the variation in the latent variable σ_t^2 in the first specification or λ_t in the second specification, and the variation in earnings-to-dividend ratios q_t. Statistics are based on data between 1976 and 2013. Dividend model parameters are estimated based on data since 1946.

5.3 Discussion

We discuss three features of our work in this section that can potentially motivate future research. First, we draw attention to the differences between the two specifications of our long run risk model. We recall that in our first specification, we use heteroskedastic volatility in real consumption growth rates as the additional state variable of our model, and in our second specification, we use the time varying correlation between shocks to real consumption growth rates and earnings-to-dividend ratios as the additional state variable. Nevertheless, as Tables 10, 11, 12, 13, and Figure 5 reveal, despite this difference, so far the asset pricing implications of the two specifications of our long run risk model are almost identical. In Figure 6, we plot the expected returns derived from the two specifications of our long run risk model, computed assuming investors learn about dividend dynamics, as well as the actual stock index returns over the next year, for the period between 1976 and 2013. Figure 6 confirms that the time series of expected returns under the two specifications almost completely overlap.

We then decompose expected returns derived from each of the two specifications of our long run risk model into model implied risk free rate and model implied risk premium.[25] In Figure 7, we plot 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 risk premium over the next year, for the period between 1976 and 2013. Interestingly, Figure 7 shows that the two specifications have completely different implications on the decomposition of expected returns into risk free rate and risk premium. That is, according to the first specification of consumption dynamics, almost all of the variation in expected returns is attributable

[25] We derive the risk free rate as a function of state variables and preference, consumption and dividend model parameters in the Appendix.

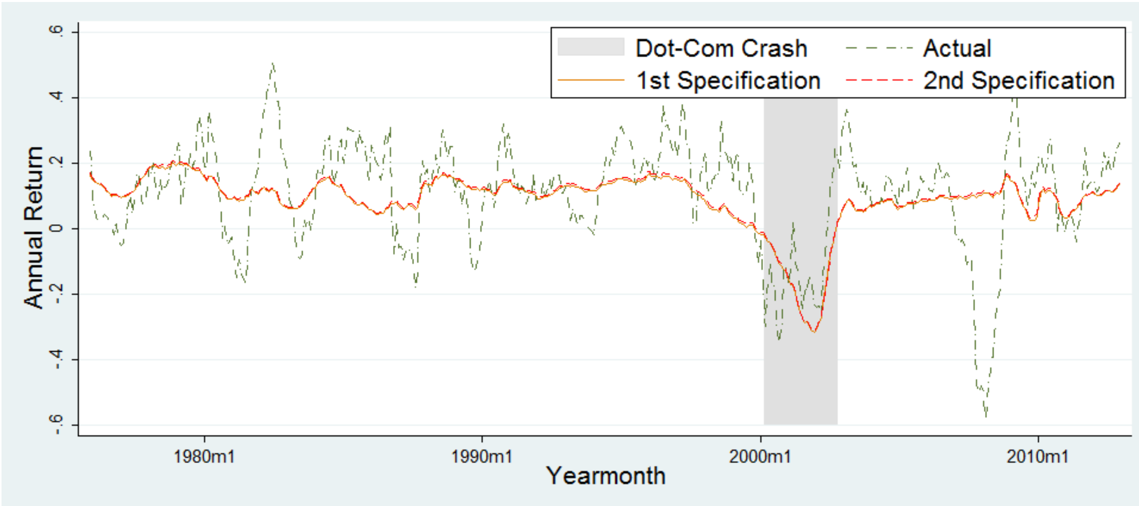


Figure 6: **Model Implied Expected Returns**. This figure plots expected returns derived from the two specifications of our long run risk model, as well as the actual stock index returns over the next year, for the period between 1975 and 2013. Dividend model parameters are estimated based on data since 1946.

to the variation in the risk free rate, whereas the risk premium hardly changes over time. To the contrary, according to the second specification of consumption dynamics, almost all of the variation in expected returns is attributable to the variation in risk premium, whereas the risk free rate hardly changes over time.

Clearly, we know from the data that the risk free rate is relatively constant over time. Thus, we can infer that, of the two specifications, the second specification of consumption dynamics is the more realistic one. 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 rate, we can use this decomposition to shed light on consumption dynamics. However, modeling consumption is not the focus of this paper, so we leave this to potential future research.

Second, we note that, according to both specifications of our long run risk model, expected returns during the Dot-Com crash are negative. We recall that, in our long run risk model, expected return is a linear function of state variables. For example, under the second specification, expected return can be written as:

$$E_t\left[r_{t+1}\right] = A_{r,0} + A_{r,1}x_t + A_{r,2}\lambda_t. \tag{47}$$

where coefficients $A_{r,\cdot}$ are always positive. So in states of the world where state variables x_t and λ_t are significantly negative, expected returns can be negative. However, our model

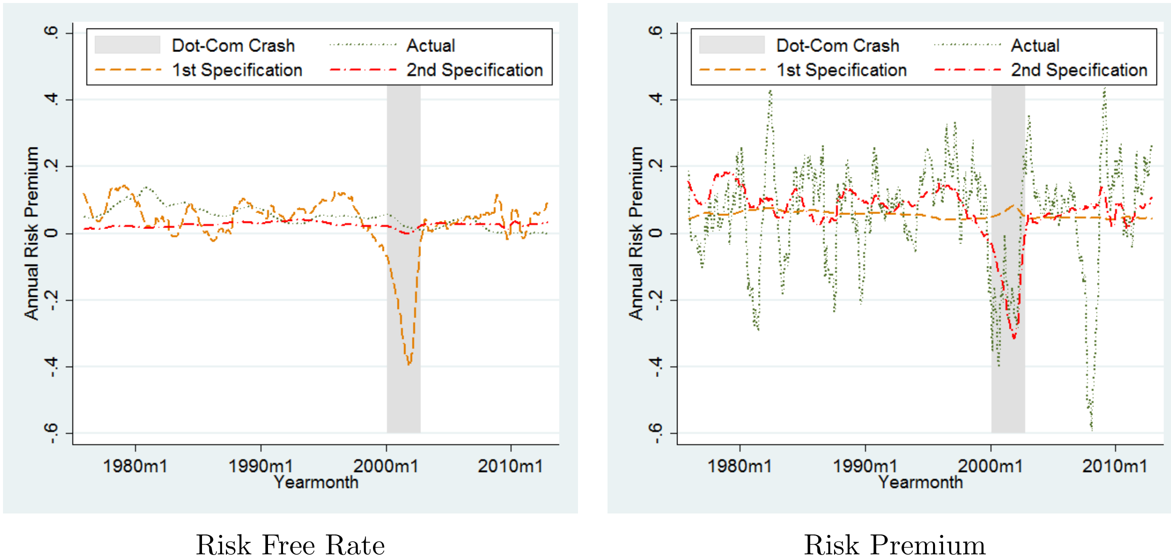


Figure 7: **Model Implied Risk Free Rate and Risk Premium.** This figure plots risk free rate and risk premium derived from the two specifications of our long run risk model, as well as the actual risk free rate and excess returns over the next year, for the period between 1975 and 2013. Dividend model parameters are estimated based on data since 1946.

does not shed any light on the deeper economic intuition behind why we estimate negative expected returns during the Dot-Com crash.

Third, we compare our estmation method to that of van Binsbergen and Koijen (2010). In that paper, the processes of dividends and price-to-dividend ratios are jointly estimated.[26] The rationale for jointly estimating these processes is that price-to-dividend ratios contain information about dividend dynamics that the econometrician cannot capture using a time series model of dividends. In other words, jointly estimating dividend and price-to-dividend ratio dynamics is the efficient approach under the assumption that investors and the econometrician have different information sets regarding dividend dynamics, so the econometrician can learn about dividend model parameters and state variables from price-to-dividend ratios. In this paper, we operate under the assumption that investors and the econometrician have the same information set regarding dividend dynamics. That is, we assume investors learn about dividend dynamics using the same model and the same data as the econometrician does. Thus, under this assumption, price-to-dividend ratios do not contain additional information that cannot be revealed by estimating our dividend model, and thus jointly estimating the processes of dividends and price-to-dividend ratios is not necessary and no more efficient than estimating dividend

[26]In van Binsbergen and Koijen (2010), both expected dividend growth rates and discount rates are modeled as stationary AR[1] processes.

model parameters and state variables using our dividend model alone. However, if investors and the econometrician's information sets regarding dividend dynamics are not exactly identical, we imagine three ways through which our estimation method can potentially be improved. First, as in van Binsbergen and Koijen (2010), one can estimate dividend and price-to-dividend ratio dynamics jointly. Second, one can estimate dividend and consumption dynamics jointly, if one is confident in a model of consumption.[27] Third, one can estimate the dividend model parameters using non-flat priors. We leave these to potential future research.

6 Conclusion

We propose a model for the dynamics of dividend growth rates that incorporates earnings information into the latent variable model of van Binsbergen and Koijen (2010). We show that the model performs well in capturing the variation in dividend growth rates, both in-sample and out-of-sample. We show that some parameters in our dividend model can be difficult to estimate with precision in finite sample. We argue that, as a result, learning about model parameters can have significant asset pricing implications.

We provide evidence that investors behave as if they learn about dividend dynamics using our model. First, we show that incorporating learning about dividend dynamics helps to forecast future stock index returns. Second, we find that changes in investors' beliefs about the persistence of dividend growth rates help to explain the variation in both long run discount rates and the term structure of discount rates. We show that the way discount rates respond to investors' beliefs about the persistence of dividend growth rates is consistent with investors' preferences for early resolution of uncertainty.

We embed learning about dividend dynamics into an equilibrium asset pricing model that features Epstein and Zin (1989) preferences and consumption dynamics from the long-run risks model of Bansal and Yaron (2004). We find that, between 1976 and 2013, our long-run risk model predicts 22.9 percent the variation in annual stock index returns. We show that, according to our model, learning about dividend dynamics contributes substantially to the variation in price-to-dividend ratios.

[27] There is also a debate on who is the marginal investor and correspondingly whose consumption should be relevant for pricing the stock index. For this, see the literature on intermediary asset pricing.

References

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.

Boudoukh, J., R. Michaely, M. Richardson, and M. R. Roberts (2007): "On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing," *Journal of Finance*, 62, 877–915.

Campbell, J. Y., and R. J. Shiller (1988a): "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. (2011): "Presidential Address: Discount Rates," *Journal of Finance*, 66, 1047–1108.

Cogley, T., and T. J. Sargent (2008): "Anticipated Utility and Rational Expectations as Approximations of Bayesian Decision Making," *International Economic Review*, 49, 185–221.

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.

Golez, B., and P. Koudijs (2014): "Four Centuries of Return Predictability," Discussion paper.

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," *American Economic Review*, 97.

Henkel, S. J., J. S. Martin, and F. Nardari (2011): "Time-Varying Short-Horizon Predictability," *Journal of Financial Economics*, 99, 560–580.

Johannes, M., L. Lochstoer, and P. Collin-Dufresne (2015): "Parameter Learning in General Equilibrium. The Asset Pricing Implications," *American Economic Review*, Forthcoming.

Johannes, M., L. A. Lochstoer, and Y. Mou (2015): "Learning about Consumption Dynamics," *Journal of Finance*, Forthcoming.

Ju, N., and J. Miao (2012): "Ambiguity, Learning, and Asset Returns," *Econometrica*, 80, 559–591.

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.

Kreps, D. M., and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, pp. 185–200.

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.

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.

Newey, W. K., and K. D. West (1987): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, pp. 777–787.

Piazzesi, M., and M. Schneider (2010): "Trend and Cycle in Bond Risk Premia," Discussion paper.

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.

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.

Weil, P. (1989): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421.

A Appendix

A.1 Derivation of Price-Dividend Ratios and Expected Returns

A.1.1 A Thought Experiment

We derive price-to-dividend ratios and expected returns implied by the equilibrium asset pricing model proposed in our thought experiment, which features dividend dynamics in (24), consumption dynamics in (25), and investors' preferences in (22). Our derivation closely follows the steps in Bansal and Yaron (2004). The log stochastic discount factor is:

$$m_{t+1} = \zeta \log(\delta) - \frac{\zeta}{\psi}\Delta\tilde{c}_{t+1} + (\zeta - 1)\tilde{s}_{t+1}. \tag{48}$$

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:

$$\tilde{s}_{t+1} = \kappa_0 + \kappa_1 z_{c,t+1} - z_{c,t} + \Delta\tilde{c}_{t+1}. \tag{49}$$

The log-linearizing constants are:

$$\kappa_0 = \log(1 + \exp(\bar{z}_c)) - \kappa_1(\bar{z}_c) \text{ and } \kappa_1 = \frac{\exp(\bar{z}_c)}{1 + \exp(\bar{z}_c)}.$$

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

$$z_{c,t} = A_{c,0} + A_{c,1}x_t. \tag{50}$$

We can write:

$$E_t[m_{t+1} + \tilde{s}_{t+1}] = \zeta \log(\delta) + \left(\zeta - \frac{\zeta}{\psi}\right)(\mu_d + x_t) + \zeta\kappa_0 + \zeta(\kappa_1 - 1)A_0 + \zeta\left(\kappa_1\omega - 1\right)A_1,$$

$$var_t\left(m_{t+1} + \tilde{s}_{t+1}\right) = \zeta^2\kappa_1^2 A_1^2 \sigma_d^2. \tag{51}$$

Using the condition $E_t[\exp(m_{t+1} + \tilde{s}_{t+1})] = 1$, we can solve for coefficients A_0 and A_1 as:

$$A_0 = \frac{1}{1-\kappa_1}\left(\log(\delta) + \left(1 - \frac{1}{\psi}\right)\mu_d + \kappa_0 + \frac{\zeta}{2}\kappa_1^2 A_1^2 \sigma_d^2\right),$$

$$A_1 = \frac{1 - \frac{1}{\psi}}{1 - \kappa_1\rho}. \tag{52}$$

Unconditonal mean of expected returns on wealth is:

$$E[\tilde{s}_{t+1}] = \kappa_0 + (\kappa_1 - 1)A_0 + \mu_d. \tag{53}$$

Because dividend and consumption dynamics are identical, the unconditional mean μ_r of expected stock index returns is:

$$\mu_r = E[\tilde{s}_{t+1}] = -\log(\delta) + \frac{\mu_d}{\psi} - \frac{\zeta}{2}\kappa_1^2 A_1^2 \sigma_d^2. \tag{54}$$

A.1.2 Full Model (First Specification)

We derive price-to-dividend ratios and expected returns implied by our long-run risk model, which features dividend dynamics in (6), consumption dynamics in (31), and investors preferences in (22). Our derivation closely follows the steps in Bansal and Yaron (2004). The log stochastic discount factor is given as:

$$m_{t+1} = \zeta \log(\delta) - \frac{\zeta}{\psi}\Delta\tilde{c}_{t+1} + (\zeta - 1)\tilde{s}_{t+1}. \tag{55}$$

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:

$$\tilde{s}_{t+1} = g_0 + g_1 z_{c,t+1} - z_{c,t} + \Delta\tilde{c}_{t+1}. \tag{56}$$

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:

$$z_{c,t} = A_{c,0} + A_{c,1}x_t + A_{c,2}\sigma_t^2. \tag{57}$$

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

$$\begin{aligned} E_t\left[m_{t+1} + \tilde{s}_{t+1}\right] =& \zeta \log(\delta) + \left(\zeta - \frac{\zeta}{\psi}\right)(\mu_c + \gamma x_t) + \zeta g_0 + \zeta\left(g_1 - 1\right)A_{c,0} + \zeta\left(g_1\rho - 1\right)x_t \\ &+ \zeta\left(g_1\varrho - 1\right)A_{c,2}\sigma_t^2 + \zeta g_1\left(1 - \varrho\right)A_{c,2}\sigma_c^2, \end{aligned}$$

$$var_t\left(m_{t+1} + \tilde{s}_{t+1}\right) = \zeta^2\left(1 - \frac{1}{\psi}\right)^2 \sigma_t^2 + \zeta^2\left(g_1 A_{c,1}\sigma_x\right)^2 + \zeta^2\left(g_1 A_{c,2}\sigma_\varsigma\right)^2. \tag{58}$$

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

$$\begin{aligned} A_{c,0} &= \frac{\log(\delta) + (1 - \frac{1}{\psi})(\mu_d - \mu_\pi) + g_0 + g_1 A_{c,2}(1-\varrho)\sigma_c^2 + \frac{1}{2}\zeta g_1^2 (A_{c,1}^2 \sigma_x^2 + A_{c,2}^2 \sigma_\varsigma^2)}{1 - g_1}, \\ A_{c,1} &= \frac{\left(1 - \frac{1}{\psi}\right)\gamma}{1 - g_1\rho}, \quad A_{c,2} = \frac{\zeta(1 - \frac{1}{\psi})^2}{2(1 - g_1\varrho)}. \end{aligned} \tag{59}$$

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

$$\begin{aligned} r_{t+1} &= \kappa_0 + \kappa_1 z_{d,t+1} - z_{d,t} + \Delta d_{t+1}, \\ \tilde{r}_{t+1} &= \kappa_0 + \kappa_1 z_{d,t+1} - z_{d,t} + \Delta \tilde{d}_{t+1}. \end{aligned} \tag{60}$$

where $\Delta\tilde{d}_{t+1}$ is real dividend growth rate. Assume that log price-to-dividend ratio is of the form:

$$z_{d,t} = A_{d,0} + A_{d,1}x_t + A_{d,2}\sigma_t^2 + A_{d,3}(q_t - \mu_q). \tag{61}$$

Then note that:

$$\begin{aligned} E_t\left[m_{t+1} + \tilde{r}_{t+1}\right] =& \zeta \log(\delta) + (\zeta - 1)(g_1 - 1) A_{c,0} + (\zeta - 1)(g_1\rho - 1) A_{c,1}x_t + (\zeta - 1)(g_1\varrho - 1) A_{c,2}\sigma_t^2 \\ &+ g_1(1-\varrho) A_{d,2}\sigma_c^2 + \left(\zeta - \frac{\zeta}{\psi} - 1\right)(\mu_c + \gamma x_t) + (\zeta - 1) g_0 + \kappa_0 + (\kappa_1 - 1) A_0 \\ &+ (\kappa_1\rho - 1) A_{d,1}x_t + (\kappa_1\varrho - 1) A_{d,2}\sigma_t^2 + \kappa_1(1-\varrho)A_{d,2}\sigma_c^2 \\ &+ (\kappa_1\theta - 1) A_{d,3}\left(q_t - \mu_q\right) + \mu_c + \frac{1}{1-\phi}x_t + \frac{\varphi - (1-\theta)\phi}{1-\phi}\left(q_t - \mu_q\right). \end{aligned}$$

$$\begin{aligned} var_t\left(m_{t+1} + \tilde{r}_{t+1}\right) =& \left(\zeta - 1 - \frac{\zeta}{\psi}\right)^2 \sigma_t^2 + \left(\frac{1}{1-\phi}\right)^2 \sigma_d^2 + \left((\zeta-1)g_1A_{c,1} + \kappa_1 A_{d,1}\right)^2 \sigma_x^2 \\ &+ \left((\zeta-1)g_1A_{c,2} + \kappa_1 A_{d,2}\right)^2 \sigma_\varsigma^2 + \left(\kappa_1 A_{d,3} + \frac{\phi}{1-\phi}\right)^2 \sigma_q^2 \\ &+ 2\left((\zeta-1)g_1A_{c,1} + \kappa_1 A_{d,1}\right)\left(\kappa_1 A_{d,3} + \frac{\phi}{1-\phi}\right)\lambda_{xq}\sigma_x\sigma_q \\ &+ \frac{2}{1-\phi}\left(\kappa_1 A_{d,3} + \frac{\phi}{1-\phi}\right)\lambda_{dq}\sigma_d\sigma_q + \frac{2}{1-\phi}\left((\zeta-1)g_1A_{c,1} + \kappa_1 A_{d,1}\right)\lambda_{dx}\sigma_d\sigma_x. \end{aligned} \tag{62}$$

Using 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:

$$A_{d,0} = \frac{\left(\begin{array}{c} \zeta \log(\delta) + (\zeta-1)g_0 + (\zeta-1)A_{c,0}(g_1-1) + ((\zeta-1)g_1 A_{c,2} + \kappa_1 A_{d,2})(1-\varrho)\sigma_c^2 \\ +\left(\zeta - \frac{\zeta}{\psi} - 1\right)\mu_c + \kappa_0 + \mu_c + \frac{1}{2}(\frac{1}{1-\phi})^2\sigma_d^2 + \frac{1}{2}((\zeta-1)g_1 A_{c,1} + \kappa_1 A_{d,1})^2\sigma_x^2 \\ +\frac{1}{2}((\zeta-1)g_1 A_{c,2} + \kappa_1 A_{d,2})^2\sigma_\varsigma^2 + \frac{1}{2}(\kappa_1 A_{d,3} + \frac{\phi}{1-\phi})^2\sigma_q^2 \\ +((\zeta-1)g_1 A_{c,1} + \kappa_1 A_{d,1})\left(\kappa_1 A_{d,3} + \frac{\phi}{1-\phi}\right)\lambda_{xq}\sigma_x\sigma_q \\ +\frac{1}{1-\phi}\left(\kappa_1 A_{d,3} + \frac{\phi}{1-\phi}\right)\lambda_{dq}\sigma_d\sigma_q + \frac{1}{1-\phi}((\zeta-1)g_1 A_{c,1} + \kappa_1 A_{d,1})\lambda_{dx}\sigma_d\sigma_x \end{array}\right)}{1-\kappa_1},$$

$$A_{d,1} = \frac{\left(\zeta - 1 - \frac{\zeta}{\psi}\right)\gamma + (\zeta-1)A_{c,1}(g_1\rho - 1) + \frac{1}{1-\phi}}{1-\kappa_1\rho},$$

$$A_{d,2} = \frac{(\zeta-1)(g_1\varrho - 1)A_{c,2} + \frac{1}{2}\left(\zeta - 1 - \frac{\zeta}{\psi}\right)^2}{1-\kappa_1\varrho}, \; A_{d,3} = \frac{\psi - (1-\theta)\phi}{(1-\kappa_1\theta)(1-\phi)}. \tag{63}$$

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

$$E_t[r_{t+1}] = A_{r,0} + A_{r,1}x_t + A_{r,2}\sigma_t^2 + A_{r,3}(q_t - \mu_q), \tag{64}$$

where:

$$A_{r,0} = \kappa_0 - (1-\kappa_1)A_{d,0} + \mu_d + \kappa_1(1-\varrho)A_{d,2}\sigma_c^2, \quad A_{r,1} = \frac{1}{1-\phi} - (1-\kappa_1\rho)A_{d,1},$$
$$A_{r,2} = -(1-\kappa_1\varrho)A_{d,2}, \quad A_{r,3} = \frac{\varphi - (1-\theta)\phi}{1-\phi} - (1-\kappa_1\theta)A_{d,3} = 0. \tag{65}$$

Expected return over the next τ period is:

$$\sum_{s=0}^{\tau-1} r_{t+s+1} = sA_{r,0} + \left(\sum_{s=1}^{\tau-1} A_{r,1}\rho^s\right)x_t + \left(\sum_{s=1}^{\tau-1} A_{r,2}\varrho^s\right)\sigma_t^2 + \left(\sum_{s=1}^{\tau-1} A_{r,2}(1-\varrho^s)\right)\sigma_c^2. \tag{66}$$

Finally, the risk free rate can be written as:

$$r_{f,t+1} = A_{f,0} + A_{f,1}x_t + A_{f,2}\sigma_t^2, \tag{67}$$

where:

$$A_{f,0} = -\zeta\log(\delta) - (\zeta - 1 - \frac{\zeta}{\psi})\mu_c - (\zeta-1)(g_0 + (g_1-1)C_0 + g_1(1-\varrho)C_2\sigma_c^2 \\ -\frac{1}{2}\left((\zeta-1)g_1C_1\right)^2\sigma_x^2 - \frac{1}{2}\left((\zeta-1)g_1C_2\sigma_\varsigma\right)^2.$$

$$A_{f,1} = -(\zeta-1-\frac{\zeta}{\psi})\gamma + (\zeta-1)(1-g_1\rho)C_1,\; A_{f,2} = (\zeta-1)(1-g_1\varrho)C_2 - \frac{1}{2}(\zeta-1-\frac{\zeta}{\psi})^2. \quad (68)$$

A.1.3 Full Model (Second Specification)

We derive price-to-dividend ratios and expected returns implied by our long-run risk model, which features dividend dynamics in (6), consumption dynamics in (37), and investors preferences in (22). Our derivation closely follows the steps in Bansal and Yaron (2004). The log stochastic discount factor is given as:

$$m_{t+1} = \zeta\log(\delta) - \frac{\zeta}{\psi}\Delta\tilde{c}_{t+1} + (\zeta-1)\tilde{s}_{t+1}. \quad (69)$$

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:

$$\tilde{s}_{t+1} = g_0 + g_1 z_{c,t+1} - z_{c,t} + \Delta\tilde{c}_{t+1}. \quad (70)$$

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:

$$z_{c,t} = A_{c,0} + A_{c,1}x_t. \quad (71)$$

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

$$E_t\left[m_{t+1} + \tilde{s}_{t+1}\right] = \zeta\log(\delta) + \left(\zeta - \frac{\zeta}{\psi}\right)(\mu_c + \gamma x_t) + \zeta g_0 + \zeta\left(g_1 - 1\right)A_{c,0} + \zeta\left(g_1\rho - 1\right)x_t,$$

$$var_t\left(m_{t+1} + \tilde{s}_{t+1}\right) = \zeta^2\left(1 - \frac{1}{\psi}\right)^2\sigma_c^2 + \zeta^2\left(g_1A_{c,1}\sigma_x\right)^2. \quad (72)$$

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

$$\begin{aligned} A_{c,0} &= \frac{\log(\delta) + (1-\frac{1}{\psi})\mu_c + g_0 + \frac{1}{2}\zeta(1-\frac{1}{\psi})^2\sigma_c^2 + \frac{1}{2}\zeta g_1^2(A_{c,1}^2\sigma_x^2 + A_{c,2}^2\sigma_\varsigma^2)}{1-g_1}, \\ A_{c,1} &= \frac{\left(1-\frac{1}{\psi}\right)\gamma}{1-g_1\rho}. \end{aligned} \tag{73}$$

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

$$\begin{aligned} r_{t+1} &= \kappa_0 + \kappa_1 z_{d,t+1} - z_{d,t} + \Delta d_{t+1}, \\ \tilde{r}_{t+1} &= \kappa_0 + \kappa_1 z_{d,t+1} - z_{d,t} + \Delta\tilde{d}_{t+1}. \end{aligned} \tag{74}$$

where $\Delta\tilde{d}_{t+1}$ is real dividend growth rate. Assume that log price-to-dividend ratio is of the form:

$$z_{d,t} = A_{d,0} + A_{d,1}x_t + A_{d,2}\lambda_t + A_{d,3}(q_t - \mu_q). \tag{75}$$

Then note that:

$$\begin{aligned} E_t\left[m_{t+1} + \tilde{r}_{t+1}\right] =& \zeta\log(\delta) + (\zeta-1)(g_1-1)A_{c,0} + (\zeta-1)(g_1\rho-1)A_{c,1}x_t + \mu_c \\ &+ \left(\zeta - \frac{\zeta}{\psi} - 1\right)(\mu_c + \gamma x_t) + (\zeta-1)g_0 + \kappa_0 + (\kappa_1-1)A_0 + (\kappa_1\rho-1)A_{d,1}x_t \\ &+ (\kappa_1\varrho-1)A_{d,2}\lambda_t + (\kappa_1\theta-1)A_{d,3}\left(q_t-\mu_q\right) + \frac{1}{1-\phi}x_t + \frac{\varphi-(1-\theta)\phi}{1-\phi}\left(q_t-\mu_q\right). \end{aligned}$$

$$\begin{aligned} var_t\left(m_{t+1} + \tilde{r}_{t+1}\right) =& \left(\zeta - 1 - \frac{\zeta}{\psi}\right)^2\sigma_c^2 + \left(\frac{1}{1-\phi}\right)^2\sigma_d^2 + \left((\zeta-1)g_1A_{c,1} + \kappa_1A_{d,1}\right)^2\sigma_x^2 \\ &+ (\kappa_1A_2)^2\sigma_\varsigma^2 + \left(\kappa_1A_{d,3} + \frac{\phi}{1-\phi}\right)^2\sigma_q^2 + 2\left(\kappa_1A_{d,3} + \frac{\phi}{1-\phi}\right)\left(\zeta - 1 - \frac{\zeta}{\psi}\right)\sigma_x\sigma_q\lambda_t \\ &+ 2\left((\zeta-1)g_1A_{c,1} + \kappa_1A_{d,1}\right)\left(\kappa_1A_{d,3} + \frac{\phi}{1-\phi}\right)\lambda_{xq}\sigma_x\sigma_q \\ &+ \frac{2}{1-\phi}\left(\kappa_1A_{d,3} + \frac{\phi}{1-\phi}\right)\lambda_{dq}\sigma_d\sigma_q + \frac{2}{1-\phi}\left((\zeta-1)g_1A_{c,1} + \kappa_1A_{d,1}\right)\lambda_{dx}\sigma_d\sigma_x. \end{aligned} \tag{76}$$

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

$$
\begin{aligned}
A_{d,0} &= \frac{\left(\begin{array}{c} \zeta\log(\delta)+(\zeta-1)g_0+(\zeta-1)A_{c,0}(g_1-1) \\ +\left(\zeta-\frac{\zeta}{\psi}-1\right)\mu_c+\kappa_0+\mu_c+\frac{1}{2}(\frac{1}{1-\phi})^2\sigma_d^2+\frac{1}{2}((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1})^2\sigma_x^2 \\ +\frac{1}{2}(\kappa_1A_{d,2})^2\sigma_\lambda^2+\frac{1}{2}(\kappa_1A_{d,3}+\frac{\phi}{1-\phi})^2\sigma_q^2 \\ +((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1})\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{xq}\sigma_x\sigma_q \\ +\frac{1}{1-\phi}\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{dq}\sigma_d\sigma_q+\frac{1}{1-\phi}((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1})\lambda_{dx}\sigma_d\sigma_x \end{array}\right)}{1-\kappa_1}, \\
A_{d,1} &= \frac{\left(\zeta-1-\frac{\zeta}{\psi}\right)\gamma+(\zeta-1)A_{c,1}(g_1\rho-1)+\frac{1}{1-\phi}}{1-\kappa_1\rho}, \\
A_{d,2} &= \frac{\frac{1}{1-\phi}\left(\zeta-1-\frac{\zeta}{\psi}\right)\sigma_c\sigma_d}{1-\kappa_1\varrho},\ A_{d,3}=\frac{\psi-(1-\theta)\phi}{(1-\kappa_1\theta)(1-\phi)}.
\end{aligned}
\tag{77}
$$

Substituting the expression for $z_{d,t}$ into $r_{t+1}=\kappa_0+\kappa_1z_{d,t+1}-z_{d,t}+\Delta d_{t+1}$ leads:

$$
E_t[r_{t+1}]=A_{r,0}+A_{r,1}x_t+A_{r,2}\lambda_t+A_{r,3}(q_t-\mu_q), \tag{78}
$$

where:

$$
\begin{aligned}
&A_{r,0}=\kappa_0-(1-\kappa_1)A_{d,0}+\mu_d, \quad A_{r,1}=\frac{1}{1-\phi}-(1-\kappa_1\rho)A_{d,1}, \\
&A_{r,2}=-(1-\kappa_1\varrho)A_{d,2}, \quad A_{r,3}=\frac{\varphi-(1-\theta)\phi}{1-\phi}-(1-\kappa_1\theta)A_{d,3}=0.
\end{aligned}
\tag{79}
$$

Expected return over the next τ period is:

$$
\sum_{s=0}^{\tau-1}r_{t+s+1}=sA_{r,0}+\left(\sum_{s=1}^{\tau-1}A_{r,1}\rho^s\right)x_t+\left(\sum_{s=1}^{\tau-1}A_{r,2}\varrho^s\right)\lambda_t+\left(\sum_{s=1}^{\tau-1}A_{r,2}(1-\varrho^s)\right)\mu_\lambda. \tag{80}
$$

Finally, the risk free rate can be written as:

$$
r_{f,t+1}=A_{f,0}+A_{f,1}x_t, \tag{81}
$$

where:

$$
\begin{aligned}
&A_{f,0}=-\zeta\log(\delta)-(\zeta-1-\frac{\zeta}{\psi})\mu_c-(\zeta-1)(g_0+(g_1-1)C_0-\frac{1}{2}\left(\zeta-1-\frac{\zeta}{\psi}\right)^2\sigma_c^2-\frac{1}{2}((\zeta-1)g_1C_1)^2\sigma_x^2. \\
&A_{f,1}=-(\zeta-1-\frac{\zeta}{\psi})\gamma+(\zeta-1)(1-g_1\rho)C_1.
\end{aligned}
\tag{82}
$$

A.2 Kalman Filter

We describe the Kalman filtering process for estimating the system of equations in (6). First, note the last equation in (6) can be estimated separately from other equations in (6) using time series regression. To estimate the first two equations, define $x'_t = x_{t-1}$ and $\epsilon'_{x,t+1} = \epsilon_{x,t}$, and re-write the remaining system of equations as:

$$\begin{aligned} \Delta d_{t+1} &= \mu_d + x'_{t+1} + \phi(\Delta e_{t+1} - \mu_d) + \varphi\left(q_t - \mu_q\right) + \sigma_d \epsilon_{d,t+1} \\ x'_{t+1} &= \rho x'_t + \sigma_x \epsilon'_{x,t+1} \\ \begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon'_{x,t+1} \end{pmatrix} &\sim \text{ i.i.d. } \mathbb{N}\left(\mathbf{0}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right). \end{aligned} \tag{83}$$

To apply the Kalman filter, let $x'_{t|s}$ denote the time-s expectation of the latent variable x'_t and $P'_{t|s}$ denote the variance of x'_t conditioning on information in time-s. Set initial conditions $x'_{0|0} = 0$ and $P'_{0|0} = \frac{\sigma_x^2}{1-\rho^2}$. We can then iterate the following system of equations:

$$\begin{aligned} x'_{t+1|t} &= \rho x'_{t|t}, \quad P'_{t+1|t} = \rho^2 P'_{t|t} + \sigma_x^2, \\ e_{t+1} &= \Delta d_{t+1} - \mu_d - \phi(\Delta e_{t+1} - \mu_d) - \varphi\left(q_t - \mu_q\right), \\ x'_{t+1|t+1} &= x'_{t+1|t} + \frac{P'_{t+1|t}}{P'_{t+1|t} + \sigma_d^2} e_{t+1}, \quad P'_{t+1|t+1} = P'_{t+1|t} - \frac{P^2_{t+1|t}}{P_{t+1|t} + \sigma_d^2}. \end{aligned} \tag{84}$$

To estimate dividend model parameters using data up to time-τ, define the log likelihood function:

$$\mathcal{L} = -\sum_{t=0}^{\tau-1} \left(\log\left(P'_{t+1|t} + \sigma_d^2\right) + \frac{e_{t+1}^2}{P'_{t+1|t} + \sigma_d^2} \right),$$

We note 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 average of the twelve log likelihood.

A.3 Bootstrap Simulation

Simulation is based on 100,000 iterations. First we simulate innovations to dividend growth rates and earnings-to-dividend ratios:

$$\begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{q,t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \lambda_{dx} & \lambda_{dq} \\ \lambda_{dx} & 1 & \lambda_{xq} \\ \lambda_{dq} & \lambda_{xq} & 1 \end{pmatrix}\right). \tag{85}$$

Dividend model parameters used for simulations are those reported in Table 2, which are estimated based on the full data sample between 1946 and 2013. From these innovations, we can simulate the latent variable x_t and earnings-to-dividend ratios iteratively as:

$$x_{t+1} = \rho x_t + \sigma_x \epsilon_{x,t+1}$$

$$q_{t+1} = \mu_q + \theta \left(q_t - \mu_q\right) + \sigma_q \epsilon_{q,t+1}. \tag{86}$$

Given the simulated time series of x_t and earnings-to-dividend ratios, we can simulate dividend and earnings growth rates iteratively as:

$$\Delta d_{t+1} = \mu_d + \frac{1}{1-\phi}\left(x_t + \phi(\Delta q_{t+1} - \mu_q) + (\varphi - \phi)\left(q_t - \mu_q\right) + \phi\sigma_q\epsilon_{q,t+1} + \sigma_d\epsilon_{d,t+1}\right)$$

$$\Delta e_{t+1} = q_{t+1} - q_t + \Delta d_{t+1}. \tag{87}$$

A.4 Modeling Real Dividend Growth Rates

We report estimated coefficients from Campbell and Shiller (1988b) regressions in (3) and (5), but replacing nominal dividend growth rates using real dividend grwoth rates instead. We report regression statistics in Table 14. We find that, as is the case with nominal rates, real dividend growth rates over the next year increase with price-to-dividend ratios, and decrease with CAPE ratios. Further, we cannot statistically reject $b_2 = -b_3$. Setting $b_2 + b_3 = 0$, we find that earnings-to-dividend ratios have a significant positive effect on future real dividend growth rates. These results are consistent with results based on nominal dividend growth rates.

We then apply our dividend model in (6) to the real dividend growth rate process, and compare the model's performance in fitting annual real dividend growth rates in-sample and predicting these rates out-of-sample to the performance of the baseline models. Based on data between 1976 and 2013, we find that, as is the case with nominal rates, our model clearly outperforms its baseline models in explaining real dividend growth rates

b_{10}	b_{11}	b_{12}	b_{13}	β_0	β_1	β_2
-0.118*	0.496***	0.098**	-0.077*	-0.046	0.509***	0.083**
(0.063)	(0.118)	(0.041)	(0.039)	(0.029)	(0.119)	(0.040)

Table 14: **Campbell and Shiller (1988b) Betas for Predicting Dividend Growth Rates (Real Rates)**: This table reports coefficients from estimating dividend growth rates using (3) and (5), based on data between 1946 and 2013. 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 $***$.

and predicting these rates out-of-sample, between 1976 and 2013.

	In-Sample		Out-of-Sample	
	Goodness-of-Fit	BIC	R^2	p-value
J&L	0.482	-5.974	0.275	0.001
vB&K	0.203	-5.544	0.166	0.012
C&S	0.271	-5.737	0.135	0.025
C&S (Restricted)	0.251	-5.709	0.198	0.006

Table 15: **Dividend Growth Rates and Expected Growth Rates (Real Rates)**. The first column of this table reports goodness-of-fit for describing dividend growth rates using our dividend model (i.e. J&L), the dividend model in van Binsbergen and Koijen (2010) (i.e. vB&K), the dividend model in Campbell and Shiller (1988b) (i.e. C&S), or its restricted version where we set $b_{12} + b_{13} = 0$. The second column reports the Bayesian information criterion. The third and fourth columns report the 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). In-sample (out-of-sample) statistics are based on data between 1946 and 2013 (1976 and 2013).

A.5 Timing of Investors Receiving Earnings Information

Throughout this paper, we assume that investors receive earnings information 3 months after fiscal quarter or year end. To show that our findings are robust to this assumption, we repeat results in Tables 6 and 9, assuming that investors instead receive earnings information 6, 9, or 12 months after fiscal quarter or year end. We report these results in Tables 16 and 17. We note that changing this assumption can affect our results through its effect on long run dividend growth expectations and investors' beliefs about the persistence

ω of dividend growth rates, both computed using dividend model parameters estimated at each point in time based on data available at the time. Nevertheless, results show that the significance of our findings that investors' learning about dividend dynamics is reflected in the returns of the stock index is robust to changes in this assumption on when investors receive earnings information.

	3 Months Lag	6 Months Lag	9 Months Lag	12 Months Lag
sy_t (Learning)	3.964*** (1.133)	3.771*** (1.114)	3.603*** (1.110)	3.554*** (1.103)
R^2	0.152	0.145	0.138	0.137

Table 16: **Stock Index Returns and Stock Yields (Timing of Investors Receiving Earnings Information)**: This table reports coefficient estimates and R-square value from regressing stock index returns over the next year on stock yields, computed assuming investors learn about dividend dynamics using our dividend model. Regression is based on data between 1976 and 2013. Dividend model parameters are estimated based on data since 1946. To estimate dividend dynamics, we assume that investors receive earnings information 3, 6, 9, or 12 months after fiscal quarter or year end. Newey and West (1987) standard errors are reported in parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.

A.6 Estimating Dividend Dynamics using a Rolling Window of Dividend Data

We show that estimating dividend model parameters based on an expanding window of past dividend data performs better than estimating those parameters based on a rolling window of past data, for the purposes of forecasting both future dividends and stock index returns. We note that expanding window refers to estimating model parameters based on all past data since 1946, and rolling window refers to estimating those parameters based on only the last h years of past data. We set h to 10, 20, or 30 years. In Table 19, we report the out-of-sample R^2 value for predicting annual dividend rates using expected dividend growth rates implied by our model, with model parameters estimated using a rolling window of h years of past data. Results confirm that the out-of-sample R-square value for predicting dividend growth rates is highest when model parameters are estimated based on an expanding window of past data. In absolute terms, however, the out-of-sample R-square value is still 27.0 (28.6) percent when parameters are estimated based on a rolling

	3 Months Lag	6 Months Lag	9 Months Lag	12 Months Lag
$\omega(t)$	-0.562**	-0.501**	-0.436***	-0.371**
	(0.129)	(0.135)	(0.149)	(0.166)
sy_t	5.545***	5.002***	4.551***	4.301***
(Learning)	(0.897)	(0.856)	(0.908)	(0.972)
R^2	0.265	0.236	0.205	0.186

Table 17: **Stock Index Returns, Stock Yields, and Investors' Beliefs about Persistence of Dividend Growth Rates (Timing of Investors Receiving Earnings Information)**. This table reports the coefficient estimates and R-square value from regressing stock index returns over the next year on investors' beliefs about the persistence ω of dividend growth rates and stock yields, computed assuming investors learn about dividend dynamics using our model. Regression is based on data between 1976 and 2013. Dividend model parameters are estimated based on data since 1946. To estimate dividend dynamics, we assume that investors receive earnings information 3, 6, 9, or 12 months after fiscal quarter or year end. Newey and West (1987) standard errors are reported in parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.

window of 20 (30) years of past data. This shows that our model's success in forecasting dividend growth rates is robust to how we set the training period.

We then show that estimating model parameters using an expanding window, rather than a rolling window, of past data also best captures investors' learning behavior. In Table 18, we repeat key results in Tables 6 and 9, but compute long run dividend growth expectations and the persistence ω of dividend growth rates using model parameters estimated at each point in time based on a rolling window of h years of past data. Results confirm that investor's learning behavior is best captured when those parameters are estimated based on an expanding window of past data.

	10 Years	20 Years	30 Years
R^2	0.083	0.270	0.286
	[0.085]	[0.001]	[0.001]

Table 18: **Dividend Growth Rates and Expected Growth Rates (Rolling Window)**: This table reports the out-of-sample R-square value for predicting dividend growth rates using our dividend model. Also reported in square parentice is the corresponding p-value from the adjusted-MSPE statistics of Clark and West (2007). Statistics are based on data between 1975 and 2013. Dividend model parameters are estimated based on a rolling window of past 10, 20, or 30 years of data.

	10 Years		20 Years		30 Years	
$\omega(t)$		-0.081		-0.150		-0.414***
		(0.084)		(0.091)		(0.136)
sy_t	1.159	1.089	2.579***	2.444***	3.209***	3.811***
(Learning)	(0.764)	(0.731)	(0.613)	(0.514)	(0.954)	(0.632)
R^2	0.048	0.060	0.164	0.184	0.135	0.206

Table 19: **Stock Index Returns, Stock Yields, and Investors' Beliefs about the Persistence of Dividend Growth Rates (Rolling Window)**: This table reports coefficient estimates and R-square value from regressing stock index returns over the next year on investors' beliefs about the persistence ω of dividend growth rates and stock yields, computed assuming investors learn about dividend dynamics using our model. Regression is based on data between 1976 and 2013. Dividend model parameters are estimated based on a rolling window of past 10, 20, or 30 years of data. Newey and West (1987) standard errors are reported in parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.