NBER WORKING PAPER SERIES

# DIVIDEND DYNAMICS, LEARNING, AND EXPECTED STOCK INDEX RETURNS

Ravi Jagannathan
Binying Liu

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

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

We thank Jonathan Berk, Jules van Binsbergen, Narayan Naik, and seminar participants at London Business School, for helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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

**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. When investors' beliefs about the persistence of dividend growth rates increase, dividend-to-price ratios increase, and short-horizon expected returns decrease after controlling for dividend-to-price ratios. These findings are consistent with investors' preferences for early resolution of uncertainty. We embed learning about dividend dynamics in an equilibrium asset pricing model. The model predicts 22.8% of the variation in annual stock index returns. Learning accounts for forty-percent of that 22.8%.

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

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

The average return on equities has been substantially higher than the average return on risk free bonds over long periods of time. Between 1946 and 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 stochastic and persistent manner. A large number of academic studies have found support for such predictable variation in the equity premium both in time series and in cross-section.[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 macroeconomic outcomes change over time and drive the variation in stock index prices and expected returns, they have full knowledge of 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 emprical evidence that investors learn and that changes in investors' beliefs about parameters describing the economy is reflected in asset prices and expected returns. Further, we show that the way asset prices and expected 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 about dividend dynamics affect asset prices and expected returns, we need a realistic

[1] See, among others, Fama and French (1988), Campbell and Shiller (1988b), Fama and French (1993), Jegadeesh and Titman (1993), Lamont (1998), Baker and Wurgler (2000), Lettau and Ludvigson (2001), Campbell and Vuolteenaho (2004), Lettau and Ludvigson (2005), Polk, Thompson, 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 (2014), Dufresne, Johannes, Lochstoer (2014), Giacoletti, Laursen, and Singleton (2014).

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 growth rates in aggregate dividends 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. Between 1946 and 2013, our model explains 55.1 percent of the variation in annual dividend growth rates in-sample and predicts 31.3 percent of the variation in annual dividend growth rates 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 the 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 avaliable at the time, model parameter estimates fluctuate, some significantly, over time as more data become avaliable. In other words, if investors estimate dividend dynamics using our model, we expect their beliefs about 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 dividend growth rates. Through this channel, changes in investors' beliefs about parameters governing the dividend process can contribute significantly to the variation in discount rates.

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 annual stock index returns. In comparison, stock yields, computed assuming that expected dividend growth rates are constant, explain only 10.2 percent of the variation in annual stock index returns. 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 asset 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' preference for early resolution of uncertainty.

We embed our dividend model into an equilibrium asset pricing model that features Epstein and Zin (1989) preferences 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 have to learn about parameters governing the dividend process, predict 22.8 percent of the variation in annual stock index returns, and learning accounts for forty-percent of the 22.8 percent. We decompose the variation in price-to-dividend ratios and find that 27.9 percent of the variation in price-to-dividend ratios is due to investors' learning about dividend dynamics.

We follow Cogley and Sargent (2009), Piazzesi and Schneider (2010), Giacoletti, Laursen, and Singleton (2014), and Johannes, Lochstoer, and Mou (2014), and define learning based on 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. Therefore, we adopt the anticipated utility approach as the more realistic 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 discuss how learning about dividend dynamics affects expectations of future dividends. In Section 3, we show that learning about dividend dynamics is reflected in prices and expected returns of the stock index. In Section 4, we argue that the way discount rates covary with investors beliefs about persistence of dividend growth rates supports investors' preference 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. 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 $x_t$, its unconditional mean $\mu_d$, and its persistence coefficient $\rho$, as follows:

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

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

$$\begin{pmatrix} \Delta d_{t+1} \\ p_{t+1} - d_{t+1} \\ p_{t+1} - \bar{e}_{t+1} \end{pmatrix} = \begin{pmatrix} b_{10} \\ b_{20} \\ b_{30} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} \Delta d_t \\ p_t - d_t \\ p_t - \bar{e}_t \end{pmatrix} + \begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{(p-d),t+1} \\ \epsilon_{(p-\bar{e}),t+1} \end{pmatrix}$$

$$\begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{(p-d),t+1} \\ \epsilon_{(p-\bar{e}),t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N} \left( \mathbf{0}, \begin{pmatrix} 1 & \lambda_{12} & \lambda_{13} \\ \lambda_{12} & 1 & \lambda_{23} \\ \lambda_{13} & \lambda_{23} & 1 \end{pmatrix} \right). \tag{3}$$

where, as in Campbell and Shiller (1998b), CAPE ratio is defined as:

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

Estimates of $b_{10}$, $b_{11}$, $b_{12}$, and $b_{13}$ from (3), based on data between 1946 and 2013, are reported in the first four columns of Table 1.[3]

Consistent with Campbell and Shiller (1988b), we find that both price-to-dividend ratios and CAPE ratios have significant effects on future dividend growth rates, but in the opposite direction. That is, increases in price-to-dividend ratios predict decreases in future dividend growth rates, but increases in CAPE ratios predict increases in future dividend growth rates. Moreover, it is clear from these coefficient estimates that $b_{12} + b_{13} = 0$ cannot be statistically rejected. For this reason, we restrict $b_{13} = -b_{12}$ and re-estimate dividend growth rates as:

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

We report estimated coefficients from (5) in the last three columns of Table 1. Results show that the $\beta_2$ 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 implies that firms have been retaining earnings in the past and are therefore expected to pay more dividends in the future.

We extend the latent variable model of van Binsbergen and Koijen (2010) based on this insight that earnings contain information about future dividends. Let $\Delta e_{t+1} = e_{t+1} - e_t$ be

[3] 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. Because overlapping monthly data are used, we report standard errors, $F$-statistics, $p$-values, and $Q$-statistics adjusted to reflect the dependence introduced by overlapping data.

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 fluctuation 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 $\mu_d$. 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:

$$\Delta e_{t+1} = \mu_d + \frac{1}{1-\phi}\left(x_t + (\varphi + \theta - 1)(q_t - \mu_q) + \sigma_d \epsilon_{d,t+1} + \sigma_q \epsilon_{q,t+1}\right). \tag{7}$$

In our model, the persistence of dividend growth rates are determined by two parameters: coefficient $\rho$ which governs the persistence of $x_t$, and coefficient $\theta$ which governs the persistence of earnings-to-dividend ratios. We can solve for expected dividend growth rates in our model as:

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

Aside from the two state variables $x_t$ and $q_t$ and their persistence coefficients $\rho$ and $\theta$, expected dividend growth rates are also functions of the unconditional means $\mu_d$ and $\mu_q$, and coefficients $\phi$ and $\varphi$ that connect earnings information to dividend dynamics.

| $b_{10}$ | $b_{11}$ | $b_{12}$ | $b_{13}$ | $\beta_0$ | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|---|---|
| -0.058 | 0.378*** | 0.147*** | -0.106*** | -0.043** | 0.397*** | 0.136*** |
| (0.056) | (0.075) | (0.034) | (0.037) | (0.019) | (0.075) | (0.034) |

Table 1: **Campbell and Shiller (1988b) Betas for Predicting Dividend Growth Rates**: This table reports coefficients from estimating dividend growth rate using regressions in (3) and (5), based on data between 1946 and 2013. Standard errors are reported in parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.

## 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 its proxy. 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. 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 $\phi$ and $\varphi$ 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 presistent 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.

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

| | | | | | |
|---|---|---|---|---|---|
| $\mu_d$ | $\phi$ | $\varphi$ | $\sigma_d$ | $\rho$ | $\sigma_x$ |
| 0.059 | 0.079 | 0.184 | 0.017 | 0.528 | 0.041 |
| (0.015) | (0.018) | (0.028) | (0.013) | (0.160) | (0.009) |
| $\mu_q$ | $\theta$ | $\sigma_q$ | $\lambda_{dx}$ | $\lambda_{dq}$ | $\lambda_{xq}$ |
| 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. Simulated standard errors are reported in parentice.

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.[4] 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 avaliable at the time. The model's forecasting performance

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

is then evaluated using the out-of-sample $R$-square value defined as:

$$R^2(\Delta d_{t+1}) = 1 - \frac{\sum_{t=T_0}^{T-1}\left(\Delta d_{t+1} - E_t[\Delta d_{t+1}]\right)^2}{\sum_{t=T_0}^{T-1}\left(\Delta d_{t+1} - \underline{\Delta d}_t\right)^2}. \tag{9}$$

where $\underline{\Delta d}_t = \frac{1}{t}\sum_{s=0}^{t-1}\Delta d_{s+1}$ is the historical average of dividend growth rates, $T$ denotes the end of the data sample, and $T_0$ denotes the end of the training period. Given 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 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.[5] 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.[6] 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 $F$-test for model significance. Results show that our model predicts 31.3 percent of the variation in annual dividend growth rates, which represent a significant improvement over the 18.5 percent and 13.5 percent from the baseline models.

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. For example, one can add additional lags of earnings-to-dividend ratios to the model.[7] Also, one can extend our model by allowing model parameters, such as the persistence $\rho$ of the latent variable $x_t$ or the standard deviation $\sigma_x$ of shocks to $x_t$, to be time varying. However, the disadvantage of incoporating 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 using a rolling

[5] 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.

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

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

| | $\Delta d_{t+1}$ | J&L | $\Delta d_{t+1} - E_t[\Delta d_{t+1}]$ 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: **Summary Statistics on Dividend Growth Rates and Expected Rates**: This table reports the 1, 2, 3, 4, and 5 years serial correlations for dividend growth rates and growth rate residuals based on our dividend model (J&L), the dividend model in van Binsbergen and Koijen (2012) and the dividend model in Campbell and Shiller (1988b). Also reported are the Ljung-Box (1973) $Q$-statistics for testing if dividend growth rates and growth rate residuals are serially correlated. Estimating dividend dynamics is based on data between 1946 and 2013. $p$-values for $Q$-statistics are reported in square parentice.

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 Appendix. In summary, we find that our model's forecasting performance is highest when model parameters are estimated using an expanding window, not a rolling window, of past data.

# 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 an investor were to estimate our model in an earlier date, she 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 avaliable at the time. In this section, we summarize how learning affects investors' beliefs about parameters governing the dividend process, assuming that investors learn about dividend dynamics using our

| | 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 | 0.248 | -5.683 | 0.135 | 0.025 |

Table 4: **Dividend Growth Rates and Model Implied 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), or the dividend model in Campbell and Shiller (1988b) (i.e. C&S). The second column reports the Bayesian information criterion. The third and fourth columns report the out-of-sample $R$-square for predicting dividend growth rates and the corresponding $p$-value from the $F$-test for model significance. In-sample (out-of-sample) statistics are based on data between 1946 and 2013 (1976 and 2013).

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-$\tau$, for $\tau$ 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 $\mu_q$ 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 $\phi$ and $\varphi$ 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 $\theta$ 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.

It is clear from Figure 1 that the persistence $\rho$ 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, $\rho$ is estimated with the highest standard error (i.e. 0.160). Investors' beliefs about $\rho$ flunctuate significantly over the sample period, especially around three periods during which beliefs about $\rho$ 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

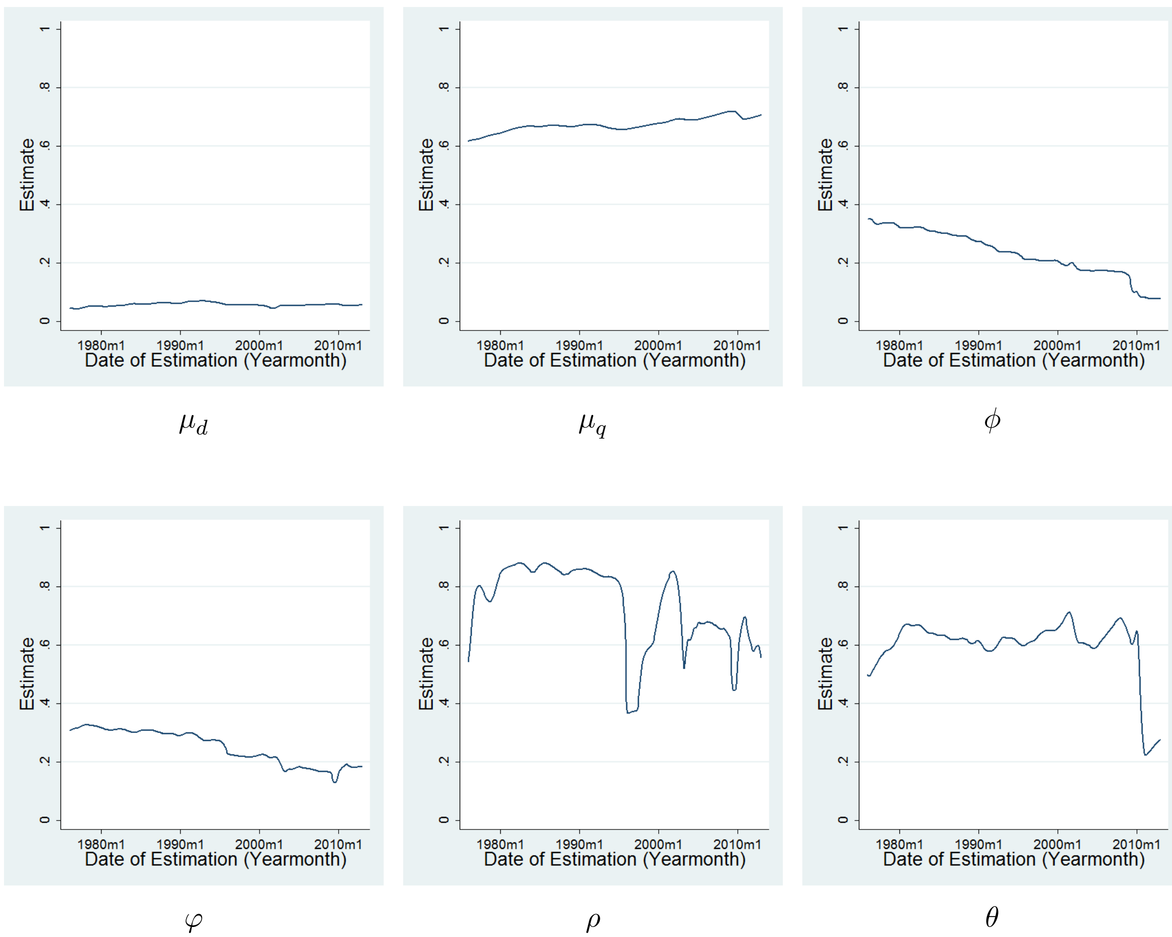


Figure 1: **Evolution of Model Parameters Estimated Out-of-Sample.** This figure plots estimates of the six model parameters in (8) that affect expected dividend growth rates, assuming that our dividend model is estimated based on data up to time-$\tau$ for $\tau$ between 1976 and 2013.

sees a gradual decrease in investors' beliefs about $\rho$ since early 1980s. For example, if we were to pick a random date between 1976 and 2013 and estimate our model based on data prior to that date, on average we would have estimated a $\rho$ of 0.734.[8] 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 $\rho$ of the latent variable $x_t$ is between 0.214 and 0.842. The same confidence interval for the persistence $\theta$ of earnings-to-dividend ratios is between 0.054 and 0.508. To quanitfy the speed of learning, following Johannes, Lochstoer and Mou (2014), for each of the six parameters that affect expected dividend growth rates, we construct a measure that is the inverse ratio between the simulated standard error assuming that the parameter is estimated based on data up to 2013 and the 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 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 reported in Table 2, we find that it is significantly more difficult to learn about $\rho$ than about any of the other five model parameters.

| $\mu_d$ | $\mu_q$ | $\phi$ | $\varphi$ | $\rho$ | $\theta$ |
|---|---|---|---|---|---|
| 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 simulated standard error assuming that the parameter is estimated based on data up to 2013 and the 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).

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

We show that learning about dividend dynamics can have important asset pricing implications. Consider the log linearized present value relationship in Campbell and Shiller (1998a):

$$p_t - d_t = \frac{\kappa_0}{1-\kappa_1} + \sum_{s=0}^{\infty} \kappa_1^s \left( E_t[\Delta d_{t+s+1}] - E_t[r_{t+s+1}] \right), \tag{10}$$

where $\kappa_0$ and $\kappa_1$ are log-linearizing constants and $r_{t+1}$ is the stock index's log return.[9] 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 (10), we can compute stock yields as:

$$\begin{aligned} sy_t &\equiv (1-\kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta r_{t+s+1}] \\ &= \kappa_0 - (1-\kappa_1)(p_t - d_t) + (1-\kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta d_{t+s+1}]. \end{aligned} \tag{11}$$

Define long run dividend growth expectation as:

$$\partial_t \equiv (1-\kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta d_{t+s+1}]. \tag{12}$$

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, using our dividend model, we can re-write (12) as:

$$\begin{aligned} \partial_t &= (1-\kappa_1) \sum_{s=0}^{\infty} \kappa_1^s \left( \mu_d + \frac{1}{1-\phi} \left( \rho^s x_t + \theta^s (\varphi - (1-\theta)\phi)(q_t - \mu_q) \right) \right) \\ &= \mu_d + \frac{1-\kappa_1}{1-\phi} \left( \frac{1}{1-\kappa_1 \rho} x_t + \frac{\varphi - (1-\theta)\phi}{1-\kappa_1 \theta} (q_t - \mu_q) \right). \end{aligned} \tag{13}$$

If a different dividend model is used instead, long run dividend growth expectations will also be different. Further, because long run dividend growth expectations are functions of dividend model parameters, it is also affected by whether model parameters are estimated

[9] Throughout this paper, 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 $\kappa_1 = 0.970$.

once based on the full data sample, or estimated at each point in time based on data avaliable at the time. In Figure 3, 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, asuming that investors learn about dividend dynamics using our model.

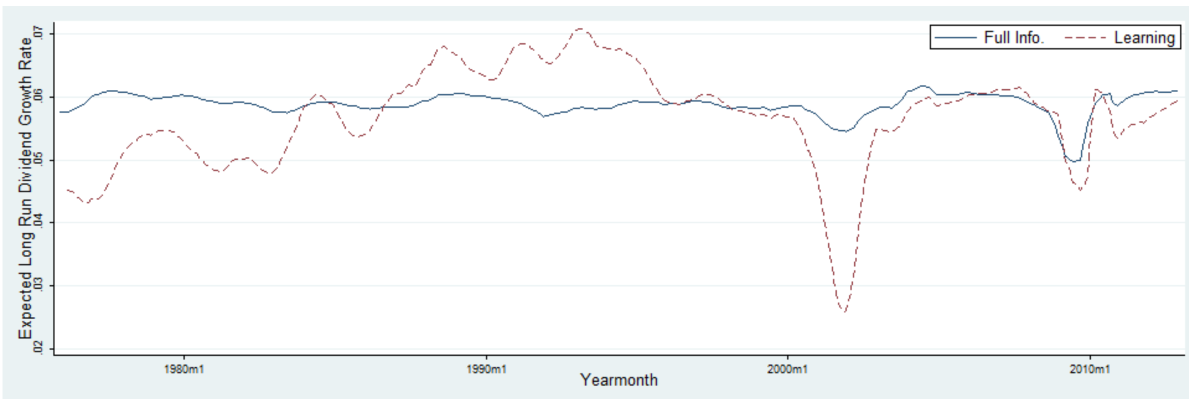


Figure 2: **Expected Long Run Dividend Growth Rates**. This figure plots long run dividend growth rate expectations $\partial_t \equiv (1-\kappa_1)\sum_{s=0}^{\infty}\kappa_1^s E_t[\Delta d_{t+s+1}]$, computed using our dividend model, for the period between 1976 and 2013. Dividend model parameters are estimated based on data since 1946. Under full information, model parameters are estimated based on the full data sample. Under learning, parameters are estimated at each point in time based on data avaliable at the time.

In Figure 3, we plot stock yields, computed using our model and assuming that model parameters are either estimated once based on the full data sample or estimated at each point in time based on data avaliable at the time:

$$sy_t = \kappa_0 - (1-\kappa_1)(p_t - d_t) + \mu_d + \frac{1-\kappa_1}{1-\phi}\left(\frac{1}{1-\kappa_1\rho}x_t + \frac{\varphi-(1-\theta)\phi}{1-\kappa_1\theta}(q_t - \mu_q)\right). \quad (14)$$

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 are only minor differences 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 rate 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 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 based on the full data sample. Under learning, parameters are estimated at each point in time based on data avaliable 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. First, we show that stock yields, computed assuming that investors learn about dividend dynamics using our model (see (12)), predict future stock index returns. To establish a baseline, note that, if we assume dividend growth rates follow a white noise process centered around $\mu_d$, stock yield can be simplified to:

$$sy_t = \kappa_0 + (1 - \kappa_1)\mu_d - (1 - \kappa_1)(p_t - d_t). \tag{15}$$

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 of Table 6. Stambaugh (1999) shows that, when variables are highly serially correlated, OLS estimators' finite-sample properties can significantly deviate from the standard regression setting. To address this issue, we report simulated $p$-values of coefficient estimates that take into account the effect of serial correlations in finite sample. See the Appendix for details on simulation. Results from Table 6 show that price-to-dividend ratios explain 10.2 percent of the variation in stock index returns over the next year. However, the coefficient estimate is only weakly significant.

We then regress stock index returns over the next year on stock yields in (14), computed assuming that investors estimate model parameters at each point in time based on data avaliable 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 dividend growth rates are assumed to be white noises. 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 next year on both price-to-dividend ratios and stock yields, and report regression statistics in the third column of Table 6. Results show that stock yields strictly 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 (14), computed assuming that investors do not learn. That is, instead of computing long run dividend growth expectations by estimating model parameters at each point in time based on data avaliable at the time, we estimate model parameters once based on the full data sample. 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 are almost no differences between the time series of stock yields, computed using our model but assuming that investors do not learn, and price-to-dividend ratios.

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, are unable to outperform price-to-dividend ratios

| | J&L | | | | vB&B | C&S |
|---|---|---|---|---|---|---|
| $p_t - d_t$ | -0.116* [0.051] | | 0.016 [0.668] | | | |
| $sy_t$ (Learning) | | 3.964** [0.013] | 4.355** [0.036] | | 3.000* [0.055] | 2.741** [0.028] |
| $sy_t$ (Full Info.) | | | | 3.753** [0.034] | | |
| $R^2$ (Returns) | 0.102 | 0.152 | 0.152 | 0.105 | 0.088 | 0.106 |
| $R^2$ (Excess Returns) | 0.090 | 0.140 | 0.141 | 0.093 | 0.075 | 0.094 |

Table 6: **Predicting Stock Index Returns using Stock Yields**: This table reports the coefficient estimates and $R$-square values 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 (2012) (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. Regressions are based on data between 1976 and 2013. Dividend model parameters are estimated based on data since 1946. Simulated $p$-values are reported in square parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.

in explaining the variation in future stock index returns.

Stock index returns combine the risk free rate and risk premium. We also report, in Table 6, the $R$-square values for predicting stock index excess returns.[10] Results show that the performance gap 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 not the risk free rate,

Recall that there are six model parameters that affect expected dividend growth rates. These parameters are the unconditional means $\mu_d$ and $\mu_q$ of dividend growth rates and earnings-to-dividend ratios, the persistence $\rho$ and $\theta$ of the latent variable $x_t$ and earnings-to-dividend ratios, and coefficients $\phi$ and $\varphi$ 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 model parameters into one subset that includes persistence

[10]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 $\frac{v\hat{a}r((r_{t+1}-r_{f,t+1})-(\hat{r}_{t+1}-r_{f,t+1}))}{v\hat{a}r(r_{t+1}-r_{f,t+1})}$.

$\rho$ 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 remaining parameters in our model. That is, parameters not subject to learning is fixed at their full sample estimated values whereas other parameters are estimated at each point in time based on data avaliable at the time. We call this partial learning. We regress stock index returns over the next year on stock yields, computed assuming parital 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. This shows that investors' learning is multi-dimensional, and not restricted to a specific parameter or few parameters. Nevertheless, we find, based on the $R$-square values, that shutting down learning about $\rho$ adversely affect return predictability more than shutting down learning about the other five parameters combined. This suggests that learning about $\rho$ has the strongest implications for asset pricing. This is also consistent with results from the previous sections that show $\rho$ is the model parameter that is hardest to learn and investors beliefs about $\rho$ fluctuates the most over time.

| | Shutting Down Learning about $\Upsilon$ | |
|---|---|---|
| | $\Upsilon = \{\mu_d, \mu_q, \phi, \varphi, \theta\}$ | $\Upsilon = \{\rho\}$ |
| $sy_t$ (Learning) | $3.806^{**}$ [0.014] | $3.854^{**}$ [0.028] |
| $R^2$ | 0.125 | 0.115 |

Table 7: **Predicting Stock Index Returns using Stock Yields (Partial Learning)**: This table reports the coefficient estimates and $R$-square values 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) and assuming investor learn about some model parameters but not others. Regressions are based on data between 1976 and 2013. Dividend model parameters are estimated based on data since 1946. Simulated $p$-values are reported in square 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 avaliable at the time, investors' beliefs about the persistence of dividend growth rates vary over time as more data become avaliable. 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 $\rho$ of the latent variabe $x_t$ and the persistence $\theta$ of earnings-to-dividend ratios. To derive a unified measure of persistence, note that, following (8), one standard deviation shocks to both $x_t$ and earnings-to-dividend ratios increase long run dividend growth expectations, defined in (11), by:[11]

$$
\begin{aligned}
&\partial_t|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\partial_t|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)\\
&=(1-\kappa_1)\sum_{s=0}^{\infty}\frac{\kappa_1^s}{1-\phi}\left(\rho^s\sigma_x+\theta^s(\varphi-(1-\theta)\phi)\sigma_q\right)\\
&=\frac{(1-\kappa_1)\sigma_x}{1-\phi}\frac{1}{1-\kappa_1\rho}+\frac{(1-\kappa_1)(\varphi-(1-\theta)\phi)\sigma_q}{1-\phi}\frac{1}{1-\kappa_1\theta}.
\end{aligned}
\tag{16}
$$

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

$$
\begin{aligned}
&\Delta d_{t+1}|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\Delta d_{t+1}|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)\\
=&\frac{1}{1-\phi}\left(\sigma_x+(\varphi-(1-\theta)\phi)\sigma_q\right).
\end{aligned}
\tag{17}
$$

The ratio between the short run and the long run effects on dividend growth rates of one

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

standard deviation shocks to both $x_t$ and earnings-to-dividend ratios is:

$$\frac{\Delta d_{t+1}|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\Delta d_{t+1}|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)}{\partial_{t+1}|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\partial_{t+1}|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)}$$
$$=\frac{\sigma_x+(\varphi-(1-\theta)\phi)\sigma_q}{\frac{(1-\kappa_1)\sigma_x}{1-\kappa_1\rho}+\frac{(1-\kappa_1)(\varphi-(1-\theta)\phi)\sigma_q}{1-\kappa_1\theta}}. \tag{18}$$

This ratio is decreasing in the persistence of dividend growth rates. That is, if dividend growth rates are more persistent, shocks to dividends have stronger effects on long run dividend growth expectations, and so this ratio is lower. Thus, we define the persistence of dividend growth rates (i.e. $\omega$) as minus this ratio, scaled so that $\omega$ is between $-1$ and 1:[12]

$$\omega=\frac{1}{\kappa_1}-\frac{(1-\kappa_1)}{\kappa_1}\cdot\frac{\Delta d_{t+1}|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\Delta d_{t+1}|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)}{\partial_{t+1}|(\epsilon_{x,t+1}=1,\epsilon_{q,t+1}=1)-\partial_{t+1}|(\epsilon_{x,t+1}=0,\epsilon_{q,t+1}=0)}$$
$$=\frac{1-\frac{\sigma_x+(\varphi-(1-\theta)\phi)\sigma_q}{\frac{\sigma_x}{1-\kappa_1\rho}+\frac{(\varphi-(1-\theta)\phi)\sigma_q}{1-\kappa_1\theta}}}{\kappa_1}. \tag{19}$$

To model investors' learning about the persistence $\omega$ of dividend growth rates, we estimate our dividend model at each point in time using data avaliable at the time. Denote $\omega(t)$ as $\omega$ estimated using data up to time-$t$. We use $\omega(t)$ as a estimate for investors' time-$t$ belief about $\omega$.

## 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 peristent.

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

[12] We only use investors' beliefs about $\omega$ in regressions, so the scaling does not affect the statistical significance of any of our estimates.

has Epstein and Zin (1989) preferences, defined recursively as:

$$U_t = \left[(1-\delta)\tilde{C}_t^{\frac{1-\alpha}{\zeta}} + \delta\left(E_t\left[U_{t+1}^{1-\alpha}\right]\right)^{\frac{1}{\zeta}}\right]^{\frac{\zeta}{1-\alpha}}, \ \zeta = \frac{1-\alpha}{1-\frac{1}{\psi}}, \tag{20}$$

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

$$m_{t+1} = -\zeta \log(\delta) - \frac{\zeta}{\psi}\Delta\tilde{c}_{t+1} + (\zeta - 1)\,\tilde{s}_{t+1}, \tag{21}$$

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

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

$$\begin{aligned} \Delta d_{t+1} - \mu_d &= y_t + \sigma_d \epsilon_{d,t+1} \\ y_{t+1} &= \omega y_t + \sigma_y \epsilon_{y,t+1} \\ \begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{y,t+1} \end{pmatrix} &\sim \text{ i.i.d. } \mathbb{N}\left(\mathbf{0}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right). \end{aligned} \tag{22}$$

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:

$$\Delta\tilde{c}_{t+1} = \Delta d_{t+1} = y_t + \sigma_d \epsilon_{d,t+1}. \tag{23}$$

We assume investors price the stock index using the Kreps (1988) anticipated utility. Assuming anticipated uility implies that investors maximize utility at each point in time assuming the current model parameter estimates are the true parameters, but then revises 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 price 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.[13] Currently, this is often used in the macroeconomics and asset pricing literature for dealing with parameter uncertainty in a dynamic setup.

Given the representative agent's preferences in (20), consumption dynamics in (23), and dividend dynamics in (22), 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, unconditional mean $\mu_r$ of stock index returns is:

$$\mu_r = E\left[r_{t+1}\right] = -\log(\delta) + \frac{\mu_d}{\psi} - \frac{\zeta\left(\psi^2-1\right)^2\sigma_d^2}{2\psi^2} - \frac{\zeta\kappa_1^2A^2\sigma_x^2}{2}.$$
$$A = \frac{1-\frac{1}{\psi}}{1-\kappa_1\omega}. \qquad (24)$$

where $\kappa_0$ and $\kappa_1$ are log-linearizing constants.[14] We note that $-\zeta A^2$ is increasing in beliefs about $\omega$ if $\zeta < 0$, and, although changes in beliefs about $\omega$ also affect $\mu_r$ through $\kappa_0$ and $\kappa_1$, effects on $\mu_r$ through the log-linearizing constants are relatively small and not of first order importance. Therefore, if the representative agent has preferences for early resolution of uncertainty, $\zeta < 0$, and $\mu_r$ is increasing in beliefs about $\omega$. On the other hand, if the representative agent prefers late resolution of uncertainty, $\mu_r$ is decreasing in beliefs about $\omega$. 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 $\omega$ of dividend growth rates and the unconditional mean $\mu_r$ of expected returns. First, we note that as $\mu_r$ increases, price-to-dividend ratios on average decrease and stock

[13] Collin-Dufresne, Johannes, and Lochstoer (2015) show that parameter uncertainties can introduce new factors with significant associated risk premiums if investors are fully Bayesians.

[14] $\kappa_0 = \log(1+\exp(\overline{p-d})) - \kappa_1(\overline{p-d})$ and $\kappa_1 = \frac{\exp(\overline{p-d})}{1+\exp(\overline{p-d})}$.

yields on average increase. To see why, we note from (10) and (11) that:

$$E\left[p_t - d_t\right] = \frac{\kappa_0 + \mu_d - \mu_r}{1 - \kappa_1},$$
$$E\left[sy_t\right] = \mu_r. \tag{25}$$

where $E\left[\cdot\right]$ denotes unconditional expectation. 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. 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 on investors' beliefs about $\omega$ 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 $\omega$ are associated with lower price-to-dividend ratios and higher stock yields, and vice versa. Between 1976 and 2013, investors' beliefs about $\omega$ 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 $\omega$ and either price-to-dividend ratios or stock yields is driven by the covariance between investors' beliefs about $\omega$ and long run dividend growth expectations. To rule out this possibility, we regress long run dividend growth expectations on investors' beliefs about $\omega$, and report regression statistics in the third column of Table 8. We find that investors' beliefs about $\omega$ do not explain long run dividend growth expectations. This confirms that investors' beliefs about $\omega$ affect price-to-dividend ratios and stock yields through the discount rate channel.

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 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 $\omega$ 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):

$$E_{t+1}[r_{t+2}] - \mu_r = \gamma\left(E_t[r_{t+1}] - \mu_r\right) + \sigma_r \epsilon_{r,t+1}. \tag{26}$$

| | $p_t - d_t$ | $sy_t$ (Learning) | $\partial_t$ |
|---|---|---|---|
| $\omega(t)$ | -2.112*** [0.010] | 0.060** [0.047] | -0.003 [0.322] |
| $R^2$ | 0.253 | 0.157 | 0.001 |

Table 8: **Stock Index Prices and Investors' Beliefs about Persistence of Dividend Growth Rates**. This table reports the coefficient estimates and $R$-square values from regressing log price-to-dividend ratios, stock yields, and long run dividend growth expectations on investors' belief about persistence $\omega$ of dividend growth rates. Regressions are based on overlapping annual data between 1976 and 2013. Estimating dividend dynamics is based on data since 1946. Simulated $p$-values are reported in square parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.

Substituting (26) into the log-linearizing present value relationship of Campbell and Shiller (1988a), we can write expected returns over the short-horizon as:

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

$$E_t[r_{t+1}] = \frac{1-\kappa_1\gamma}{1-\kappa_1}sy_t - \frac{\kappa_1(1-\gamma)\mu_r}{1-\kappa_1}. \tag{27}$$

Although $\kappa_1$ and $\kappa_2$ are functions of $\mu_r$, the effects of changes in $\mu_r$ on these log-linearizing constants are relatively small and not of first order importance. Thus, we note from (27) that, after controlling for price-to-dividend ratios or stock yields, expected stock index returns over the short-horizon are decreasing in $\mu_r$. Intuitively, this is because, when $\mu_r$ increases, in order to justify the same stock yield, expected returns over the short-horizon must decrease sufficiently to compensate for the effect of an increase in $\mu_r$ 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 $\omega$ and either price-to-dividend ratios or stock yields. 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 $\omega$ predict lower stock index returns over the short-horizon, and vice versa. We find that, between 1976 and 2013, stock yields and investors' beliefs about $\omega$ together explain as much as 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 including investors beliefs about $\omega$ as an additional regressor strengthens the predictive performance of price-to-dividend ratios (stock yields) from statistical significance at the 90 (95) percent confidence level to statistical significance at the 99 (99) percent confidence level. We also report the $R$-square values for predicting excess returns and find those to be comparable.

| | | |
|---|---|---|
| $\omega(t)$ | -0.644** [0.014] | -0.562** [0.027] |
| $p_t - d_t$ | -0.193*** [0.002] | |
| $sy_t$ (Learning) | | 5.448*** [0.002] |
| $R^2$ (Returns) | 0.235 | 0.265 |
| $R^2$ (Excess Returns) | 0.224 | 0.255 |

Table 9: **Stock Index Returns and Investors' Beliefs about Persistence of Dividend Growth Rates.** This table reports the coefficient estimates and $R$-square values from regressing stock index returns on investors' belief about persistence $\omega$ of dividend growth rate, log price-to-dividend ratios, and stock yields. Regressions are based on overlapping annual data between 1976 and 2013. Estimating dividend dynamics is based on data since 1946. Simulated $p$-values are reported in square 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 expected 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 and consumption dynamics from the long-run risk model of Bansal and Yaron (2004). We assume there is a representative agent who has Epstein and Zin (1989) preferences, which is defined in (20). Following Bansal and Yaron (2004), we set $\delta$ to $0.998^{12}$, $\alpha$ to 10, and $\psi$ to 1.5.[15]

We assume dividend dynamics are described by our dividend model, given in (6). As in Bansal and Yaron (2004), we assume expected growth rates in real consumption follow an AR[1] process and allow volatility in consumption growth rates to be time varying. That is, we describe real consumption growth rates using the following system of equations:

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

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

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

Because we do not use 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 $\mu_\pi = 0.036$.[16] We assume that the latent variable $x_t$ in real consumption growth rates is the same as the latent variable in dividend growth rates. To set $\gamma$, we note that, in Bansal and Yaron (2004), the unconditional standard deviation of expected real consumption growth rates is $12 \cdot 0.044 \cdot 0.0078$. To match this, $\gamma$ must be set to $\sigma_x\left(12 \cdot 0.044 \cdot 0.0078\right)^{-1}$. In Bansal and Yaron (2004), the persistence $\rho$ of $x_t$ is set to $0.975^{12}$. One common criticism

[15]The original Bansal and Yaron (2004) calibration is based on monthly observations, so whenever necessary we convert them to their annualized equivalents.

[16]The choice of 0.036 is to match the average inflation rate between 1946 and 2013.

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 diffcult to find support in the data.[17] By estimating $\rho$ and other model parameters in real time from dividend data, our approach is not subject to this criticism. In fact, to the contrary, this criticism serves as the rationale for why we expect learning to be important. Finally, we follow Bansal and Yaron (2004) and set $\mu_\varsigma$ to $12 \cdot 0.0078^2$, $\varrho$ to $0.987^{12}$, and $\sigma_\varsigma$ to $12^2 \cdot 0.23 \cdot 10^{-5}$. We solve 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 $\sigma_t^2$, 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:

$$p_t - d_t = A_{d,0} + A_{d,1} x_t + A_{d,2} \sigma_t^2 + A_{d,3} \left(q_t - \mu_q\right). \tag{29}$$

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

$$E_t[r_{t+1}] = A_{r,0} + A_{r,1} x_t + A_{r,2} \sigma_t^2, \tag{30}$$

where coefficients $A_{d,\cdot}$ and $A_{r,\cdot}$, derived in the Appendix, are functions of parameters governing investors' preferences, consumption dynamics, and dividend dynamics.

## 5.2 Estimation

We describe how investors, whose preferences and consumption dynamics are governed by (20) and (28), learn about dividend dynamics. Our approach is similar to van Binsbergen and Koijen (2010) in that we estimate dividend and discount rate processes jointly. We assume that, at each time-$\tau$, investors observe the entire history of dividend growth rates and price-to-dividend ratios.

$$\begin{aligned} \Delta d_{t+1} - \mu_d(\tau) &= x_t + \phi(\tau) \cdot (\Delta e_{t+1} - \mu_d) + \varphi(\tau) \cdot \big(q_t - \mu_q(\tau)\big) + \sigma_d(\tau) \cdot \epsilon_{d,t+1} \\ p_t - d_t &= A_{d,0}(t) + A_{d,1}(t) \cdot x_t + A_{d,2}(t) \cdot \sigma_t^2 + A_{d,3}(t) \cdot \big(q_t - \mu_q(t)\big), \end{aligned} \tag{31}$$

where $\mu_d(\tau)$ is investors' time-$\tau$ belief about the unconditional mean $\mu_d$ of dividend growth rates. That is, $\mu_d(\tau)$ denotes $\mu_d$ estimated based on data up to time-$\tau$. The same notation applies to other dividend model parameters. Similarly, coefficients $A_{d,\cdot}(\tau)$ are functions of dividend model parameters estimated based on data up to time-$\tau$. We call (31) the measurement equations. The measurement equations are functions of the three state variables: two latent (i.e. $x_t$ and $\sigma_t^2$) and one observed (i.e. $q_t$). Transition equations for

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

the two latent variables $x_t$ and $\sigma_t^2$ are:

$$x_{t+1} = \rho(\tau) \cdot x_t + \sigma_x(\tau) \cdot \epsilon_{x,t+1}$$
$$\sigma_{t+1}^2 - \mu_\varsigma(\tau) = \varrho(\tau) \cdot \big(\sigma_t^2 - \mu_\varsigma(\tau)\big) + \sigma_\varsigma(\tau) \cdot \epsilon_{\varsigma,t+1}. \tag{32}$$

Transition equation for the observed state variable is:

$$q_{t+1} - \mu_q(\tau) = \theta(\tau) \cdot \big(q_t - \mu_q(\tau)\big) + \sigma_q(\tau) \cdot \epsilon_{q,t+1}. \tag{33}$$

Because the second equation in (31) has no error term, we can substitute it into the second equation in (32). This means that we can reduce the system of equations in (31), (32), and (33) into:

$$\Delta d_{t+1} - \mu_d(\tau) = x_t + \phi(\tau) \cdot (\Delta e_{t+1} - \mu_d(\tau)) + \varphi(\tau) \cdot \big(q_t - \mu_q(\tau)\big) + \sigma_d(\tau) \cdot \epsilon_{d,t+1}$$
$$\frac{(p_{t+1} - d_{t+1}) - A_{d,0}(t+1) + A_{d,1}(t+1) \cdot x_{t+1} + A_{d,3}(t+1) \cdot \big(q_{t+1} - \mu_q(t+1)\big)}{A_{d,2}(t+1)} - \mu_\varsigma$$
$$= \varrho \left( \frac{(p_t - d_t) - A_{d,0}(t) + A_{d,1}(t) \cdot x_t + A_{d,3}(t) \cdot \big(q_t - \mu_q(t)\big)}{A_{d,2}(t)} - \mu_\varsigma \right) + \sigma_\varsigma \cdot \epsilon_{\varsigma,t+1}$$
$$x_{t+1} = \rho(\tau) \cdot x_t + \sigma_x(\tau) \cdot \epsilon_{x,t+1}$$
$$q_{t+1} - \mu_q(\tau) = \theta(\tau) \cdot \big(q_t - \mu_q(\tau)\big) + \sigma_q(\tau) \cdot \epsilon_{q,t+1}. \tag{34}$$

We compute the likelihood of the asset pricing model using Kalman filters (Hamilton (1994)) and estimate dividend model paremeters

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

based on maximum-likelihood. See the Appendix for details. In estimating our model, we closely follow the steps in van Binsbergen and Koijen (2010). We note that, although stock index returns do not enter into this system of equations, expected returns can be computed by substituting model parameter estimates into (30).

## 5.3 Empirical Results

### 5.3.1 Time Variation in Expected Returns

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

$$R^2(r_{t+1}) = 1 - \frac{\sum_{t=T_0}^{T-1} \left(r_{t+1} - E_t[r_{t+1}]\right)^2}{\sum_{t=T_0}^{T-1} \left(r_{t+1} - \underline{r}_t\right)^2}. \tag{35}$$

where $\underline{r}_t = \frac{1}{t}\sum_{s=0}^{t-1} r_{s+1}$ is the historical average of stock index returns, $T$ denotes the end of the data sample, and $T_0$ denotes 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 of our dividend model are estimated at each point in time based on data avaliable at the time, parameters governing preferences and consumption dynamics are fixed and can be forward looking. In the first and second columns of Table 10, we report the quasi out-of-sample $R$-square value for predicting annual stock index returns using expected returns in (30), computed assuming investors learn about dividend dynamics based on the system of equations in (34), and the corresponding $p$-value from the $F$-test for model significance. We find that, between 1976 and 2013, our long-run risk model predicts 22.8 percent of the variation in annual stock index returns. We also report the quasi out-of-sample $R$-square value for predicting excess returns and find it to be comparable.

| | Learning | | Full Info. | |
|---|---|---|---|---|
| | $R^2$ | $p$-value | $R^2$ | $p$-value |
| Returns | 0.228 | 0.003 | 0.131 | 0.038 |
| Excess Returns | 0.221 | 0.004 | 0.126 | 0.044 |

Table 10: **Stock Index Returns and Model Implied Expected Returns.** This table reports the out-of-sample $R$-square value from predicting stock index returns or excess returns using expected returns implied by the asset pricing model, assuming investors either learn, or do not learn, about dividend dynamics. Also reported is the corresponding $p$-value from the $F$-test for model significance. Statistics are based on data between 1976 and 2013. Estimating the asset pricing model is based on data since 1946.

As a point of comparison, we also report the $R$-square values for predicting stock index returns and excess returns using expected returns in (30), computed assuming that investors do not learn. That is, instead of estimating dividend model parameters at each point in time based on data avaliable at the time, we estimate model parameters once based on the full data sample. We find that the $R$-square value for predicting annual stock index returns drops to 13.1 percent without learning. Thus, learning about dividend dynamics accounts for over forty-percent of the 22.8 $R$-square value.

To examine the robustness of this forecasting performance in sub-samples of the data, following Goyal and Welch (2008), we define the cumulative sum of squared errors difference (SSED) as:

$$SSED_t = \sum_{s=T_0}^{t-1} \left( (r_{s+1} - E_t[r_{s+1}])^2 - (r_{s+1} - \underline{r}_s)^2 \right). \tag{36}$$

We plot SSED in Figure 4. We note that if the forecasting performance of the model implied expected returns 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-sample of the data, we should see a significant drawback in SSED in that sub-sample. Figure 4 shows that our long-run risk model's forecasting performance is relatively stable and robust between 1976 and 2013. However, consistent with Nardari (2011) and Golez and Koudijs (2015), we find that most of the forecasting performance is realized during the NBER recessions.

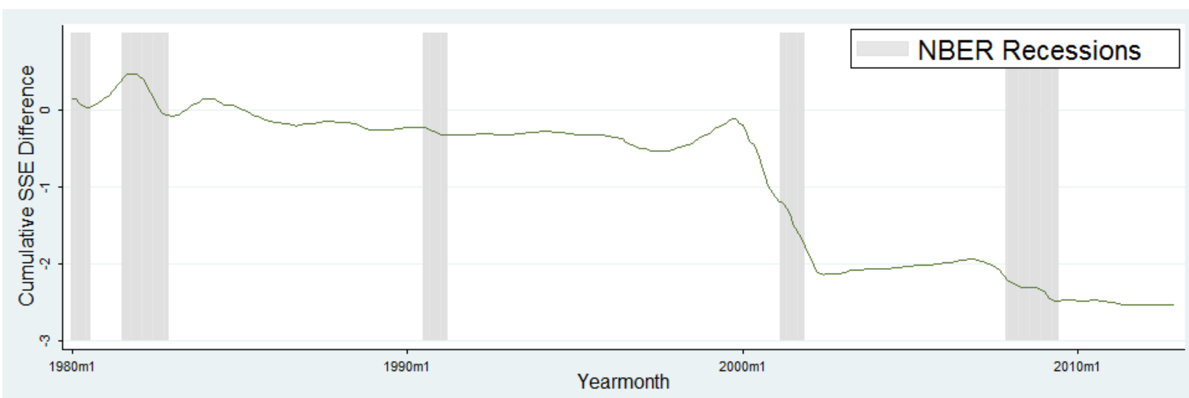


Figure 4: **Stock Index Returns and Model Implied Expected Returns (Cumulative SSE Difference).** This figure plots the cumulative sum of squared errors difference for the period between 1976 and 2013. Estimating the asset pricing model is based on data since 1946.

### 5.3.2 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 (2012) 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, we can label (29) as:

$$p_t - d_t = \underbrace{A_{d,0}(t)}_{1)} + \underbrace{A_{d,1}(t) \cdot x_t}_{2)} + \underbrace{A_{d,2}(t) \cdot \sigma_t^2}_{3)} + \underbrace{A_{d,3}(t) \cdot \left(q_t - \mu_q(t)\right)}_{4)}. \tag{37}$$

So 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 $\sigma_t^2$, or 4) the variation in earnings-to-dividend ratios. This means that we can decompose the sample variance of price-to-dividend ratios into:

$$\begin{aligned} v\hat{a}r\left(p_t - d_t\right) = & \underbrace{c\hat{o}v\left(p_t - d_t, A_{d,0}(t)\right)}_{1)} + \underbrace{c\hat{o}v\left(p_t - d_t, A_{d,1}(t) \cdot x_t\right)}_{2)} \\ & + \underbrace{c\hat{o}v\left(p_t - d_t, A_{d,2}(t) \cdot \sigma_t^2\right)}_{3)} + \underbrace{c\hat{o}v\left(p_t - d_t, A_{d,3}(t) \cdot \left(q_t - \mu_q(t)\right)\right)}_{4)}. \end{aligned} \tag{38}$$

where $v\hat{a}r(\cdot)$ denotes sample variance and $c\hat{o}v(\cdot)$ denotes sample covariance.

We report the decomposition results in Table 11. Results show that 72.1 percent the variation in price-to-dividend ratios is driven by the variation in $\sigma_t^2$, which affects discount rates but not cash flow expectations. This is consistent with what has been documented in the existing literature. Nevertheless, we find that a substantial 23.5 percent of the variation in price-to-dividend ratios can be attributed to changes in beliefs about parameters governing the dividend process, and the remaining 4.4 percent is due to changes in state variables that directly affect expected dividend growth rates. In other words, we find 27.9 percent of the variation in price-to-dividend ratios is due to learning about dividend dynamics.

| Learning | $x_t$ | $\sigma_t^2$ | $q_t$ |
|---|---|---|---|
| 0.235 | 0.032 | 0.721 | 0.012 |

Table 11: **Decomposing the Variation in Price-to-Dividend Ratios**. This figure 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 $x_t$, the variation in $\sigma_t^2$, and the variation in $q_t$. Statistics are based on data between 1976 and 2013. Estimating the dividend and discount rate processes is based on data since 1946.

# 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 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 persistence of dividend growth rates is consistent with investors' preference 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 our long-run risk model predicts 22.8 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.

# References

Alvarez, F., and U. J. Jermann (2005): "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73, 1977–2016.

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., D. Kiku, and A. Yaron (2010): "Long Run Risks, the Macroeconomy, and Asset Prices," *American Economic Review*, 100, 542–546.

——— (2012): "An Empirical Evaluation of the Long-Run Risks Model for Asset Prices," *Critical Finance Review*, 1, 183–221.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Beeler, J., and J. Y. Campbell (2012): "The Long Run Risks Model and Aggregate Asset Prices: An Empirical Assessment," *Critical Finance Review*, 1, 141–182.

Breeden, D. T. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–296.

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–255.

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 S. B. Thompson (2008): "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?," *Review of Financial Studies*, 22, 1509–1531.

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.

Cochrane, J. H. (2008): "The Dog That Did Not Bark: A Defense of Return Predictability," *Review of Financial Studies*, 21, 1533–1575.

——— (2011): "Presidential Address: Discount Rates," *Journal of Finance*, 66, 1047–1108.

Collin-Dufresne, P., M. Johannes, and L. A. Lochstoer (2013): "Parameter Learning in General Equilibrium: The Asset Pricing Implications," *Working Paper*.

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*.

Dew-Becker, I. (2014): "Long Run Risk is the Worst-Case Scenario," *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.

Gabaix, X. (2012): "Variable Rare Disasters: An Eactly Solved Framework for Ten Puzzles in Macro-Finance," *Quarterly Journal of Economics*, 127, 645–700.

Goffe, W. L., G. D. Ferrier, and J. Rogers (1994): "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60, 65–99.

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., J. C. Heaton, and N. Li (2008): "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116, 260–302.

Jagannathan, R., and M. Srikant (2011): "Price Dividend Ratio Factors: Proxies for Long Run Risk," *Working Paper*.

Kelly, B., and S. Pruitt (2013): "Market Expectations in the Cross-Section of Present Values," *Journal of Finance*, 68, 1721–1756.

Koijen, R. S., H. Lustig, S. van Nieuwerburgh, and A. Verdelhan (2010): "Long Run Risk, the Wealth-Consumption Ratio, and the Temporal Pricing of Risk," *American Economic Review*, 100, 552–556.

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.

——— (2008): "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, 21, 1607–1652.

Lettau, M., and S. van Nieuwerburgh (2008): "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, 21, 1607–1652.

Lewellen, J. (2004): "Predicting Returns with Financial Ratios," *Journal of Financial Economics*, 74, 209–235.

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.

Polk, C., S. Thompson, and T. Vuolteenaho (2006): "Cross-Sectional Forecasts of the Equity Premium," *Journal of Finance*, 55, 2219–2257.

van Binsbergen, J. H., and R. S. Koijen (2010): "Predictive Regression: A Present-Value Approach," *Journal of Finance*, 65, 1439–1471.

——— (2012): "On the Timing and Pricing of Dividends," *American Economic Review*, 102, 1596–1618.

Wachter, J. A. (2005): "Solving Models with External Habit," *Finance Research Letters*, 2, 210–226.

# 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 (22), consumption dynamics in (23), and investors' preferences in (20). Our derivation closely follows the steps in Bansal and Yaron (2004). The log stochastic discount factor is:

$$m_{t+1} = \zeta \log(\delta) - \frac{\zeta}{\psi} \Delta \tilde{c}_{t+1} + (\zeta - 1)\tilde{s}_{t+1}. \tag{39}$$

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:

$$\tilde{s}_{t+1} = \kappa_0 + \kappa_1 z_{c,t+1} - z_{c,t} + \Delta \tilde{c}_{t+1}. \tag{40}$$

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:

$$z_{c,t} = A_{c,0} + A_{c,1} x_t. \tag{41}$$

We can write:

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

$$var_t\left(m_{t+1} + \tilde{s}_{t+1}\right) = \left(\zeta - \frac{\zeta}{\psi}\right)^2 \sigma_d^2 + \zeta^2\kappa_1^2 A_1^2 \sigma_x^2. \tag{42}$$

Since $m_{t+1}$ is the stochastic discount factor, $E_t[\exp(m_{t+1} + \tilde{s}_{t+1})] = 1$. This means that:

$$\log\left(E_t[\exp(m_{t+1} + \tilde{s}_{t+1})]\right) = E_t[m_{t+1} + \tilde{s}_{t+1}] + \frac{1}{2} var_t(m_{t+1} + \tilde{s}_{t+1}) = 0. \tag{43}$$

So, we can solve for coefficients $A_0$ and $A_1$ as:

$$A_0 = \frac{1}{1-\kappa_1}\left(\log(\delta) + \left(1-\frac{1}{\psi}\right)\mu_d + \kappa_0 + \frac{\zeta}{2}\left(1-\frac{1}{\psi}\right)\sigma_d^2 + \frac{\zeta}{2}\kappa_1^2 A_1^2 \sigma_x^2\right),$$
$$A_1 = \frac{1-\frac{1}{\psi}}{1-\kappa_1\rho}. \tag{44}$$

Unconditonal mean of expected returns on wealth is:

$$E[\tilde{s}_{t+1}] = \kappa_0 + (\kappa_1 - 1)A_0 + \mu_d. \tag{45}$$

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

$$\mu_r = E[\tilde{s}_{t+1}] = -\log(\delta) + \frac{1}{\psi}\mu_d - \frac{\zeta}{2}\left(1-\frac{1}{\psi}\right)\sigma_d^2 - \frac{\zeta}{2}\kappa_1^2 A_1^2 \sigma_x^2. \tag{46}$$

### A.1.2 Full Model

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 (28), and investors preferences in (20). Our derivation closely follows the steps in Bansal and Yaron (2004). The log stochastic discount factor is given as:

$$m_{t+1} = \zeta \log(\delta) - \frac{\zeta}{\psi}\Delta\tilde{c}_{t+1} + (\zeta - 1)\tilde{s}_{t+1}. \tag{47}$$

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:

$$\tilde{s}_{t+1} = g_0 + g_1 z_{c,t+1} - z_{c,t} + \Delta\tilde{c}_{t+1}. \tag{48}$$

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:

$$z_{c,t} = A_{c,0} + A_{c,1}x_t + A_{c,2}\sigma_t^2 \tag{49}$$

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

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

$$var_t\left(m_{t+1} + \tilde{s}_{t+1}\right) = \zeta^2 \left(1 - \frac{1}{\psi}\right)^2 \sigma_t^2 + \zeta^2\left(g_1 A_{c,1} \sigma_x\right)^2 + \zeta^2\left(g_1 A_{c,2} \sigma_\varsigma\right)^2. \tag{50}$$

Since $m_{t+1}$ is the stochastic discount factor, $E_t[\exp(m_{t+1} + \tilde{s}_{t+1})] = 1$. This means that:

$$\log\left(E_t[\exp(m_{t+1} + \tilde{s}_{t+1})]\right) = E_t[m_{t+1} + \tilde{s}_{t+1}] + \frac{1}{2} var_t(m_{t+1} + \tilde{s}_{t+1}) = 0. \tag{51}$$

So, we can solve for coefficients $A_{c,0}$, $A_{c,1}$, and $A_{c,2}$ as:

$$A_{c,0} = \frac{\log(\delta) + (1 - \frac{1}{\psi})(\mu_d - \mu_\pi) + g_0 + g_1 A_{c,2}(1 - \varrho)\mu_\varsigma + \frac{1}{2}\zeta g_1^2 (A_{c,1}^2 \sigma_x^2 + A_{c,2}^2 \sigma_\varsigma^2)}{1 - g_1},$$
$$A_{c,1} = \frac{\left(1 - \frac{1}{\psi}\right)\gamma}{1 - g_1 \rho}, \quad A_{c,2} = \frac{\zeta(1 - \frac{1}{\psi})^2}{2(1 - g_1 \varrho)}. \tag{52}$$

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:

$$r_{t+1} = \kappa_0 + \kappa_1 z_{d,t+1} - z_{d,t} + \Delta d_{t+1}, \\ \tilde{r}_{t+1} = \kappa_0 + \kappa_1 z_{d,t+1} - z_{d,t} + \Delta \tilde{d}_{t+1}. \tag{53}$$

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

$$z_{d,t} = A_{d,0} + A_{d,1} x_t + A_{d,2} \sigma_t^2 + A_{d,3}(q_t - \mu_q) \tag{54}$$

Then note that:

$$E_t\left[m_{t+1} + \tilde{r}_{t+1}\right] = \zeta \log(\delta) + (\zeta - 1)\left(g_1 - 1\right) A_{c,0} + (\zeta - 1)\left(g_1 \rho - 1\right) A_{c,1} x_t \\ + (\zeta - 1)\left(g_1 \varrho - 1\right) A_{c,2} \sigma_t^2 + g_1 \left(1 - \varrho\right) A_{d,2} \mu_\varsigma + \left(\zeta - \frac{\zeta}{\psi} - 1\right)(\mu_c + \gamma x_t) \\ + (\zeta - 1) g_0 + \kappa_0 + (\kappa_1 - 1) A_0 + (\kappa_1 \rho - 1) A_{d,1} x_t + (\kappa_1 \varrho - 1) A_{d,2} \sigma_t^2 \\ + \kappa_1 (1 - \varrho) A_{d,2} \mu_\varsigma + (\kappa_1 \theta - 1) A_{d,3}\left(q_t - \mu_q\right) \\ + (\kappa_1 \vartheta - 1) A_{d,4}\left(\Delta \pi_t - \mu_\pi\right) + \mu_c + \frac{1}{1 - \phi} x_t + \frac{\varphi - (1 - \theta)\phi}{1 - \phi}\left(q_t - \mu_q\right).$$

$$var_t\left(m_{t+1}+\tilde{r}_{t+1}\right)=\left(\zeta-1-\frac{\zeta}{\psi}\right)^2\sigma_t^2+\left(\frac{1}{1-\phi}\right)^2\sigma_d^2+\left((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1}\right)^2\sigma_x^2$$
$$+\left((\zeta-1)g_1A_{c,2}+\kappa_1A_{d,2}\right)^2\sigma_\varsigma^2+\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)^2\sigma_q^2$$
$$+2\left((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1}\right)\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{xq}\sigma_x\sigma_q$$
$$+\frac{2}{1-\phi}\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{dq}\sigma_d\sigma_q+\frac{2}{1-\phi}\left((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1}\right)\lambda_{dx}\sigma_d\sigma_x. \tag{55}$$

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:

$$A_{d,0}=\frac{\left(\begin{array}{c}\zeta\log(\delta)+(\zeta-1)g_0+(\zeta-1)A_{c,0}(g_1-1)+((\zeta-1)g_1A_{c,2}+\kappa_1A_{d,2})(1-\varrho)\mu_\varsigma\\+\left(\zeta-\frac{\zeta}{\psi}-1\right)(\mu_c+\gamma x_t)+\kappa_0+\mu_c+\frac{1}{2}(\frac{1}{1-\phi})^2\sigma_d^2+\frac{1}{2}((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1})^2\sigma_x^2\\+\frac{1}{2}((\zeta-1)g_1A_{c,2}+\kappa_1A_{d,2})^2\sigma_\varsigma^2+\frac{1}{2}(\kappa_1A_{d,3}+\frac{\phi}{1-\phi})^2\sigma_q^2\\+((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1})\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{xq}\sigma_x\sigma_q\\+\frac{1}{1-\phi}\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{dq}\sigma_d\sigma_q+\frac{1}{1-\phi}\left((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1}\right)\lambda_{dx}\sigma_d\sigma_x\end{array}\right)}{1-\kappa_1},$$
$$A_{d,1}=\frac{\left(\zeta-1-\frac{\zeta}{\psi}\right)\gamma+(\zeta-1)A_{c,1}(g_1\rho-1)+\frac{1}{1-\phi}}{1-\kappa_1\rho},$$
$$A_{d,2}=\frac{(\zeta-1)(g_1\varrho-1)A_{c,2}+\frac{1}{2}\left(\zeta-1-\frac{\zeta}{\psi}\right)^2}{1-\kappa_1\varrho},\ A_{d,3}=\frac{\psi-(1-\theta)\phi}{(1-\kappa_1\theta)(1-\phi)}. \tag{56}$$

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

$$E_t[r_{t+1}]=A_{r,0}+A_{r,1}x_t+A_{r,2}\sigma_t^2+A_{r,3}(q_t-\mu_q), \tag{57}$$

where:

$$A_{r,0}=\kappa_0-(1-\kappa_1)A_{d,0}+\mu_d+\kappa_1(1-\varrho)A_{d,2}\mu_\varsigma,\quad A_{r,1}=\frac{1}{1-\phi}-(1-\kappa_1\rho)A_{d,1},$$
$$A_{r,2}=-(1-\kappa_1\varrho)A_{d,2},\quad A_{r,3}=\frac{\varphi-(1-\theta)\phi}{1-\phi}-(1-\kappa_1\theta)A_{d,3}=0. \tag{58}$$

## A.2 Kalman Filter

### A.2.1 Estimating Model Parameters using Dividend Dynamics

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:

$$\begin{aligned}\Delta d_{t+1} &= \mu_d + x'_{t+1} + \phi(\Delta e_{t+1} - \mu_d) + \varphi\left(q_t - \mu_q\right) + \sigma_d \epsilon_{d,t+1} \\ x'_{t+1} &= \rho x'_t + \sigma_x \epsilon'_{x,t+1} \\ \begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon'_{x,t+1} \end{pmatrix} &\sim \text{ i.i.d. } \mathbb{N}\left(\mathbf{0}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).\end{aligned} \tag{59}$$

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:

$$\begin{aligned}x'_{t+1|t} &= \rho x'_{t|t}, \quad P'_{t+1|t} = \rho^2 P'_{t|t} + \sigma_x^2, \\ e_{t+1} &= \Delta d_{t+1} - \mu_d - \phi(\Delta e_{t+1} - \mu_d) - \varphi\left(q_t - \mu_q\right), \\ x'_{t+1|t+1} &= x'_{t+1|t} + \frac{P'_{t+1|t}}{P'_{t+1|t} + \sigma_d^2} e_{t+1}, \quad P'_{t+1|t+1} = P'_{t+1|t} - \frac{P^2_{t+1|t}}{P_{t+1|t} + \sigma_d^2}.\end{aligned} \tag{60}$$

To estimate dividend model parameters using data up to time $\tau$, define the log likelihood function $\mathcal{L} = \mathcal{L}_x + \mathcal{L}_q$, where:

$$\begin{aligned}\mathcal{L}_x &= -\sum_{t=0}^{\tau-1} \left(\log\left(P'_{t+1|t} + \sigma_d^2\right) + \frac{e_{t+1}^2}{P'_{t+1|t} + \sigma_d^2}\right), \\ \mathcal{L}_q &= -\sum_{t=0}^{\tau-1} \left(\log(\sigma_q^2) + \left(\frac{q_{t+1} - \theta q_t - (1-\theta)\mu_q}{\sigma_q}\right)^2\right).\end{aligned} \tag{61}$$

That is, $\mathcal{L}_x$ is the log likelihood function for $\{x_{t+1}\}_{t=0}^{\tau-1}$ and $\mathcal{L}_q$ is for $\{q_{t+1}\}_{t=0}^{\tau-1}$.

A caveat in our implementation of Kalman filter is 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. Because, for convenience, this approach ignores the MA(11) structure of the residuals, we refer to this approach as a quasi maximum-likelihood approach,

### A.2.2 Estimating Model Parameters using Both Dividend and Discount Rate Dynamics (Full Information)

We now impose the pricing restriction given in (20) while estimating dividend model parameters. It is important to note that how the pricing restriction should be imposed depends on whether investors learn about dividend dynamics. First, consider the senario where investors have full information of model parameters and do not have to learn. Under this senario, we can re-write price-to-dividend ratio in (29) as:

$$p_t - d_t = A_{d,0} + A_{d,1}x_t + A_{d,2}\sigma_t^2 + A_{d,3}(q_t - \mu_q). \tag{62}$$

Note that under full information, $\mu_q$ and coefficients $A_{d,\cdot}$ are no longer functions of time because investors price assets every period based on the same set of dividend model parameters. So rearranging (29) and substitute it into the second equation of (32) gives:

$$\begin{aligned}&\frac{(p_{t+1} - d_{t+1}) - A_{d,0} - A_{d,1}x_{t+1} - A_{d,3}(q_{t+1} - \mu_q)}{A_{d,2}} - \mu_c \\ &= \varrho\left(\frac{(p_t - d_t) - A_{d,0} - A_{d,1}x_t - A_{d,3}(q_t - \mu_q)}{A_{d,2}} - \mu_c\right) + \sigma_\varsigma\epsilon_{\varsigma,t+1}.\end{aligned} \tag{63}$$

Let $\hat{z}_{d,t} = \frac{(p_t-d_t)-A_{d,0}}{A_{d,2}} - \mu_\varsigma$ and $\hat{q}_t = q_t - \mu_q$, we can re-write (63) as:

$$\begin{aligned}\hat{z}_{d,t+1} &= \frac{A_{d,1}}{A_{d,2}}x_{t+1} + \frac{A_{d,3}}{A_{d,2}}\hat{q}_{t+1} - \varrho\left(\hat{z}_{d,t} + \frac{A_{d,1}}{A_{d,2}}x_t + \frac{A_{d,3}}{A_{d,2}}\hat{q}_t\right) + \sigma_\varsigma\epsilon_{\varsigma,t+1} \\ &= \varrho\hat{z}_{d,t} + \frac{A_{d,1}}{A_{d,2}}(\rho - \varrho)\,x_t + \frac{A_{d,3}}{A_{d,2}}(\theta - \varrho)\,\hat{q}_t + \frac{A_{d,1}}{A_{d,2}}\sigma_x\epsilon_{x,t+1} + \frac{A_{d,3}}{A_{d,2}}\sigma_q\epsilon_{q,t+1} + \sigma_\varsigma\epsilon_{\varsigma,t+1}\end{aligned} \tag{64}$$

Let $\Delta\hat{d}_t = \Delta d_t - \mu_d$, we can substitute $\Delta e_{t+1} - \mu_d = q_{t+1} - q_t + \Delta\hat{d}_{t+1}$ into the first equation in (29) and write:

$$\Delta\hat{d}_{t+1} = \frac{1}{1-\phi}x_t + \frac{\varphi - (1-\theta)\phi}{1-\phi}\hat{q}_t + \frac{1}{1-\phi}\sigma_d\epsilon_{d,t+1} + \frac{\phi}{1-\phi}\sigma_q\epsilon_{q,t+1}. \tag{65}$$

Following the steps in van Binsbergen and Koijen (2010), define the expanded state vector:

$$\mathbf{X_t} = \begin{pmatrix} x_{t-1} \\ \epsilon_{x,t} \\ \epsilon_{q,t} \\ \epsilon_{\varsigma,t} \\ \epsilon_{d,t} \end{pmatrix}. \tag{66}$$

We note that the expanded state vector satisfies: $\mathbf{X_{t+1}} = \mathbf{F X_t} + \mathbf{\Gamma U_{t+1}}$, where:

$$\mathbf{F} = \begin{pmatrix} 1 & \sigma_x & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{\Gamma} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{U_t} = \begin{pmatrix} \epsilon_{x,t} \\ \epsilon_{q,t} \\ \epsilon_{\varsigma,t} \\ \epsilon_{d,t} \end{pmatrix}. \tag{67}$$

Define the vector of observables:

$$\mathbf{Y_t} = \begin{pmatrix} \hat{z}_{d,t} \\ \Delta \hat{d}_t \\ \hat{q}_t \end{pmatrix}, \tag{68}$$

then the measurement equation can be expressed as:

$$\mathbf{Y_{t+1}} = \mathbf{G Y_t} + \mathbf{H X_{t+1}}. \tag{69}$$

where:

$$\mathbf{G} = \begin{pmatrix} \varrho & 0 & \frac{A_{d,3}}{A_{d,2}}(\theta - \varrho) \\ 0 & 0 & \frac{\varphi - (1-\theta)\phi}{1-\phi} \\ 0 & 0 & \theta \end{pmatrix},$$

$$\mathbf{H} = \begin{pmatrix} \frac{A_{d,1}}{A_{d,2}}(\rho - \varrho) & \frac{A_{d,1}}{A_{d,2}}\sigma_x & \frac{A_{d,3}}{A_{d,2}}\sigma_q & \sigma_\varsigma & 0 \\ \frac{1}{1-\phi} & 0 & \frac{\phi}{1-\phi}\sigma_q & 0 & \frac{1}{1-\phi}\sigma_d \\ 0 & 0 & \sigma_q & 0 & 0 \end{pmatrix}. \tag{70}$$

Define $\mathbf{\Sigma} = var(\mathbf{U})$. Set the initial conditions:

$$\mathbf{X_{0|0}} = \mathbf{0}, \quad \mathbf{P_{0|0}} = \begin{pmatrix} \frac{\sigma_x^2}{1-\rho^2} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} \end{pmatrix}. \tag{71}$$

We can then iterate the following system of equations:

$$
\begin{aligned}
\mathbf{X_{t+1|t}} &= \mathbf{F X_{t|t}}, \quad \mathbf{P_{t+1|t}} = \mathbf{F P_{t|t} F^T} + \mathbf{\Gamma \Sigma \Gamma^T}, \\
\boldsymbol{\epsilon}_{\mathbf{t+1}} &= \mathbf{Y_{t+1}} - \mathbf{G Y_t} - \mathbf{H X_{t+1|t}}, \quad \mathbf{S_{t+1}} = \mathbf{H P_{t+1|t} H^T}, \\
\mathbf{K_{t+1}} &= \mathbf{P_{t+1|t} H^T S_{t+1}^{-1}}, \\
\mathbf{X_{t+1|t+1}} &= \mathbf{X_{t+1|t}} + \mathbf{K_{t+1}} \boldsymbol{\epsilon}_{\mathbf{t+1}}, \quad \mathbf{P_{t+1|t+1}} = \left(\mathbf{I} - \mathbf{K_{t+1} H}\right) \mathbf{P_{t+1|t}}.
\end{aligned} \tag{72}
$$

The log likelihood function is:

$$
\mathcal{L} = -\sum_{t=0}^{\tau-1} \left( \log(\det(\mathbf{S_{t+1}})) + \boldsymbol{\epsilon}_{\mathbf{t+1}}^{\mathbf{T}} \mathbf{S_{t+1}^{-1}} \boldsymbol{\epsilon}_{\mathbf{t+1}} \right). \tag{73}
$$

### A.2.3 Estimating Model Parameters using Both Dividend and Discount Rate Dynamics (Learning)

Next, consider the senario where investors have to learn about model parameters from past data. Recall that, under this senario, price-to-dividend ratio is given in (29) as:

$$
p_t - d_t = A_{d,0} + A_{d,1}(t) \cdot x_t + A_{d,2}(t) \cdot \sigma_t^2 + A_{d,3}(t) \cdot (q_t - \mu_q(t)), \; t \in \{0, .., \tau\}. \tag{74}
$$

Assuming investors have to learn about dividend dynamics, $\mu_q(t)$ and coefficients $A_{d,\cdot}(t)$ are functions of time because investors price the stock index in time $t$ using dividend model parameters estimated based on data up to time $t$. In other words, only the price-to-dividend ratio in time $\tau$ is a function of model parameters estimated using data up to time $\tau$. This means that only the price-to-dividend ratios in time $\tau$ helps the econometrician estimate model parameters using data up to time $\tau$. Rearranging (29) gives:

$$
\sigma_\tau^2 = \frac{(p_\tau - d_\tau) - A_{d,0}(\tau) - A_{d,1}(\tau) - A_{d,3}(\tau)(q_\tau - \mu_q(\tau))}{A_{d,2}(\tau)}. \tag{75}
$$

where $A = A_{d,\cdot}(\tau)$ are functions model parameters estimated based on data up to time $\tau$. Because $\sigma_\tau^2$ is normally distributed with mean $\mu_\varsigma$ and standard deviation $\frac{\sigma_\varsigma^2}{1-\varrho^2}$, where $\mu_\varsigma$, $\varrho$, and $\sigma_\varsigma$ are set exogenously based on the calibration of Bansal and Yaron (2004), we can write the log likelihood function for estimating model parameters based on data up to time $\tau$ as $\mathcal{L} = \mathcal{L}_x + \mathcal{L}_q + \mathcal{L}_\sigma$, where $\mathcal{L}_x$ and $\mathcal{L}_q$ are given in (61) and:

$$
\mathcal{L}_\sigma = -\frac{(1-\varrho^2)(\sigma_\tau - \mu_\varsigma)}{\sigma_\varsigma^2}. \tag{76}
$$

## A.3 Simulation

The Null hypothesis, which we reject throughout this paper, is that discount rates are unpredictable (i.e. expected returns are constant). To simulate stock index data under this Null, first simulate innovations to dividend growth rates and earnings-to-dividend ratios:

$$\begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{q,t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \lambda_{dx} & \lambda_{dq} \\ \lambda_{dx} & 1 & \lambda_{xq} \\ \lambda_{dq} & \lambda_{xq} & 1 \end{pmatrix}\right). \tag{77}$$

Dividend model parameters used for simulation 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:

$$\begin{aligned} x_{t+1} &= \rho x_t + \sigma_x \epsilon_{x,t+1} \\ q_{t+1} &= \mu_q + \theta\left(q_t - \mu_q\right) + \sigma_q \epsilon_{q,t+1}. \end{aligned} \tag{78}$$

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

$$\begin{aligned} \Delta d_{t+1} &= \mu_d + \frac{1}{1-\phi}\left(x_t + \phi(\Delta q_{t+1} - \mu_q) + (\varphi - \phi)\left(q_t - \mu_q\right) + \sigma_d \epsilon_{d,t+1}\right) \\ \Delta e_{t+1} &= q_{t+1} - q_t + \Delta d_{t+1}. \end{aligned} \tag{79}$$

To simulate price-to-dividend ratios, recall the Campbell and Shiller (1988a) log-linearized present value relationship:

$$p_t - d_t = \frac{\kappa_0}{1-\kappa_1} + \sum_{s=0}^{\infty} \kappa_1^s \left(E_t[\Delta d_{t+s+1}] - E_t[r_{t+s+1}]\right). \tag{80}$$

Let $\mu_r$ be the constant expected returns of the stock index under the Null that expected returns are constant, the present value relationship can be simplied to:

$$\begin{aligned} p_t - d_t &= \frac{\kappa_0 + \mu_r}{1-\kappa_1} + \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta d_{t+s+1}] \\ &= \frac{\kappa_0 - \mu_r + \mu_d}{1-\kappa_1} + \frac{1}{1-\phi}\left(\frac{1}{1-\kappa_1\rho} x_t + \frac{\varphi - (1-\theta)\phi}{1-\kappa_1\theta}(q_t - \mu_q)\right). \end{aligned} \tag{81}$$

We set $\mu_r = 0.105$ to match average annual stock index return between 1946 and 2013. Finally, we simulate stock index returns from:

$$r_{t+1} = \kappa_0 + \kappa_1(p_{t+1} - d_{t+1}) - (p_t - d_t) + \Delta d_{t+1}. \tag{82}$$

## A.4 On When Investors Receive Earnings Information

In 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 12 and 13. Note that changing this assumption can affect our results through its effect on long run dividend growth expectations and investors' beliefs about persistence $\omega$ of dividend growth rates. Nevertheless, results show that the significance of our findings that investors' learning about dividend dynamics is reflected in the expected returns of the stock index is robust to changes in this assumption.

| | 3 Months Lag | 6 Months Lag | 9 Months Lag | 12 Months Lag |
|---|---|---|---|---|
| $sy_t$ (Learning) | $4.543^{**}$ [0.012] | $4.327^{**}$ [0.016] | $4.128^{**}$ [0.019] | $4.052^{**}$ [0.022] |
| $R^2$ | 0.151 | 0.146 | 0.139 | 0.137 |

Table 12: **Predicting Stock Index Return using Stock Yield (Additional Lags to Earnings Information)**: This table reports coefficient estimates and $R$-square values from regressing stock index return over the year on stock yield, computed from long run dividend growth expectations implied by our dividend model and assuming investors learn about model parameters. Regressions are based on data between 1976 and 2013. Estimating dividend dynamics is based on data since 1946. When estimating dividend dynamics, we assume that investors receive earnings information 6,9, or 12 months after fiscal quarter or year end. Simulated $p$-values are reported in square parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.

## A.5 Estimating Dividend Dynamics using a Rolling Window of Dividend Data

We show that estimating dividend model parameters out-of-sample using an expanding window of dividend data performs better than using a rolling window of data, for the

| | 3 Months Lag | 6 Months Lag | 9 Months Lag | 12 Months Lag |
|---|---|---|---|---|
| $\omega(t)$ | -0.644** | -0.624** | -0.576** | -0.501** |
| | [0.014] | [0.017] | [0.025] | [0.042] |
| $p_t - d_t$ | -0.193*** | -0.191*** | -0.184*** | -0.173*** |
| | [0.002] | [0.002] | [0.003] | [0.006] |
| $R^2$ | 0.235 | 0.220 | 0.202 | 0.179 |

Table 13: **Stock Index Returns and Investors' Beliefs about Persistence of Dividend Growth Rates (Additional Lags to Earnings Information).** This table reports the coefficient estimates and $R$-square values from regressing stock index returns over the next year on investors' belief about persistence $\omega$ of dividend growth rate and log price-to-dividend ratios. Regressions are based on data between 1976 and 2013. Estimating dividend dynamics is based on data since 1946. When estimating dividend dynamics, we assume that investors receive earnings information 6,9, or 12 months after fiscal quarter or year end. Simulated $p$-values are reported in square parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.

purposes of both forecasting future dividends and stock index returns. In Table 15, we report the out-of-sample $R^2$ values for predicting annual dividend growth rates using expected growth rates implied by our model, with model parameters estimated using a rolling window of past 10, 20, or 30 years of data. Results confirm that the $R^2$ value for predicting dividend growth rates is highest when model parameters are estimated using an expanding window. In absolute terms, however, out-of-sample $R^2$ value is still 27.0 (28.6) percent when parameters are estimated using a rolling window of 20 (30) years. This shows that our model's superior performance 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 best captures investors' learning behavior. In Table 14, we repeat key results in Tables 6 and 9, but use a rolling window of past 10, 20, or 30 years, instead of an expanding window, when estimating dividend dynamics out-of-sample. Results confirm that investor's learning behavior is best captured when model parameters are estimated using an expanding window of past data.

| 10 Years | | 20 Years | | 30 Years | |
|---|---|---|---|---|---|
| $R^2$ | $p$-value | $R^2$ | $p$-value | $R^2$ | $p$-value |
| 0.083 | 0.085 | 0.270 | 0.001 | 0.286 | 0.001 |

Table 14: **Predicting Dividend Growth Rates using Model Implied Expected Growth Rates (Rolling Window Estimation)**: This table reports the out-of-sample $R$-square for predicting dividend growth rates using expected growth rates implied by our model and the corresponding $p$-value from the $F$-test for model significance.Statistics are based on data between 1975 and 2013. Estimating dividend dynamics is based on a rolling window of past 10, 20, or 30 years of data.

| | 10 Years | | 20 Years | | 30 Years | |
|---|---|---|---|---|---|---|
| $sy_t$ (Learning) | 1.231 [0.201] | | 2.579*** [0.005] | | 3.209** [0.018] | |
| $\omega(t)$ | | -0.111 [0.350] | | -0.222 [0.221] | | -0.682** [0.010] |
| $p_t - d_t$ | | -0.054 [0.204] | | -0.118** [0.037] | | -0.201*** [0.001] |
| $R^2$ | 0.053 | 0.107 | 0.164 | 0.144 | 0.135 | 0.253 |

Table 15: **Stock Index Returns, Stock Yields, and Investors' Beliefs about Persistence of Dividend Growth Rates (Rolling Window Estimation)**: This table reports coefficient estimates and $R$-square values from regressing stock index return over the next year on stock yield, log price-to-dividend ratios, and investors' beliefs about persistence $\omega$ of dividend growth rates. Regressions are based on data between 1976 and 2013. Estimating dividend dynamics is based on a rolling window of past 10, 20, or 30 years of data. Simulated $p$-values are reported in square parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.