Dividend Dynamics, Learning, and Expected Stock Index Returns*

Ravi Jagannathan
Northwestern University and NBER

Binying Liu
Northwestern University

April 14, 2016

Abstract

We show that, in a perfect and efficient securities market, an asset pricing model that better describes investors' behavior should predict stock index returns better than other models. We propose a dividend model that predicts, out-of-sample, 31.3 percent of the variation in annual dividend growth rates between 1976 and 2013. We demonstrate that, when learning about dividend dynamics is incorporated into a long run risks model, the model predicts, out-of-sample, 22.3 percent of the variation in annual stock index returns during the same time frame. This supports the view that both investors' aversion to long run risks in dividends and learning about these risks play critical roles in determining asset prices and expected returns.

*We thank Jonathan Berk, Jules van Binsbergen, Wayne Ferson, Lawrence Harris, Gerard Hoberg, Narayan Naik, and seminar participants at London Business School and University of Sourthern California, for helpful comments and suggestions.

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 risks in equities that require such a large premium and why the risk free rate is so low. A common feature in most of these models is that the risk premium on equities does not remain constant over time, but varies in a systematic and stochastic manner. A large number of academic studies have found support for such predictable variation in the equity premium.[1] This led Lettau and Ludvigson (2001) to conclude "it is now widely accepted that excess returns are predictable by variables such as price-to-dividend ratios."

Goyal and Welch (2008) argue that variables such as price-to-dividend ratios, although successful in predicting stock index returns in-sample, fail to predict returns out-of-sample. The difference between in-sample and out-of-sample prediction is the assumption made on investors' information set. Traditional dynamic equilibrium asset pricing models assume that, while investors' beliefs about investment opportunities and economic conditions change over time and drive the variation in stock index prices and expected returns, they have full knowledge of the parameters describing the economy. For example, these models assume that investors know the true model and model parameters governing consumption and dividend dynamics. However, as Hansen (2007) argues, "this assumption has been only a matter of analytical convenience" and is unrealistic in that it requires us to "burden the investors with some of the specification problems that challenge the econometrician". Motivated by this insight, a recent but growing literature has focused on the role of learning in asset pricing models. Timmermann (1993) and Lewellen and Shanken (2002) demonstrate, via simulations, that parameter uncertainty can lead to excess predictability and volatility in stock returns. Johannes, Lochstoer, and Mou (2015) propose a Markov-switching model for consumption dynamics and shows that learning about the consumption process is reflected in asset prices and returns. Croce, Lettau, and Ludvigson (2014) show that a bounded rationality limited information long-run risks model can generate a downward-sloping equity term structure. Johannes, Lochstoer, and

[1] See, among others, Campbell and Shiller (1988b), Fama and French (1993), Lamont (1998), Baker and Wurgler (2000), Lettau and Ludvigson (2001), Campbell and Vuolteenaho (2004), Lettau and Ludvigson (2005), Polk, Thompson, and Vuolteenaho (2006), Ang and Bekaert (2007), van Binsbergen and Koijen (2010), Kelly and Pruitt (2013), van Binsbergen, Hueskes, Koijen, and Vrugt (2013), Li, Ng, and Swaminathan (2013), and Da, Jagannathan, and Shen (2014), Glosten, Jagannathan, and Runkle (1993), Breen, Glosten, and Jagannathan (1989).

Collin-Dufresne (2015) provide the theoretical foundation that parameter learning can be a source of long-run risk under Bayesian learning. We add to this literature.

The contributions of our paper, which distinguish it from the existing literature on the interaction between learning and asset pricing, is as follows. First, we show that, when equity markets are perfect and efficient, an asset pricing model that is closer to the true asset pricing model, i.e. the model that better describes investors' behavior, should predict stock index returns better. This provides the theoretical foundation for the economic significance of return predictability in assessing an asset pricing model. Then, we show that, when learning about dividend dynamics is incorporated into a long run risks model, the model's ability to predict annual stock index returns substantially improves from an R-square value of 11.8 percent to 22.3 percent, out-of-sample. This addresses the Goyal and Welch (2008) critique and significantly revises upward the degree of return predictability documented in the existing literature. Further, this lend support to the view that both investors' aversion to long run risks and learning about these risks play important roles in determining asset prices and returns.[2]

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

We document that uncertainties about parameters in our dividend model, especially the parameter governing the persistence of the latent variable, are high and resolve slowly. That is, these uncertainties remain substantial even at the end of our 68 years data sample, suggesting that learning about dividend dynamics is difficult. Further, when

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

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

We provide evidence that investors behave as if they learn about dividend dynamics and price stocks using our model. First, 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 time variation in stock index returns over the next year. In comparison, stock yields, assuming no learning, explain only 10.5 percent of the same variation. Next, we embed our dividend model into an dynamic equilibrium asset pricing model that features Epstein and Zin (1989) preferences, which capture preferences for the timing of resolution of uncertainty, and consumption dynamics from the long-run risk model of Bansal and Yaron (2004). We refer to this model as our long-run risk model. We find that, between 1976 and 2013, expected returns derived from our long-run risk model, assuming that investors learn about the parameters governing the dividend process, predict 22.3 percent of the variation in annual stock index returns. Learning accounts for about half of the 22.3 percent. The incremental contribution of learning to stock return predictability is statistically significant.

We follow Cogley and Sargent (2008), Piazzesi and Schneider (2010), and Johannes, Lochstoer, and Mou (2015), and define learning based on the anticipated utility of Kreps (1998), where agents update using Bayes' law but optimize myopically in that they do not take into account uncertainties associated with learning in their decision making process. That is, anticipated utility assumes agents form expectations not knowing that their beliefs will continue to evolve going forward in time as the model keeps updating. Given the relative complexity of our asset pricing model and the multi-dimensional nature of learning, we find that solving our model with parameter uncertainties as additional risk factors is too computationally prohibitive.[3] Therefore, we adopt the anticipated utility approach as the more convenient alternative.

The rest of this paper is organized as follows. In Section 1, we introduce our dividend model and evaluate its performance in capturing dividend dynamics. In Section 2, we show that investors beliefs about dividend model parameters can vary significantly over time as a result of Kreps' learning about dividend dynamics. In Sections 3, we explain that a model's performance in predicting future stock index returns can be used as a criterion to evaluate asset pricing models. We then demonstrate that a model that incorporates Kreps' learning better captures the time variation of stock index returns and explain why such a finding provides us insight into investors' preferences and the role of learning in describing investors' behavior. In Section 4, 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.

Define d_t as log aggregate nominal dividend of the stock index and $\Delta d_{t+1} = d_{t+1} - d_t$ as 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), we fit our model to the nominal dividend process. As shown in Boudoukh, Michaely, Richardson, and Roberts (2007), equity issuance and repurchase tend to be sporadic and random compared to cash dividends. For this reason, we focus on modeling the cash dividend process and treat equity issuance

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

and repurchase as unpredictable. In this model, time-t is defined in years to control for potential seasonality in dividend payments and expected nominal dividend growth rates follow a stationary AR[1] process and are functions of the latent variable x_t, the unconditional mean μ_d of dividend growth rates, and the persistence coefficient ρ of x_t, 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 nominal price of the stock index, e_t as log nominal earnings, π_t as log consumer price index, and, following Campbell and Shiller (1988b), consider the following vector-autoregression for annual nominal dividend growth rates, log price-to-dividend ratios, and CAPE ratios:

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

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

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

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

We report estimates of b_{10}, b_{11}, b_{12}, and b_{13} from (3), based on data between 1946 and 2013, in the first four columns of Table 1.[4] Consistent with Campbell and Shiller (1988b), we find that both price-to-dividend ratios and CAPE ratios have significant effects on future dividends, but in the opposite direction. That is, increases in price-to-dividend ratios predict increases in future dividend growth rates, but increases in CAPE ratios predict decreases in future dividend growth rates. Interestingly, we see from Table 1 that $b_{12} + b_{13} = 0$ cannot be statistically rejected. For this reason, we restrict $b_{13} = -b_{12}$ and re-estimate annual dividend growth rates as:

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

[4] Throughout this paper, we report results based on overlapping monthly data. That is, in each month, we fit or predict dividend growth rates and stock index returns over the next 12 months. We report standard errors, F-statistics, p-values, and Q-statistics adjusted to reflect the dependence introduced by overlapping monthly data.

We note that the stock index price does not appear in (5). We report estimated coefficients from (5) in the last three columns of Table 1. Results show that the β_2 estimate is highly statistically significant, suggesting that expected dividend growth rates respond to the log ratios between historical earnings and dividends. Intuitively, high earnings relative to dividends implies that firms have been retaining earnings in the past and so are expected to pay more dividends in the future.

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

Table 1: **Campbell and Shiller (1988b) Betas for Predicting Dividend Growth Rates**: This table reports coefficients from estimating dividend growth rates using (3) and (5), based on data between 1946 and 2013. Newey and West (1987) adjusted standard errors are reported in parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.

We extend (1) based on this insight that earnings contain information about future dividends. Define $\Delta e_{t+1} = e_{t+1} - e_t$ as log nominal earnings growth rate and $q_t = e_t - d_t$ as log earnings-to-dividend ratio, our dividend model can be described by the following system of equations:

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

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

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

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

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

This means that, aside from the two state variables, expected dividend growth rates are functions of the unconditional means μ_d and μ_q of dividend growth rates and earnings-to-dividend ratios, the persistence ρ and θ of the latent variable x_t and earnings-to-dividend ratios, and coefficients ϕ and φ that connect earnings information to dividend dynamics. We note that earnings dynamics is not modeled explicitly in (6). However, we can solve for nominal earnings growth rates from the processes for dividend growth rates and earnings-to-dividend ratios:[5]

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

1.1 Data and Estimation

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

We compute the likelihood of our dividend model using Kalman filters (Hamilton (1994)) and estimate model parameters,

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

based on maximum-likelihood. See Appendix A.3 for details. Table 2 reports model parameter estimates based on data between 1946 and 2013. Standard errors are based on bootstrap simulation. Previous works have suggested a regime shift in dividend dynamics before and after World War II. Fama and French (1988) note that dividends are more smoothed in the post-war period. Chen, Da, and Priestley (2012) argue that the lack

[5] One can verify that, substituting (6) and (9) into $q_{t+1} = e_{t+1} - d_{t+1}$ yields the third equation in (5).

of predictability in dividend growth rates by price-to-dividend ratios in the post-war period is attributable to this dividend smoothing behavior. Consistent with our intuition, both ϕ and φ that connect earnings information to dividend dynamics are estimated to be positive and highly statistically significant. That is, high contemporaneous earnings growth rates imply high dividend growth rates, and high past earnings-to-dividend ratios imply high dividend growth rates. The annual persistence of earnings-to-dividend ratios is estimated to be 0.281. The latent variable x_t is estimated to be more persistent at 0.528. In summary, there is a moderate level of persistence in nominal dividend growth rates between 1946 and 2013 based on estimates from our model.

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

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

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

In the first column of Table 4, we report the goodness-of-fit for describing nominal dividend growth rates using our dividend model, based on data between 1946 and 2013. We find that our model explains 55.0 percent of the variation in annual nominal dividend growth rates, which represents a significant improvement over the baseline models. We

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

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

know that at least part of this improved fit comes from adding more parameters to existing models and is thus mechanical. Thus, to address the concern that our model overfits the data, and that our model uses more data than the baseline models, we also assess our model based on how it forecasts dividend growth rates out-of-sample. That is, instead of fitting the model based on the full data sample, we predict dividend growth rates at each point in time based on data available at the time. Forecasting performance is then evaluated using the out-of-sample R-square value as defined in Goyal and Welch (2008):

$$R^2 = 1 - \frac{\frac{1}{T-T_0+1}\sum_{t=T_0}^{T-1}\left(\Delta d_{t+1} - E_t[\Delta d_{t+1}|\text{Dividend Model}]\right)^2}{\frac{1}{T-T_0+1}\sum_{t=T_0}^{T-1}\left(\Delta d_{t+1} - \hat{\mu}_d(t)\right)^2}, \tag{10}$$

where $\hat{\mu}_d(t)$ is the average of dividend growth rates up to time-t:

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

where we use time-0 to denote the start of the data sample, time-T_0 to denote the end of the training period, and time-T to denote the end of the data sample. Due to the relative complexity of our model, we use the first 30 years of our data sample as the training period so that out-of-sample prediction is for the period between 1976 and 2013. Throughout this paper, for predictive analysis, we assume investors have access to earnings information 3 months after fiscal quarter or year end. The choice of 3 months is based on Securities and Exchange Commission (SEC) rules since 1934 that require public companies to file 10-Q reports no later than 45 days after fiscal quarter end and 10-K reports no later than 90 days after fiscal year end.[6] To show that our findings are robust to this assumption, we repeat the main results of this paper in Appendix A.1, 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.[7] In the third and fourth columns of Table 4, we report the out-of-sample R-square value for predicting annual dividend growth rates and the corresponding p-value from the adjusted-MSPE statistic of Clark and West (2007). Results show that our model predicts 31.3 percent of the variation in annual nominal dividend growth rates, which is a significant improvement over the R-square values of 18.5 percent and 13.5 percent from the baseline models. Interestingly, we note that imposing the restriction that $b_{12} + b_{13} = 0$ in (3), so that price drops out of the equation, significantly improves the out-of-sample forecasting performance of the Campbell and Shiller (1988b) model, lending additional support for our decision to impose this restriction.

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

Table 4: **Dividend Growth Rates and Expected Growth Rates**. The first and second columns of this table report goodness-of-fit for describing nominal dividend growth rates using our dividend model (i.e. J&L), the dividend model in van Binsbergen and Koijen (2010) (i.e. vB&K), the dividend model in Campbell and Shiller (1988b) (i.e. C&S), or its restricted version where we set $b_{12} + b_{13} = 0$, and the corresponding p-value. The second column reports the Bayesian information criterion. The third and fourth columns report the out-of-sample R-square value for predicting dividend growth rates and the corresponding p-value from the adjusted-MSPE statistic of Clark and West (2007). In-sample (out-of-sample) statistics are based on data between 1946 and 2013 (1976 and 2013).

To examine that the difference in dividend growth rates predictability between using

[6] 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.

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

our dividend model and using one of the baseline models is statistically significant, we define incremental R-square value for predicting dividend growth rates using our model over one of the two alternative models as:

$$(\mathrm{I})R^2 = 1 - \frac{\frac{1}{T-T_0+1}\sum_{t=T_0}^{T-1}\left(\Delta d_{t+1} - E_t[\Delta d_{t+1}|\mathrm{J\&L}]\right)^2}{\frac{1}{T-T_0+1}\sum_{t=T_0}^{T-1}\left(\Delta d_{t+1} - E_t[\Delta d_{t+1}|\text{Baseline Model}]\right)^2}, \tag{12}$$

and report statistics in Table 5. If this incremental R-square value is significantly positive, it suggests that our dividend model is an improvement over baseline models in capturing the variation in annual dividend growth rates. We note that the superior performance of our dividend model over the alternatives is statistically significant at the 95% confidence level.

Incremental R-squared of Our Dividend Model vs a Baseline Model			
vB&K		C&S	
R^2	p-value	R^2	p-value
0.132	0.048	0.201	0.012

Table 5: **Performance Difference across Dividend Models.** This table reports the out-of-sample incremental R-squared, defined in (9), for predicting nominal dividend growth rates over the next year between using expected rates implied by our dividend model or one of the two alternative models, i.e. the model in van Binsbergen and Koijen (2010) (i.e. vB&K) and the model in Campbell and Shiller (1988b) (i.e. C&S), and the corresponding p-value from the adjusted-MSPE statistic of Clark and West (2007). Statistis are based on data between 1976 and 2013.

Although results in this section show that our model is successful in capturing the time variation in nominal dividend growth rates both in-sample and out-of-sample, we recognize that it inevitably simplifies the true process governing dividend dynamics. One can add additional lags of earnings-to-dividend ratios to the model.[8] Also, one can extend our model by allowing model parameters, such as the persistence ρ of the latent variable x_t or the standard deviation σ_x of shocks to x_t, to be time varying. However, the disadvantage of incorporating such extensions is that a more complicated model is also more difficult to estimate with precision in finite sample. For example, one way to assess whether accounting for the possibilities of time varying model parameters improves our model's out-of-sample forecasting performance is to estimate model parameters based on a rolling window, rather than an expanding window, of past data, so that observations from the

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

distant past are not used to estimate model parameters. We provide this analysis in Appendix A.1. In summary, we find that our model's forecasting performance is not improved by estimating model parameters based on a rolling window of past data.

1.2 Inflation and Real Rates

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

$$\Delta\pi_{t+1} - \mu_\pi = \eta\left(\Delta\pi_t - \mu_\pi\right) + \sigma_\pi \epsilon_{\pi,t+1},\ \epsilon_{\pi,t+1} \sim \mathbb{N}(0,1). \tag{13}$$

We can then derived the expression for expected real dividend growth rates based on expected nominal rates and inflation rates as:

$$E_t[\Delta\tilde{d}_{t+s+1}] = (\mu_d - \mu_\pi) + \frac{\rho^s}{1-\phi} x_t + \frac{\theta^s(\varphi - (1-\theta)\phi)}{1-\phi}(q_t - \mu_q) - \eta^{s+1}\left(\Delta\pi_t - \mu_\pi\right),\ \forall s \geq 0. \tag{14}$$

where $\Delta\tilde{d}_t = \Delta d_t - \Delta\pi_t$ denotes real dividend growth rate.

μ_π	η	σ_π
0.038	0.555	0.026
(0.013)	(0.109)	(0.020)

Table 6: **Inflation Model Parameters**: This table reports estimated parameters from our dividend model, based on data between 1946 and 2013. Bootstrap simulated standard errors are reported in parentice. Simulation is based on 100,000 iterations.

In Table 7, we repeat the statistics reported in Table 4, but for fitting and predicting annual real, rather than nominal, dividend growth rates. We find that our model also outperforms the alternative baseline models in capturing the variation in annual real dividend growth rates, i.e. 48.2 percent of the variation in real dividend growth rates in-sample and 27.2 percent of the variation out-of-sample.

To provide more details on how various types of shocks to dividend growth rates and inflation at a given time affect expected real dividend growth rates going forward, based on our model and estimated parameters, reported in Table 3, we consider an one unit

	In-Sample		Out-of-Sample	
	Goodness-of-Fit	p-value	R^2	p-value
J&L	0.482	0.000	0.272	0.001
vB&K	0.203	0.000	0.166	0.012
C&S	0.281	0.000	0.157	0.015
C&S (Restricted)	0.273	0.000	0.217	0.003

Table 7: **Dividend Growth Rates and Expected Growth Rates (Real Rates)**. The first and second columns of this table report goodness-of-fit for describing real dividend growth rates using our dividend model (i.e. J&L), the dividend model in van Binsbergen and Koijen (2010) (i.e. vB&K), the dividend model in Campbell and Shiller (1988b) (i.e. C&S), or its restricted version where we set $b_{12} + b_{13} = 0$, and our inflation model, and the corresponding p-valuel. The second column reports the Bayesian information criterion. The third and fourth columns report the out-of-sample R-square value for predicting dividend growth rates and the corresponding p-value from the adjusted-MSPE statistic of Clark and West (2007). In-sample (out-of-sample) statistics are based on data between 1946 and 2013 (1976 and 2013).

change to $\epsilon_{d,t}$, $\epsilon_{x,t}$, $\epsilon_{q,t}$, or $\epsilon_{\pi,t}$ and show how such a change affects expected real dividend growth rates up to 10 years into the future, i.e. time-t to time-$(t + 10)$. We report these impulse response functions in Figure 3. We note that shocks $\epsilon_{x,t}$, $\epsilon_{q,t}$, and $\epsilon_{\pi,t}$ have varying degrees of persistent effects on future dividend growth rates with shocks to $\epsilon_{x,t}$ being most persistent. On the other hand, shock $\epsilon_{d,t}$ is purely transient. Further, shock $\epsilon_{\pi,t}$ affects future real dividend growth expectations negatively so that an increase in inflation rate reduces future expected real dividend growth rates.

2 Parameter Uncertainty and Learning

The difference between in-sample and out-of-sample prediction is the assumption made on investors' information set. Model parameters reported in Table 2 are estimated using data up to 2013, so they reflect investors' knowledge of dividend dynamics at the end of 2013. That is, if investors were to estimate our model in an earlier date, they would have estimated a set of parameter values different from those reported in Table 2. This is a result of investors' knowledge of dividend dynamics evolving as more data become available. We call this learning. That is, we use learning to refer to investors estimating model parameters at each point in time based on data available at the time. In this section,

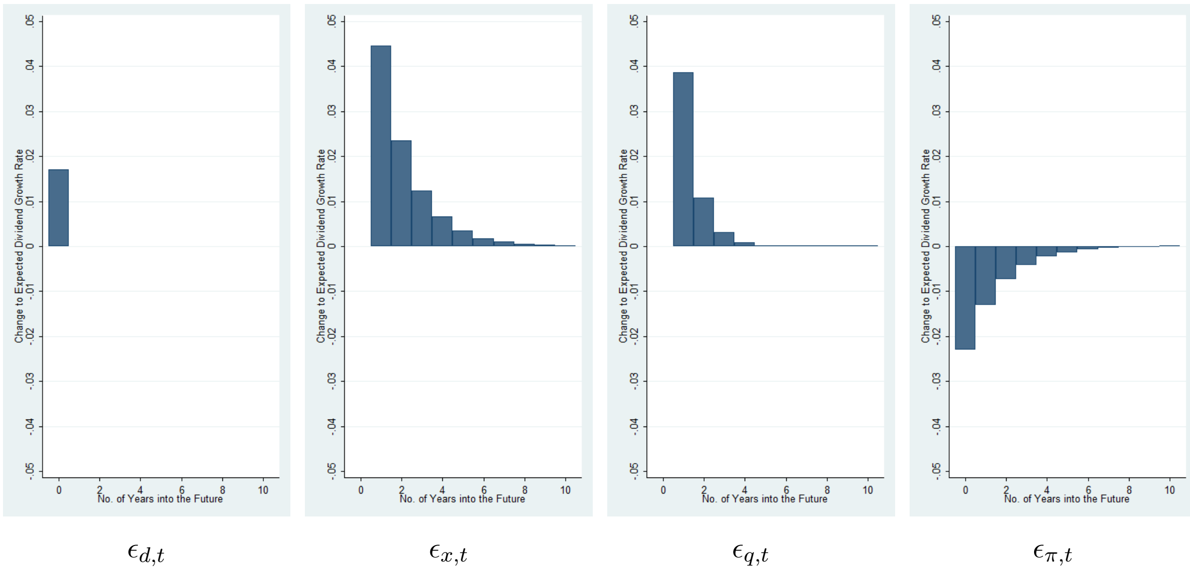
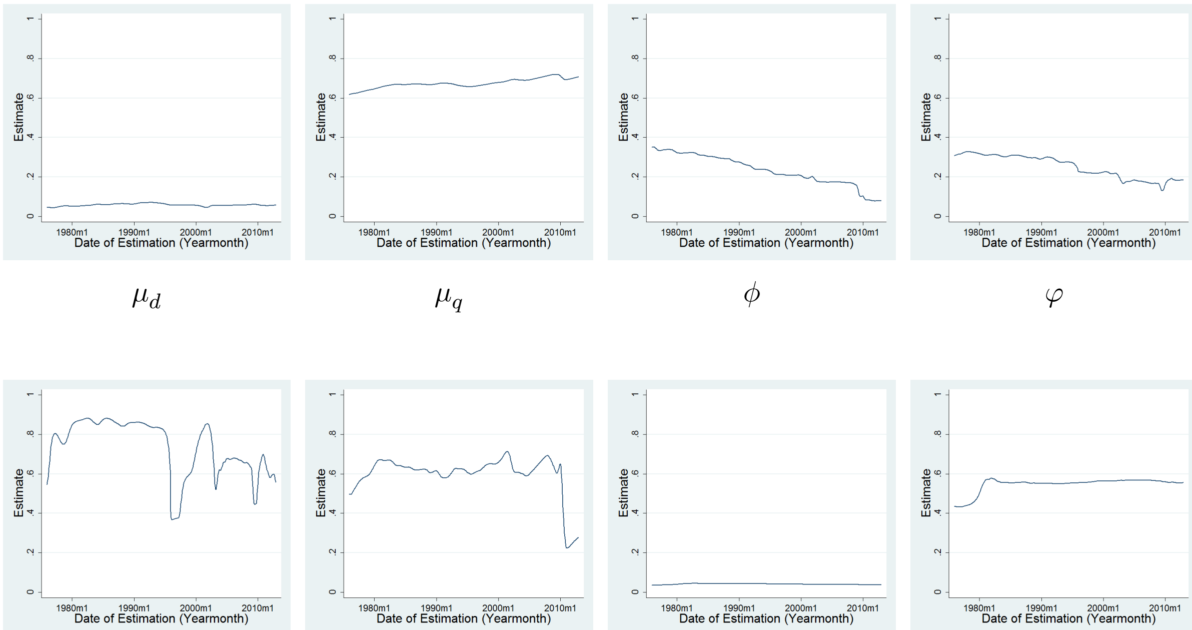


Figure 1: **Impulse Response Functions on Shocks that Affect Real Dividend Growth Rates.** This figure plots the changes to expected dividend growth rates over the next 10 years due to a unit change in shocks to dividend growth rates: $\epsilon_{d,t}$, $\epsilon_{x,t}$, $\epsilon_{q,t}$, and $\epsilon_{\pi,t}$.

we summarize how learning affects investors' beliefs about the parameters governing the dividend process, assuming that investors learn about dividend dynamics using our model. We then show that learning can have significant asset pricing implications.

In Figure 2, we report estimates of the eight model parameters in (14) that affect expected dividend growth rates, assuming that our model is estimated based on data up to time-τ, for τ between 1976 and 2013. There are several points we take away from Figure 2. First, there is a gradual upward drift in investors' beliefs about the unconditional mean μ_q of earnings-to-dividend ratios. This suggests that firms have been paying a smaller fraction of earnings as cash dividends in recent decades. Second, there are gradual downward drifts in investors' beliefs about ϕ and φ that connect earnings information to dividend dynamics. This means that dividends have become more smoothed over time. Third, a sharp drop in investors' beliefs about the persistence θ of earnings-to-dividend ratios towards the end of our data sample is due to the abnormally low earnings reported in late 2008 and early 2009 as a result of the financial crisis and the strong stock market recovery that followed. Also, estimates of inflation model parameters are relatively stable over time.

It is clear from Figure 2 that the persistence ρ of the latent variable x_t is the parameter hardest to learn and least stable over time. This observation is consistent with results reported in Table 2, which show that, of all model parameters, ρ is estimated with the



μ_d μ_q ϕ φ

ρ θ μ_π η

Figure 2: **Evolution of Dividend and Inflation Model Parameters.** This figure plots estimates of the eight dividend model parameters that affect expected real dividend growth rates, assuming that these parameters are estimated based on data up to time-τ for τ between 1976 and 2013.

highest standard error (i.e. 0.160). Investors' beliefs about ρ fluctuate significantly over the sample period, especially around three periods during which beliefs about ρ sharply drop. The first is at the start of dot-com bubble between 1995 and 1998. The second is during the crash of that bubble in late 2002 and early 2003. The third is during the financial crisis in late 2008 and early 2009. Further, there is also a long term trend that sees a gradual decrease in investors' beliefs about ρ since early 1980s. For example, if we were to pick a random date between 1976 and 2013 and estimate our model based on data up to that date, on average we would have estimated a ρ of 0.734.[9] This would be significantly higher than the 0.528 reported in Table 2 that is estimated using the full data sample.

We can infer, from standard errors reported in Table 2, that learning about dividend dynamics is a slow process. That is, even with 68 years of data, there are still significant uncertainties surrounding the estimates of some model parameters. For example, the 95 percent confidence interval for the persistence ρ of the latent variable x_t is between 0.214 and 0.842. The same confidence interval for the persistence θ of earnings-to-dividend

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

ratios is between 0.054 and 0.508. To quantify the speed of learning, following Johannes, Lochstoer, and Mou (2015), for each of the eight parameters that affect expected dividend growth rates, we construct a measure that is the inverse ratio between the bootstrap simulated standard error assuming that the parameter is estimated based on data up to 2013 and the bootstrap simulated standard error assuming that the parameter is estimated based on 10 additional years of data (i.e. if the parameter were estimated in 2023). See Appendix A.2 for details on bootstrap simulation. In other words, this ratio reports how much an estimated parameter's standard error would reduce if investors were to have 10 more years of data. So the closer this ratio is to 1, the more difficult it is for investors to learn about the parameter. In Table 8, 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 2 and those reported in Table 2, we find that it is more difficult to learn about ρ than about any of the other five model parameters.

μ_d	μ_q	ϕ	φ	ρ	θ	μ_π	η
0.924	0.924	0.926	0.928	0.951	0.920	0.923	0.918

Table 8: **Speed of Learning about Model Parameters**: This table reports the speed of learning for the six model parameters that affect expected dividend growth rates. Speed of learning is defined as the inverse ratio between the bootstrap simulated standard error assuming that the parameter is estimated based on data up to 2013 and the bootstrap simulated standard error assuming that the parameter is estimated based on 10 additional years of data (i.e. if the parameter were estimated in 2023). Simulation is based on 100,000 iterations.

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

$$p_t - d_t = \frac{\kappa_0}{1-\kappa_1} + \sum_{s=0}^{\infty} \kappa_1^s \left(E_t[\Delta d_{t+s+1}] - E_t[r_{t+s+1}] \right), \tag{15}$$

where κ_0 and κ_1 are log-linearizing constants and r_{t+1} is the stock index's log return.[10] 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

[10] 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 $\kappa_1 = 0.970$.

current price of the stock index. That is, rearranging (15), we can write stock yield as:

$$
\begin{aligned}
sy_t &\equiv (1-\kappa_1)\sum_{s=0}^{\infty}\kappa_1^s E_t[\Delta r_{t+s+1}] \\
&= \kappa_0 - (1-\kappa_1)(p_t - d_t) + (1-\kappa_1)\sum_{s=0}^{\infty}\kappa_1^s E_t[\Delta d_{t+s+1}]. \qquad (16)
\end{aligned}
$$

Define long run dividend growth expectation as:

$$
\partial_t \equiv (1-\kappa_1)\sum_{s=0}^{\infty}\kappa_1^s E_t[\Delta d_{t+s+1}]. \qquad (17)
$$

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 (17) as:

$$
\begin{aligned}
\partial_t &= (1-\kappa_1)\sum_{s=0}^{\infty}\kappa_1^s\left(\mu_d + \frac{\rho^s}{1-\phi}x_t + \frac{\theta^s(\varphi-(1-\theta)\phi)}{1-\phi}(q_t-\mu_q)\right) \\
&= \mu_d + \frac{1-\kappa_1}{1-\phi}\left(\frac{1}{1-\kappa_1\rho}x_t + \frac{\varphi-(1-\theta)\phi}{1-\kappa_1\theta}(q_t-\mu_q)\right). \qquad (18)
\end{aligned}
$$

If a different dividend model is used, long run dividend growth expectations will also be different. For example, if we assume that dividend growth rates follow a white noise process centered around μ_d, we can re-write (17) instead as $\partial_t = \mu_d$. Further, because long run dividend growth expectations are functions of dividend model parameters, it is also affected by whether model parameters are estimated once based on the full data sample, or estimated at each point in time based on data available at the time. The first case corresponds to investors having full knowledge of the parameters describing the dividend process, whereas the second case corresponds to investors having to learn about dividend dynamics. In this paper, we assume that when investors learn, they do not impose a prior and use the classical estimator of model parameters. We relax this assumption and consider investors' learning under non-flat priors in Appendix A.1. We document similar results. 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, assuming that investors learn about dividend dynamics using our model.

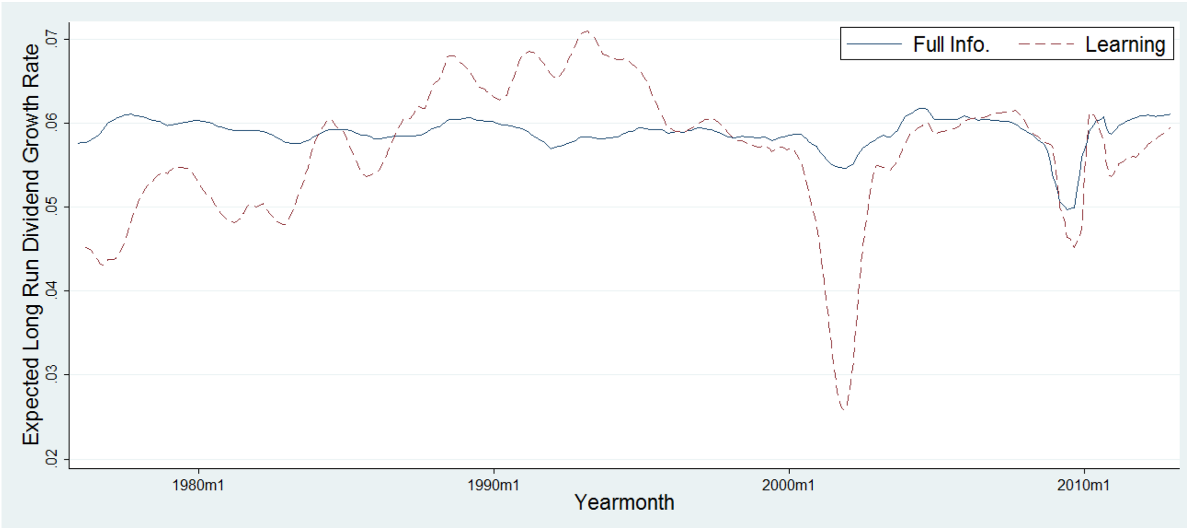


Figure 3: **Expected Long Run Dividend Growth Rates.** This figure plots long run dividend growth expectations, computed using our dividend model, for the period between 1976 and 2013. Dividend model parameters are estimated based on data since 1946. Under full information, model parameters are estimated once based on the full data sample. Under learning, those parameters are estimated at each point in time based on data available at the time.

In Figure 4, we plot stock yields, computed by substituting (18) into (16):

$$sy_t = \kappa_0 - (1-\kappa_1)(p_t - d_t) + \mu_d + \frac{1-\kappa_1}{1-\phi}\left(\frac{1}{1-\kappa_1\rho}x_t + \frac{\varphi - (1-\theta)\phi}{1-\kappa_1\theta}(q_t - \mu_q)\right). \tag{19}$$

Dividend model parameters are either estimated once based on the full data sample or estimated at each point in time based on data available at the time. We also plot price-to-dividend ratios in Figure 4, and scale price-to-dividend ratios to allow for easy comparison to stock yields. We find that there is almost no noticeable difference between the time series of price-to-dividend ratios and stock yields, computed assuming that investors do not learn. This suggests that the variation in long run dividend growth expectations, assuming that investors do not learn, is minimal relative to the variation in price-to-dividend ratios, so the latter dominates the variation in stock yields. However, assuming that investors have to learn, we find significant differences between the time series of price-to-dividend ratios and stock yields.

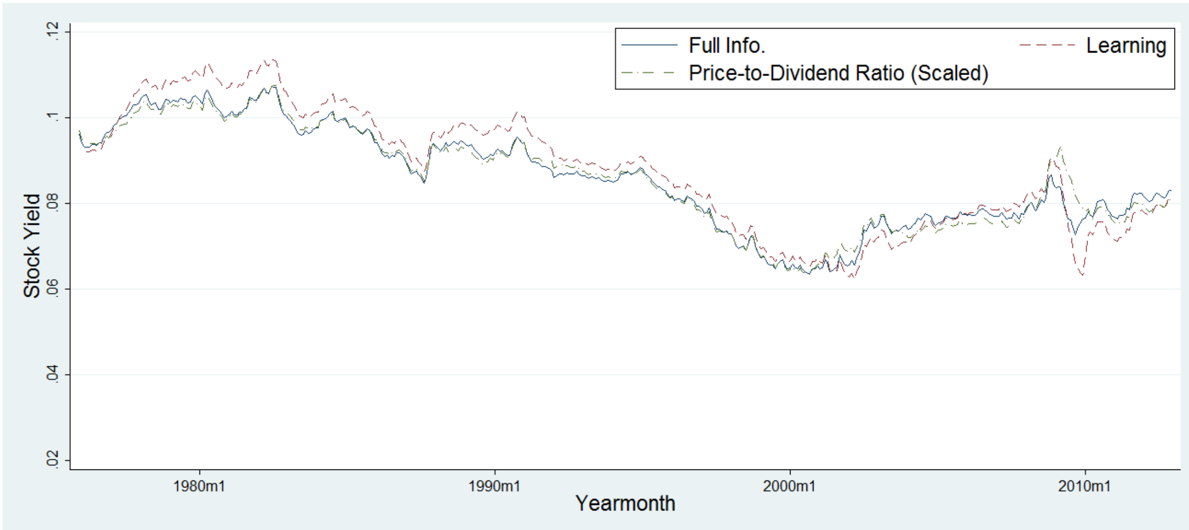


Figure 4: **Stock Yields**. This figure plots stock yields sy_t, computed using our dividend model, and log price-to-dividend ratios (scaled) for the period between 1976 and 2013. Dividend model parameters are estimated based on data since 1946. Under full information, model parameters are estimated once based on the full data sample. Under learning, those parameters are estimated at each point in time based on data available at the time.

3 Learning about Dividend Dynamics and Investor' Revealed Preferences

Results in the previous section show that parameters in our dividend model can be difficult to estimate with precision in finite sample. As a result, we argue that learning about model parameters can have significant asset pricing implications. This claim is based on the assumption that our model captures investors' expectations about future dividends. That is, we assume that investors behave as if they learn about dividend dynamics using our model. In this section, we present evidence that supports this assumption. We show that stock yields, computed assuming that investors learn about dividend dynamics using our model (see (19)), predict future stock index returns. To establish a baseline, note that, if we assume dividend growth rates follow a white noise process centered around μ_d, stock yield can be simplified to:

$$sy_t = \kappa_0 - (1 - \kappa_1)(p_t - d_t) + \mu_d. \tag{20}$$

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 9. Standard errors reported are Newey and West (1987) adjusted.[11] Results from Table 9 show that, between 1976 and 2013, price-to-dividend ratios explain 10.2 percent of the variation in stock index returns over the next year.

We then regress stock index returns over the next year on stock yields in (19), computed assuming that investors estimate model parameters at each point in time based on data available at the time. We report regression statistics in the second column of Table 9. The R-square value from this regression is 15.2 percent. We note that the only difference between this regression and the baseline regression is the assumption on dividend dynamics. That is, we assume that investors learn about dividend dynamics using our model in this regression, whereas in the baseline regression we assume that expected dividend growth rates are constant. This means that we can attribute the increase in the R-square value from 10.2 percent to 15.2 percent to our modeling of learning about dividend dynamics. We also run a bivariate regression of stock index returns over the next year on both price-to-dividend ratios and stock yields, computed assuming that investors learn about dividend dynamics using our model, and report regression statistics in the third column of Table 9. Results show that stock yields, computed assuming that investors learn about dividend dynamics using our model, 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 (19), computed assuming that investors do not learn. That is, instead of estimating model parameters at each point in time based on data available at the time, we estimate those parameters once based on the full data sample. So at every point in time, the same parameter estimates are used to compute stock yields. We report regression statistics in the fourth column of Table 9. 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 4, which show that there is almost no noticeable difference between the time series of price-to-dividend ratios and stock yields, computed using our model but assuming that investors do not learn.

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

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

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

Table 9: **Stock Index Returns and Stock Yields**: This table reports the coefficient estimates and R-square value from regressing stock index returns over the next year on log price-to-dividend ratios and stock yields, computed using our dividend model (i.e. J&L), the dividend model in van Binsbergen and Koijen (2010) (i.e. vB&K), or the dividend model in Campbell and Shiller (1988b) (i.e. C&S), and assuming investors have to learn (i.e. Learning), or do not learn (i.e. Full Info.), about model parameters. Regression is based on data between 1976 and 2013. Dividend model parameters are estimated based on data since 1946. Newey and West (1987) standard errors are reported in parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.

dynamics using either of the baseline models. We report regression statistics in the fifth and sixth columns of Table 9. We find that stock yields, computed assuming that investors learn using either of the baseline models, also perform roughly as well as price-to-dividend ratios, 10.5 percent versus 10.2 percent, in explaining future dividend growth rates.

We note that stock index returns combine the risk free rate and risk premium. To investigate whether the gain in return predictability is for predicting the risk free rate or the risk premium, in the last row of Table 9, we report the R-square value for predicting stock index excess returns.[12] Results show that the gap in forecasting performance between stock yields, computed assuming that investors learn about dividend dynamics using our model, and price-to-dividend ratios is entirely for predicting the risk premium and is not for predicting the risk free rate.

As defined in (19), stock yield is a measure of long run expected returns, i.e. the

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

weighted average discount rates of all future cash flows. Further, (19) is purely a mathematical identify, so computing stock yield does not require the econometrician to make assumptions on the asset pricing model, i.e. the stochastic discount factor, used by investors. In other words, the econometrician does not need to assume whether investors have constant relative risk aversion (CRRA) preferences or Epstein and Zin (1989) preferences in computing stock yield. However, to derive the stock index's expected return next period from stock yield, the econometrician also needs to fix the term structure of expected returns, and so more information or assumption is needed. More specifically, different asset pricing models have different implications on the term structure of expected returns, and thus implies different expected returns next period conditioning on stock yield.

For the rest of this section, we first provide the theoretical foundation that how well stock index returns can be predicted using an asset pricing model's implied expected returns can be used to evaluate the asset pricing model. Then, we incorporate learning about dividend dynamics into a long run risks model and show that as much as 22.3 percent of the variation in annual stock index returns can be predicted using our long run risks model.

3.1 Return Predictability and Assessing Asset Pricing Models

The criterion we propose to assess an asset pricing model is the deviation of that candidate model's implied expected returns from the expected returns implied by the true model. The true model here is defined as the asset pricing model that best describes the behavior of the marginal investor who prices that asset in a frictionless and efficient market. We focus on the stock market index as the asset in question, proxied by the S&P 500 stock index. Let $\mathcal{M}_i$ be a candidate model, $\mathcal{M}_0$ be the unobserved true asset pricing model, r_t be log return of the stock index, $E_t[r_{t+1}|\mathcal{M}_i]$ be the $\mathcal{M}_i$-endowed-investor's expected next period stock return, and $E_t[r_{t+1}|\mathcal{M}_0]$ be expected return under the true model. The following definition defines the best asset pricing model, i.e. the candidate model that is closet to the true model) as the model that minimizes the mean squared difference between its model implied expected returns and the expected returns implied by the true model.

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

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

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

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

$$1 - \frac{E\left[(r_{t+1} - E_t[r_{t+1}|\mathcal{M}_i])^2\right]}{E\left[(r_{t+1} - \hat{\mu}_r(t))^2\right]} > 1 - \frac{E\left[(r_{t+1} - \mathbb{E}_t[r_{t+1}|\mathcal{M}_j])^2\right]}{E\left[(r_{t+1} - \hat{\mu}_r(t))^2\right]}$$

where $\hat{\mu}_r(t) = \frac{1}{t}\sum_{s=0}^{t-1} r_{s+1}$ is the average of stock index returns up to time-t. We leave all proofs to Appendix A.4. In other words, we can denote:

$$R^2(\mathcal{M}_i) = 1 - \frac{\frac{1}{T-T_0+1}\sum_{t=T_0}^{T-1}(r_{t+1} - E_t[r_{t+1}|\mathcal{M}_i])^2}{\frac{1}{T-T_0+1}\sum_{t=T_0}^{T-1}(r_{t+1} - \hat{\mu}_r(t))^2} \tag{21}$$

as the R-square value for how well can an investor endowed with a candidate model $\mathcal{M}_i$ predict next period stock index returns. Based on Proposition 1, we employ it as an estimate of how close the candidate model $\mathcal{M}_i$ is to the true model.

3.2 Return Predictability and Learning in a Dynamic Equilibrium Asset Pricing Model

To derive the term structure of expected returns, we need to build a dynamic equilibrium asset pricing model and that requires us to specify investors' preferences and consumption dynamics. We merge dividend model with Epstein and Zin (1989) investors' preferences and persistent dividend growth rates as modeled in Bansal and Yaron (2004).

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

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

where $\tilde{C}_t$ is real consumption, ψ is the elasticity of intertemporal substitution (EIS), and α is the coefficient of risk aversion. We note that, the representative agent prefers early resolution of uncertainty if $\zeta < 0$ and prefers late resolution of uncertainty if $\zeta > 0$.[13] Log of the intertemporal marginal rate of substitution (IMRS) is:

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

where $\tilde{s}_{t+1}$ denotes real return of the representative agent's wealth portfolio. We assume that the expected growth rates in real consumption follow an AR[1] process and allow volatility in consumption growth rates to be time varying. In other words, we describe real consumption growth rates using the following system of equations:

$$\begin{aligned} \Delta \tilde{c}_{t+1} - \mu_c &= \frac{1}{\gamma(1-\phi)} x_t + \sigma_t \epsilon_{c,t+1} \\ \sigma_{t+1}^2 - \sigma_c^2 &= \varrho \left(\sigma_t^2 - \sigma_c^2 \right) + \sigma_\varsigma \epsilon_{\varsigma,t+1}. \end{aligned} \tag{24}$$

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

$$\begin{pmatrix} \epsilon_{c,t+1} \\ \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{\varsigma,t+1} \\ \epsilon_{q,t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N} \left(\mathbf{0}, \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \lambda_{dx} & 0 & \lambda_{dq} \\ 0 & \lambda_{dx} & 1 & 0 & \lambda_{xq} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & \lambda_{dq} & \lambda_{xq} & 0 & 1 \end{pmatrix} \right). \tag{25}$$

Because we do not use actual consumption data in this paper, the correlations that involve shocks $\epsilon_{c,t}$ or $\epsilon_{\varsigma,t}$ to the 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. We assume that the latent variable x_t in real consumption growth rates is the same as the latent variable in dividend growth rates. We recall that dividend growth rates in our model have the functional form given in (8). So γ is the dividend leverage parameter. We set it to 5. The primary effect of this parameter is on the unconditional mean of the equity premium. In

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

Bansal and Yaron (2004), the persistence ρ of the latent variable x_t is set to 0.975 at the monthly frequency. A common criticism of the long-run risk model has always been that it requires a small but highly persistent component in consumption and dividend growth rates that is difficult to find support in the data.[14] This criticism serves as the rationale for why we expect learning to be important. To calibrate the dynamics of consumption volatility, we follow Bansal and Yaron (2004), who set σ_c to 0.0078, ϱ to 0.987, and σ_ς to $0.23 \cdot 10^{-5}$ at the monthly frequency. We convert these to their annual equivalents. We note that our long run risk model differs from the setup in Bansal and Yaron (2004) in that we shut down heteroskedastic volatility in the dividend process. We do this in order to incorporate our dividend model, which is estimated under homoskedasticity.

We solve this specification of our long-run risk model in Appendix 1.6. In solving this model, we closely follow the steps in Bansal and Yaron (2004). The model consists of three state variables: 1) the latent variable x_t, 2) the latent variable σ_t^2, and 3) earnings-to-dividend ratios. We can solve for price-to-dividend ratio in this model as a linear function of the three state variables:

$$p_t - d_t = A_{d,0} + A_{d,1} x_t + A_{d,2} \sigma_t^2 + A_{d,3} \left(q_t - \mu_q\right) + A_{d,4}(\Delta \pi_t - \mu_\pi). \tag{26}$$

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

$$E_t[r_{t+1}] = A_{r,0} + A_{r,1} x_t + A_{r,2} \sigma_t^2 + A_{r,4}(\Delta \pi_t - \mu_\pi), \tag{27}$$

where coefficients $A_{d,\cdot}$ and $A_{r,\cdot}$, derived in Appendix A.5, are functions of the parameters governing investors' preferences, consumption dynamics, and dividend dynamics. We note that, substituting (26) into (27), we can avoid estimating time varying consumption volatility directly and instead write expected return over the short horizon as a function state variables that can be estimated from dividend dynamics and price-to-dividend ratio:

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

$$A_0 = \frac{A_{r,0} A_{d,2} - A_{r,2} A_{d,0}}{A_{d,2}}, \quad A_1 = \frac{A_{r,1} A_{d,2} - A_{r,2} A_{d,1}}{A_{d,2}} \tag{28}$$

$$A_2 = \frac{A_{r,2}}{A_{d,2}}, \quad A_3 = -\frac{A_{r,2} A_{d,3}}{A_{d,2}}, \quad A_4 = \frac{A_{r,4} A_{d,2} - A_{r,2} A_{d,4}}{A_{d,2}}. \tag{29}$$

In Figure 5, we plot the expected returns in (28), computed either assuming investors learn or that they have full information, along with realized returns over the next year, for the period between 1976 and 2013.

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

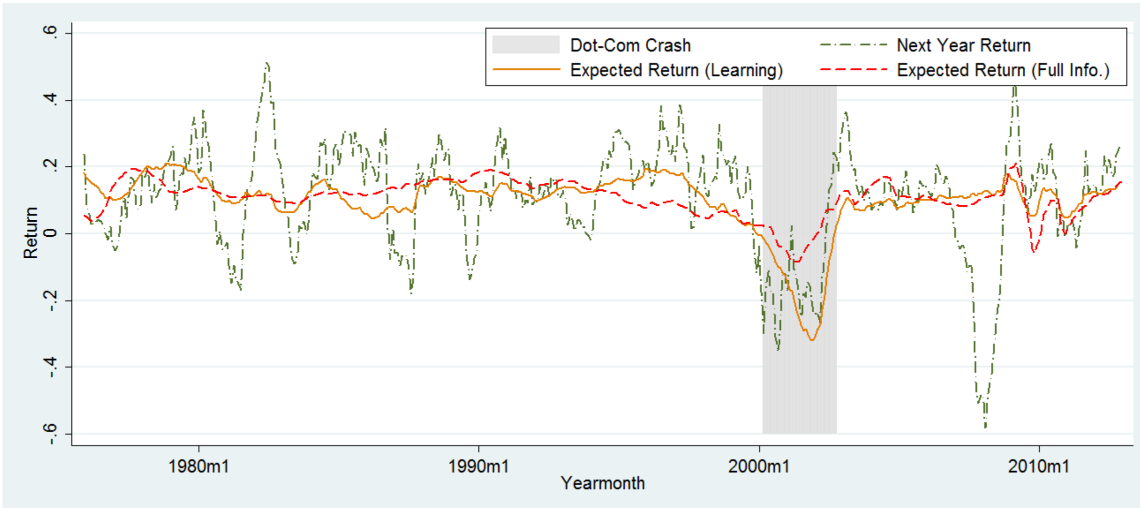


Figure 5: **Returns and Model Implied Expected Returns.** This figure plots expected returns derived from our long run risks model, as well as the actual stock index returns over the next year, for the period between 1976 and 2013. Dividend model parameters are estimated based on data since 1946.

Using expected returns in (28), we examine how our long run risks model, assuming investors learn about dividend model parameters, performs in predicting future stock index returns. We measure forecasting performance using the out-of-sample R-square value in (21). We use the first 30 years of the data sample as the training period and compute the out-of-sample R-square value using data between 1976 and 2013. In Table 10, we report the quasi out-of-sample R-square value for predicting annual stock index returns using expected returns, computed assuming investors learn about dividend dynamics using our long-run risk model. That is, we estimate dividend model parameters at each point in time based on data available at the time and substitute these parameters into (28) to compute model implied expected returns. We find that, between 1976 and 2013, our long-run risk model predicts as much as 22.3 percent of the variation in annual stock index returns.

To isolate the incremental contribution of learning to the reported R-square value in predicting stock index returns, we recalculate $A_{r,\cdot}$ and $A_{d,\cdot}$ in (28) using dividend model parameters estimated using the full data sample, and compute long run risks model implied expected returns, assuming full information , i.e. $E_t[r_{t+1}|\text{Full Info.}]$. We then report the out-of-sample R-square value for predicting stock index returns using our long run risks model, assuming full information, in the second column of Table 10. We see that, under full information, the R-square value reduces from 22.3 percent to 11.8 percent, i.e. learning acounts for about half of the return predictability documented.

To examine the statistical significance of this difference, we report, in the third column of Table 10, the incremental R-square value defined as:

$$(\mathrm{I})R^2 = 1 - \frac{\frac{1}{T-T_0+1}\sum_{t=T_0}^{T-1}\left(r_{t+1} - E_t[r_{t+1}]|\text{Learning}]\right)^2}{\frac{1}{T-T_0+1}\sum_{t=T_0}^{T-1}\left(r_{t+1} - E_t[r_{t+1}|\text{Full Info.}]\right)^2}. \tag{30}$$

Results from Table 10 show that the incremental gain in return predictability that is attributable to learning is statistically significant at the 95 percent confidence level.

Learning		Full Info.		Difference	
R^2	p-value	R^2	p-value	R^2	p-value
0.223	0.003	0.118	0.037	0.101	0.033

Table 10: **Predictive Performance under Learning and Full Information**. This table reports the out-of-sample incremental R-squared, defined in (30), for predicting stock index returns over the next year between using our long run risks model implied expected returns, assuming investors either learn about dividend model parameters or have full information, and the corresponding p-value from the adjusted-MSPE statistic of Clark and West (2007). Statistis are based on ata between 1976 and 2013.

To provide some details on our long-run risk model's forecasting performance, we follow Goyal and Welch (2008) and define the cumulative sum of squared errors difference (SSED) for between predicting stock index return using long run risks implied expected returns, assuming learning, and using historical mean returns as:

$$\text{SSED}_t = \sum_{s=T_0}^{t-1}\left(r_{s+1} - E_t[r_{s+1}|\text{Learning}]\right)^2 - \sum_{s=T_0}^{t-1}\left(r_{s+1} - \hat{\mu}_r(t)\right)^2. \tag{31}$$

The SSED is ploted in Figure 6. If the forecasting performance of our long-run risk model is stable and robust, we should observe a steady but constant decline in SSED. Instead, if the forecasting performance is especially poor in certain sub-period of the data, we should see a significant drawback in SSED during that sub-period. Figure 6 shows that our model's forecasting performance is consistent through the majority of the data sample. In detail, about one third of the forecasting performance is realized during the first two decades of the data sample, about two-third is realized during the few years surrounding the dot-com crash, and relatively flat over the last decade.

Further, to isolate the contribution of learning about dividend dynamics to SSED, we plot, in the left subfigure of Figure 6, the incremental SSED predicting stock index return using long run risks implied expected returns, assuming learning, and using those

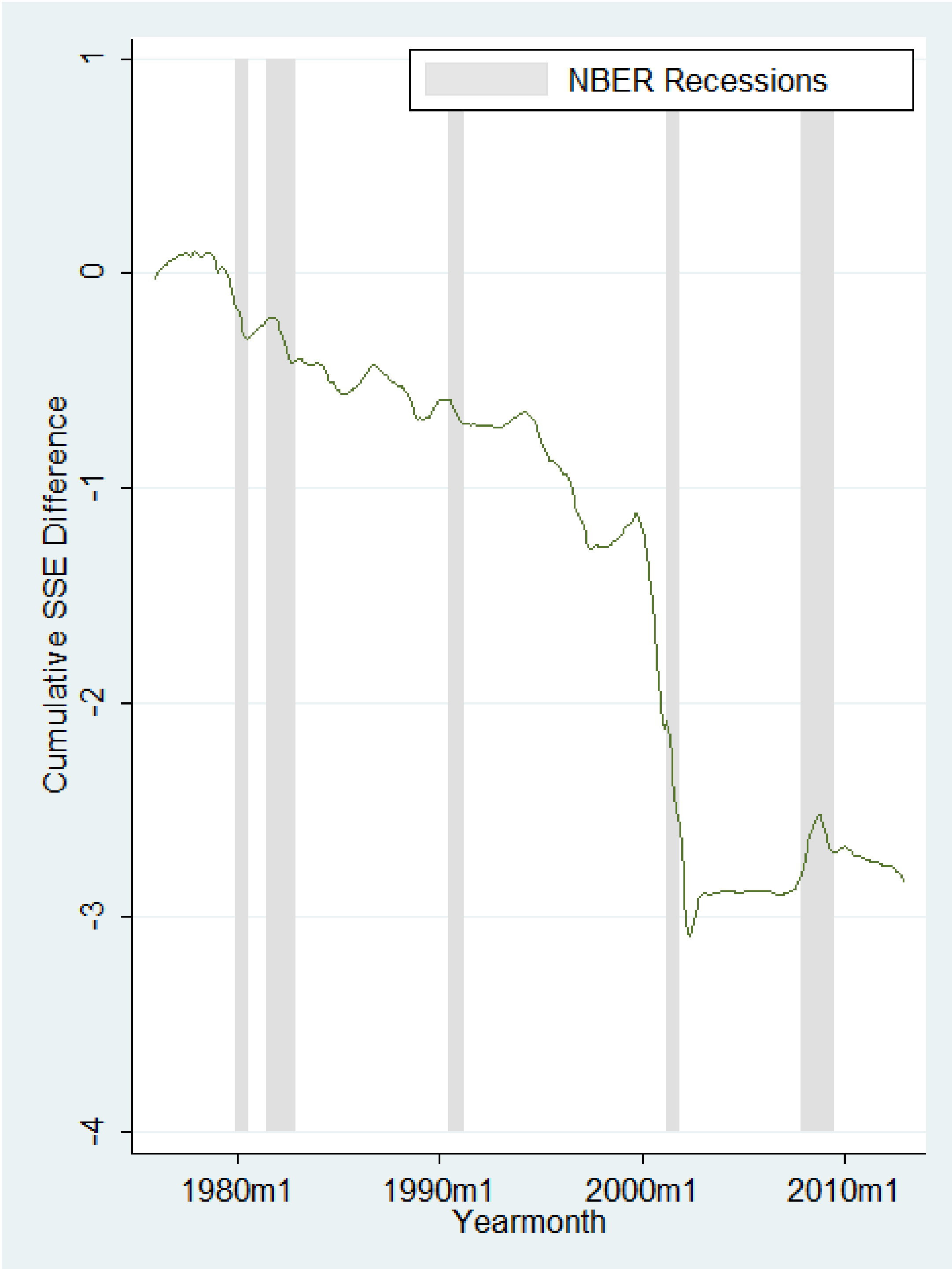


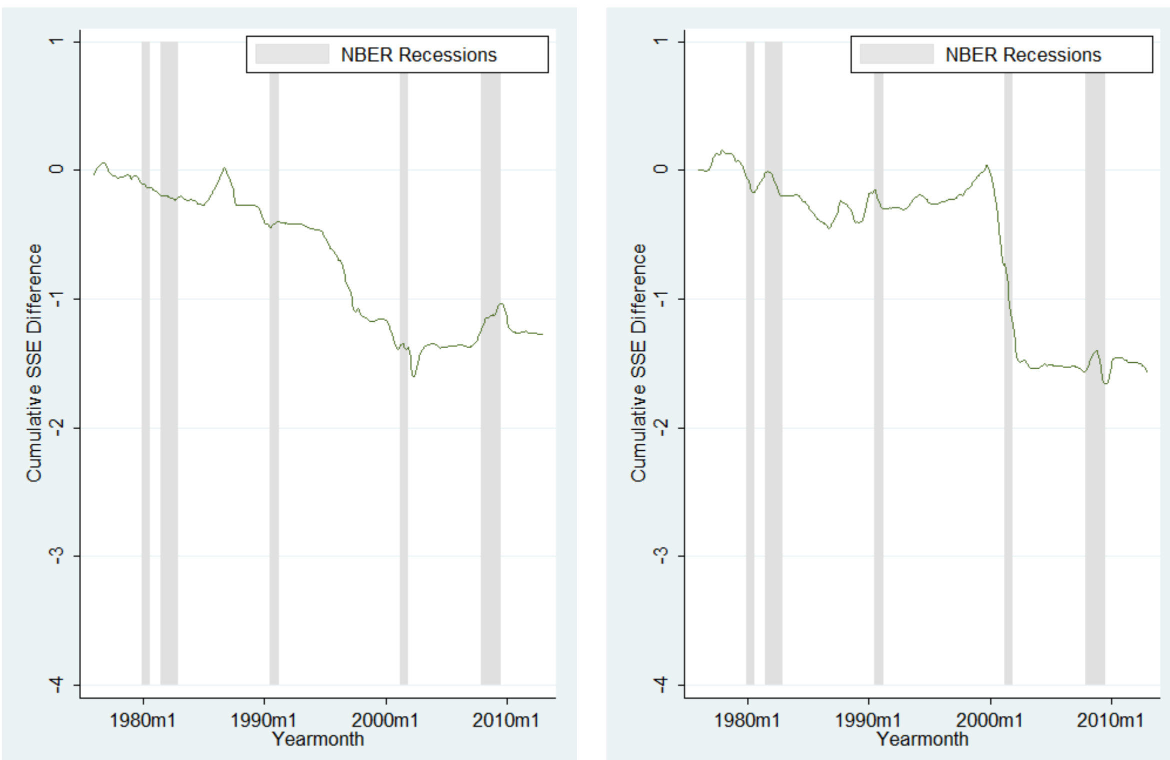
Figure 6: **Stock Index Returns and Model Implied Expected Returns (Cumulative SSE Difference).** This figure plots the cumulative sum of squared errors difference for the period between 1976 and 2013. Dividend model parameters are estimated based on data since 1946.

assuming full information:

$$(\mathrm{I})\mathrm{SSED}_t = \sum_{s=T_0}^{t-1} \left(r_{s+1} - E_t[r_{s+1}|\text{Learning}]\right)^2 - \sum_{s=T_0}^{t-1} \left(r_{s+1} - E_t[r_{s+1}|\text{Full Info.}]\right)^2 . \tag{32}$$

We note that the incremental gain in return predictability due to learning is consistent throught most of the data sample, except for the last decade, when the contribution of learning is flat . Further, there is no noticable difference in the relative importance of learning during the dot-com crash compared to other periods in the data sample.

In the right subfigure of Figure 7, we plot the difference that makes up the rest of SSED, i.e. $\mathrm{SSED}_t - (\mathrm{I})\mathrm{SSED}_t$. These results show that the predictive performance of our long run risks model, assuming investors do not learn about dividends, is almost exclusively realized during the few years surrounding the crash of the dot com bubble.



Learning − Full Info. Full Info.

Figure 7: **Decomposition of Cumulative SSE Difference**. This figure plots the cumulative sum of squared errors difference (SSED), decomposed into (I)SSED and SSED − (I)SSED, for the period between 1976 and 2013. Dividend model parameters are estimated based on data since 1946.

3.2.1 The Dot-Com Crash

Figures 5 and 6 suggest that the dot-com crash plays an especially important role in the return predictability results. To emphasize that the predictive performance of our long run risks model, assuming learn, and the incremental contribution of learning to return predictability, is not exclusively restricted to the few years surrounding the dot com crash, we recalculate the R-squared value and incremental R-squared value reported in Table 10, but for the dot com crash alone and for the subsample excluding the dot com crash. These results are reported in Table 11. We find that the predictive performance of our long run risks model, assuming learning, is statistically significant excluding the dot com crash, so is the incremental contribution of learning to this predictive performance.

	Full Info.		Learning		Learning − Full Info.	
	R^2	p-value	R^2	p-value	R^2	p-value
10/1999 - 09/2002	0.640	0.364	0.714	0.298	0.198	0.585
Other Periods	-0.024	1.000	0.097	0.070	0.121	0.058

Table 11: **Predictive Performance under Learning and Full Information**. This table reports the out-of-sample incremental R-squared, defined in (30), for predicting stock index returns over the next year between using our long run risks model implied expected returns, assuming investors either learn about dividend model parameters or have full information, and the corresponding p-value from the adjusted-MSPE statistic of Clark and West (2007). Statistis are based on two subsamples of the data: 1) the dot com crash between March 2000 and October 2002 and 2) between 1976 and 2013, excluding the dot com crash.

3.2.2 An Alternative Specification of the Long Run Risks Model

To show that our findings are not restricted to one specification of the long run risks model, we consider predicting stock index returns using an alternative model specification in this subsection. In this alternative specification, we assume consumption volatility to be constant over time. We note that in our first specification, time varying volatility in real consumption growth rates serves as an additional state variable, so to make up for the lost state variable, we assume the correlation between shocks $\epsilon_{c,t+1}$ to real consumption growth rates and shocks $\epsilon_{q,t+1}$ to earnings-to-dividend ratios to be time varying.[15] In other words, this specification allows for the conditional covariance of real consumption and dividend growth rates to vary over time. We note that this conditional covariance

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

can be negative in some states of the world, i.e. dividends can be hedges against shocks to consumption. This gives us our second specification of consumption dynamics, which can be described by the following system of equations:

$$\begin{aligned} \Delta\tilde{c}_{t+1} - \mu_c &= \frac{1}{\gamma(1-\phi)} x_t + \sigma_c \epsilon_{c,t+1} \\ \lambda_{t+1} - \mu_\lambda &= \varrho\left(\lambda_t - \mu_\lambda\right) + \sigma_\lambda \epsilon_{\lambda,t+1}, \end{aligned} \tag{33}$$

where λ_t is the correlation between $\epsilon_{c,t+1}$ in (33) and $\epsilon_{q,t+1}$ in (6) and serves as the additional state variable in this specification of consumption dynamics. Thus, the correlation matrix for shocks to dividend and real consumption dynamics can be written as:

$$\begin{pmatrix} \epsilon_{c,t+1} \\ \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{\lambda,t+1} \\ \epsilon_{q,t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N}\left(\mathbf{0}, \begin{pmatrix} 1 & 0 & 0 & 0 & \lambda_t \\ 0 & 1 & \lambda_{dx} & 0 & \lambda_{dq} \\ 0 & \lambda_{dx} & 1 & 0 & \lambda_{xq} \\ 0 & 0 & 0 & 1 & 0 \\ \lambda_t & \lambda_{dq} & \lambda_{xq} & 0 & 1 \end{pmatrix}\right). \tag{34}$$

We set the consumption dynamics parameters σ_c and ϱ to be the same as those in the first specification. We assume the unconditional mean μ_λ of the latent variable λ_t to be 0 and the standard deviation of shocks to λ_t to be 0.033 at the monthly frequency.[16] Because this specification of consumption dynamics has not been adopted in the existing literature, our choice of σ_λ can appear arbitrary. However, we note that our results are not sensitive to setting σ_λ to 0.033.[17] We solve this specification of our long run risk model in the Appendix. In solving this model, we closely follow the steps in Bansal and Yaron (2004). The model consists of three state variables: 1) the latent variable x_t, 2) the latent variable λ_t, and 3) earnings-to-dividend ratios. We can solve for price-to-dividend ratio in this model as a linear function of the three state variables:

$$p_t - d_t = A_{d,0} + A_{d,1} x_t + A_{d,2} \lambda_t + A_{d,3}\left(q_t - \mu_q\right) + A_{d,4}\left(\Delta\pi_t - \mu_\pi\right). \tag{35}$$

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

$$E_t[r_{t+1}] = A_{r,0} + A_{r,1} x_t + A_{r,2} \lambda_t + A_{r,4}\left(\Delta\pi_t - \mu_\pi\right), \tag{36}$$

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

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

where coefficients $A_{d,\cdot}$ and $A_{r,\cdot}$, derived in the Appendix, are functions of the parameters governing investors' preferences, consumption dynamics, and dividend dynamics. We note that, substituting (35) into (36), we can avoid estimating time varying correlation between shocks to real consumption growth rates and shocks to earnings-to-dividend ratios directly and instead write expected return over the short horizon as a function state variables that can be estimated from dividend dynamics and price-to-dividend ratio:

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

$$A_0 = \frac{A_{r,0}A_{d,2} - A_{r,2}A_{d,0}}{A_{d,2}}, \quad A_1 = \frac{A_{r,1}A_{d,2} - A_{r,2}A_{d,1}}{A_{d,2}}, \tag{37}$$

$$A_2 = \frac{A_{r,2}}{A_{d,2}}, \quad A_3 = -\frac{A_{r,2}A_{d,3}}{A_{d,2}}, \quad A_4 = \frac{A_{r,4}A_{d,2} - A_{r,2}A_{d,4}}{A_{d,2}}. \tag{38}$$

In Table 12, we report the out-of-sample R-square value for predicting stock index returns using this alternative long run risks model. We find very similar results. Between 1976 and 2013, this alternative long run risks model, assuming learning, predicts as much as 21.6 percent of the variation in annual stock index returns. Learning accounts for over half of this 12 percent. In Appendix A.1, we provide further discussions regarding the differences between the two long run risks model specifications.

Learning		Full Info.		Difference	
R^2	p-value	R^2	p-value	R^2	p-value
0.226	0.003	0.118	0.037	0.102	0.033

Table 12: **Predictive Performance under Learning and Full Information (Alternative Specification).** This table reports the out-of-sample incremental R-squared, defined in (30), for predicting stock index returns over the next year between using our long run risks model implied expected returns, assuming investors either learn about dividend model parameters or have full information, and the corresponding p-value from the adjusted-MSPE statistic of Clark and West (2007). Statistis are based on ata between 1976 and 2013.

4 Conclusion

In this paper, we develop a time series model for stock index dividend growth rates that combines the insights from van Binsbergen and Koijen (2010) and Campbell and Shiller (1988b). We show that the model performs well in capturing the variation in dividend growth rates. We find that some parameters in our dividend model are difficult to estimate with precision in finite sample. As a consequence, learning about dividend model parameters significantly changes investors beliefs about future dividends and the nature of the long run risks in the economy.

We show how to evaluate the economic and statistical significance of learning about parameters in the dividend process in determining asset prices and returns. We argue that a better asset pricing model should forecast returns better. We find that a long run risks model that incorporates learning about dividend dynamics is surprisingly successful in forecasting stock index returns. While the long run risks model, assuming learning, explains 22.9 percent of the variation in annual stock index returns, shutting down learning reduces the R-square value to 10.4 percent. This drop in R-square value is statistically significant.

These results also highlight the importance of investors aversion to long run risks for understanding asset prices.

References

Ang, A., and G. Bekaert (2007): "Stock Return Predictability: Is It There?," *Review of Financial Studies*, 20, 651–707.

Baker, M., and J. Wurgler (2000): "The Equity Share in New Issues and Aggregate Stock Returns," *Journal of Finance*, 55, 2219–2257.

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

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

Breen, W., L. R. Glosten, and R. Jagannathan (1989): "Economic significance of predictable variations in stock index returns," *The Journal of Finance*, 44(5), 1177–1189.

Campbell, J. Y., and R. J. Shiller (1988a): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of financial studies*, 1, 195–228.

Campbell, J. Y., and R. J. Shiller (1988b): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195–227.

Campbell, J. Y., and T. Vuolteenaho (2004): "Bad Beta, Good Beta," *Journal of Finance*, 55, 2219–2257.

Chen, L., Z. Da, and R. Priestley (2012): "Dividend Smoothing and Predictability," *Management Science*, 58, 1834–1853.

Clark, T. E., and K. D. West (2007): "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics*, 138, 291–311.

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

Croce, M. M., M. Lettau, and S. C. Ludvigson (2014): "Investor Information, Long-Run Risk, and the Term Structure of Equity," *Working Paper*.

Da, Z., R. Jagannathan, and J. Shen (2014): "Growth Expectations, Dividend Yields, and Future Stock Returns," *Working Paper*.

Epstein, L. G., and S. E. Zin (1989): "Substitution, Risk Aversion, and the Intertemporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

Fama, E. F., and K. R. French (1988): "Dividend Yield and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–25.

Fama, E. F., and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks," *The journal of finance*, 48(5), 1779–1801.

Goyal, A., and I. Welch (2008): "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455–1508.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press.

Hansen, L. P. (2007): "Beliefs, Doubts and Learning. Valuing Macroeconomic Risk," *American Economic Review*, 97.

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

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

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

Kreps, D. M. (1998): "Anticipated Utility and Dynamic Choice," *Econometric Society Monographs*, 29, 242–274.

Lamont, O. (1998): "Earnings and Expected Returns," *Journal of Finance*, 53, 1563–1587.

Lettau, M., and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56, 815–849.

Lettau, M., and S. C. Ludvigson (2005): "Expected Returns and Expected Dividend Growth," *Journal of Financial Economics*, 76, 583–626.

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

Lewellen, J., and J. Shanken (2002): "Learning, Asset-Pricing Tests, and Market Efficiency," *The Journal of Finance*, 57, 1113–1145.

Li, Y., D. T. Ng, and B. Swaminathan (2013): "Predicting Market Returns Using Aggregate Implied Cost of Capital," *Journal of Financial Economics*, 110, 419–436.

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

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

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

Stambaugh, R. F. (1999): "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421.

Timmermann, A. G. (1993): "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices," *The Quarterly Journal of Economics*, pp. 1135–1145.

van Binsbergen, J., W. Hueskes, R. Koijen, and E. Vrugt (2013): "Equity Yields," *Journal of Financial Economics*, 110, 503–519.

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

A Appendix

A.1 Additional Evidence of Robustness

A.1.1 Timing of Investors Receiving Earnings Information

Throughout this paper, we assume that investors receive earnings information 3 months after fiscal quarter or year end. To show that our findings are robust to this assumption, we repeat results in this paper, assuming that investors instead receive earnings information 6, 9, or 12 months after fiscal quarter or year end. We report these results in Tables 13. We note that changing this assumption can affect our results through its effect on long run dividend growth expectations and investors' beliefs about the persistence ω of dividend growth rates, both computed using dividend model parameters estimated at each point in time based on data available at the time. Nevertheless, results show that the significance of our findings that investors' learning about dividend dynamics is reflected in the returns of the stock index is robust to changes in this assumption on when investors receive earnings information.

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

Table 13: **Stock Index Returns and Stock Yields (Timing of Investors Receiving Earnings Information)**: This table reports coefficient estimates and R-square value from regressing stock index returns over the next year on stock yields, computed assuming investors learn about dividend dynamics using our dividend model. Regression is based on data between 1976 and 2013. Dividend model parameters are estimated based on data since 1946. To estimate dividend dynamics, we assume that investors receive earnings information 3, 6, 9, or 12 months after fiscal quarter or year end. Newey and West (1987) standard errors are reported in parentice. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.

A.1.2 Estimating Dividend Dynamics using a Rolling Window of Dividend Data

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

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

Table 14: **Dividend Growth Rates and Expected Growth Rates (Rolling Window)**: This table reports the out-of-sample R-square value for predicting dividend growth rates using our dividend model. Also reported in square parentice is the corresponding p-value from the adjusted-MSPE statistics of Clark and West (2007). Statistics are based on data between 1975 and 2013. Dividend model parameters are estimated based on a rolling window of past 10, 20, or 30 years of data.

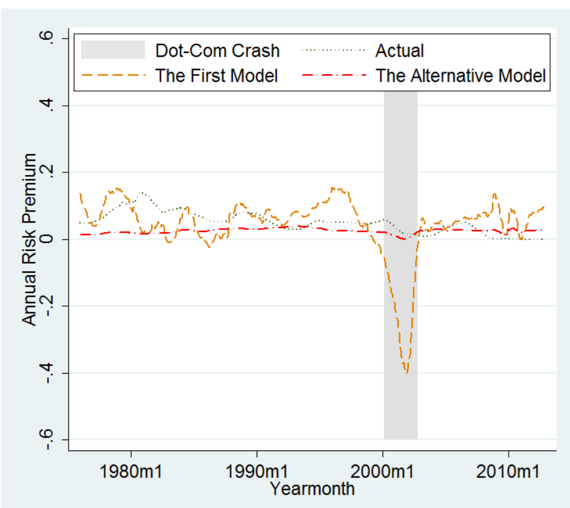
Second, we note that, according to both specifications of our long run risk model, expected returns during the Dot-Com crash are negative. We recall that, in our long run risk model, expected return is a linear function of state variables. For example, under the second specification, expected return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

$$E_t\left[r_{t+1}\right] = A_{r,0} + A_{r,1}x_t + A_{r,2}\lambda_t. \tag{39}$$

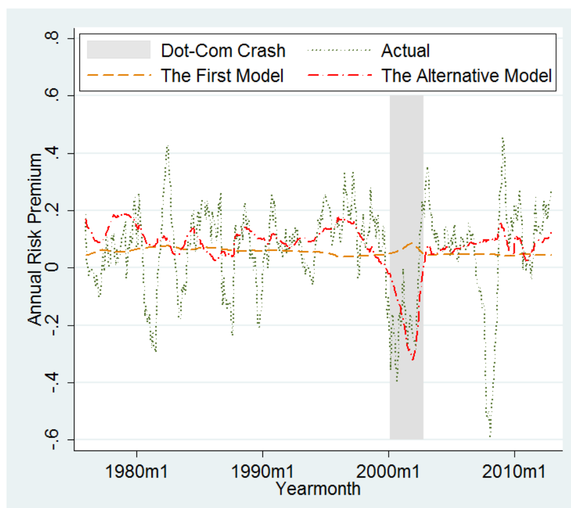
where coefficients $A_{r,\cdot}$ are always positive. So in states of the world where state variables x_t and λ_t are significantly negative, expected returns can be negative. However, our model does not shed any light on the deeper economic intuition behind why we estimate negative expected returns during the Dot-Com crash.

A.1.3 Two Long Run Risks Models

We decompose expected returns derived from each of the two specifications of our long run risk model into model implied risk free rate and model implied risk premium.[18] In Figure 8, we plot risk free rate and risk premium derived from the two specifications of our long run risk model, as well as the actual risk free rate and risk premium over the next year, for the period between 1976 and 2013. Interestingly, Figure 7 shows that the two specifications have completely different implications on the decomposition of expected returns into risk free rate and risk premium. That is, according to the first specification of consumption dynamics, almost all of the variation in expected returns is attributable to the variation in the risk free rate, whereas the risk premium hardly changes over time. To the contrary, according to the second specification of consumption dynamics, almost all of the variation in expected returns is attributable to the variation in risk premium, whereas the risk free rate hardly changes over time.



Risk Free Rate



Risk Premium

Figure 8: **Model Implied Risk Free Rate and Risk Premium.** This figure plots risk free rate and risk premium derived from the two specifications of our long run risk model, as well as the actual risk free rate and excess returns over the next year, for the period between 1975 and 2013. Dividend model parameters are estimated based on data since 1946.

Clearly, we know from the data that the risk free rate is relatively constant over time. Thus, we can infer that, of the two specifications, the second specification of consumption dynamics is the more realistic one. In other words, because different models of the

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

consumption process can have different implications for the decomposition of expected returns into risk premium and risk free rate, we can use this decomposition to shed light on consumption dynamics. However, modeling consumption is not the focus of this paper, so we leave this to potential future research.

A.2 Kalman Filter

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

$$\begin{aligned}
\Delta d_{t+1} &= \mu_d + x'_{t+1} + \phi(\Delta e_{t+1} - \mu_d) + \varphi\left(q_t - \mu_q\right) + \sigma_d \epsilon_{d,t+1} \\
x'_{t+1} &= \rho x'_t + \sigma_x \epsilon'_{x,t+1} \\
\begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon'_{x,t+1} \end{pmatrix} &\sim \text{ i.i.d. } \mathbb{N}\left(\mathbf{0}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).
\end{aligned} \tag{40}$$

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:

$$\begin{aligned}
x'_{t+1|t} &= \rho x'_{t|t}, \quad P'_{t+1|t} = \rho^2 P'_{t|t} + \sigma_x^2, \\
e_{t+1} &= \Delta d_{t+1} - \mu_d - \phi(\Delta e_{t+1} - \mu_d) - \varphi\left(q_t - \mu_q\right), \\
x'_{t+1|t+1} &= x'_{t+1|t} + \frac{P'_{t+1|t}}{P'_{t+1|t} + \sigma_d^2} e_{t+1}, \quad P'_{t+1|t+1} = P'_{t+1|t} - \frac{P^2_{t+1|t}}{P_{t+1|t} + \sigma_d^2}.
\end{aligned} \tag{41}$$

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

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

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

A.3 Bootstrap Simulation

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

$$\begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{q,t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \lambda_{dx} & \lambda_{dq} \\ \lambda_{dx} & 1 & \lambda_{xq} \\ \lambda_{dq} & \lambda_{xq} & 1 \end{pmatrix}\right). \tag{42}$$

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

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

$$q_{t+1} = \mu_q + \theta\left(q_t - \mu_q\right) + \sigma_q \epsilon_{q,t+1}. \tag{43}$$

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

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

$$\Delta e_{t+1} = q_{t+1} - q_t + \Delta d_{t+1}. \tag{44}$$

A.4 Proof for Proposition 1

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

$$\begin{aligned} & var(\mathbb{E}_t[r_{t+1}|\mathcal{M}_0] - \mathbb{E}_t[r_{t+1}|\mathcal{M}_i]) \\ = & \, var(r_{t+1} - \mathbb{E}_t[r_{t+1}|\mathcal{M}_i]) + var(\epsilon_{t+1}) - 2 \cdot cov(r_{t+1} - \mathbb{E}_t[r_{t+1}|\mathcal{M}_i], \epsilon_{t+1}) \\ = & \, var(r_{t+1} - \mathbb{E}_t[r_{t+1}|\mathcal{M}_i]) + var(\epsilon_{t+1}) + 2 \cdot cov(\mathbb{E}_t[r_{t+1}|\mathcal{M}_i], \epsilon_{t+1}) - 2 \cdot cov(r_{t+1}, \epsilon_{t+1}) \\ = & \, var(r_{t+1} - \mathbb{E}_t[r_{t+1}|\mathcal{M}_i]) + var(\epsilon_{t+1}) - 2 \cdot cov(r_{t+1}, \epsilon_{t+1}) \; . \end{aligned}$$

Last equality assumes frictionless and efficient market and investors having rational expectations. As a result, marginal investor's investment decisions are based on all information available and so ϵ_{t+1} is orthogonal to any variable that is time-t measurable. $var(\epsilon_{t+1})$

and $cov(r_{t+1}, \epsilon_{t+1})$ are independent of the candidate model i and so:

$$\begin{aligned}
&var(\mathbb{E}_t[r_{t+1}|\mathcal{M}_0] - \mathbb{E}_t[r_{t+1}|\mathcal{M}_i]) < var(\mathbb{E}_t[r_{t+1}|\mathcal{M}_0] - \mathbb{E}_t[r_{t+1}|\mathcal{M}_j]) \\
\Leftrightarrow\ & var(r_{t+1} - \mathbb{E}_t[r_{t+1}|\mathcal{M}_i]) < var(r_{t+1} - \mathbb{E}_t[r_{t+1}|\mathcal{M}_j]) \\
\Leftrightarrow\ & 1 - \frac{var(r_{t+1} - \mathbb{E}_t[r_{t+1}|\mathcal{M}_i])}{var(r_{t+1})} > 1 - \frac{var(r_{t+1} - \mathbb{E}_t[r_{t+1}|\mathcal{M}_j])}{var(r_{t+1})}\ .
\end{aligned}$$

A.5 Derivation of Price-Dividend Ratios and Expected Returns

A.5.1 The First Long Run Risks 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 (24), and investors preferences in (22). Our derivation closely follows the steps in Bansal and Yaron (2004). The log stochastic discount factor is given as:

$$m_{t+1} = \zeta \log(\delta) - \frac{\zeta}{\psi} \Delta \tilde{c}_{t+1} + (\zeta - 1)\tilde{s}_{t+1}. \tag{45}$$

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:

$$\tilde{s}_{t+1} = g_0 + g_1 z_{c,t+1} - z_{c,t} + \Delta \tilde{c}_{t+1}. \tag{46}$$

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:

$$z_{c,t} = A_{c,0} + A_{c,1} x_t + A_{c,2} \sigma_t^2. \tag{47}$$

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

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

$$var_t\left(m_{t+1}+\tilde{s}_{t+1}\right)=\zeta^2\left(1-\frac{1}{\psi}\right)^2\sigma_t^2+\zeta^2\left(g_1A_{c,1}\sigma_x\right)^2+\zeta^2\left(g_1A_{c,2}\sigma_\varsigma\right)^2. \tag{48}$$

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

$$\begin{aligned}A_{c,0}&=\frac{\log(\delta)+(1-\frac{1}{\psi})(\mu_d-\mu_\pi)+g_0+g_1A_{c,2}(1-\varrho)\sigma_c^2+\frac{1}{2}\zeta g_1^2(A_{c,1}^2\sigma_x^2+A_{c,2}^2\sigma_\varsigma^2)}{1-g_1},\\ A_{c,1}&=\frac{\left(1-\frac{1}{\psi}\right)\gamma}{1-g_1\rho},\quad A_{c,2}=\frac{\zeta(1-\frac{1}{\psi})^2}{2(1-g_1\varrho)}.\end{aligned} \tag{49}$$

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

$$\begin{aligned}r_{t+1}&=\kappa_0+\kappa_1z_{d,t+1}-z_{d,t}+\Delta d_{t+1},\\ \tilde{r}_{t+1}&=\kappa_0+\kappa_1z_{d,t+1}-z_{d,t}+\Delta\tilde{d}_{t+1}.\end{aligned} \tag{50}$$

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

$$z_{d,t}=A_{d,0}+A_{d,1}x_t+A_{d,2}\sigma_t^2+A_{d,3}(q_t-\mu_q). \tag{51}$$

Then note that:

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

$$\begin{aligned}var_t\left(m_{t+1}+\tilde{r}_{t+1}\right)=&\left(\zeta-1-\frac{\zeta}{\psi}\right)^2\sigma_t^2+\left(\frac{1}{1-\phi}\right)^2\sigma_d^2+\left((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1}\right)^2\sigma_x^2\\ &+\left((\zeta-1)g_1A_{c,2}+\kappa_1A_{d,2}\right)^2\sigma_\varsigma^2+\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)^2\sigma_q^2\\ &+2\left((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1}\right)\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{xq}\sigma_x\sigma_q\\ &+\frac{2}{1-\phi}\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{dq}\sigma_d\sigma_q+\frac{2}{1-\phi}\left((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1}\right)\lambda_{dx}\sigma_d\sigma_x.\end{aligned} \tag{52}$$

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:

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

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

$$A_{d,2}=\frac{(\zeta-1)(g_1\varrho-1)A_{c,2}+\frac{1}{2}\left(\zeta-1-\frac{\zeta}{\psi}\right)^2}{1-\kappa_1\varrho},\ A_{d,3}=\frac{\psi-(1-\theta)\phi}{(1-\kappa_1\theta)(1-\phi)}. \tag{53}$$

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

$$E_t[r_{t+1}]=A_{r,0}+A_{r,1}x_t+A_{r,2}\sigma_t^2+A_{r,3}(q_t-\mu_q), \tag{54}$$

where:

$$A_{r,0}=\kappa_0-(1-\kappa_1)A_{d,0}+\mu_d+\kappa_1(1-\varrho)A_{d,2}\sigma_c^2,\quad A_{r,1}=\frac{1}{1-\phi}-(1-\kappa_1\rho)A_{d,1},$$
$$A_{r,2}=-(1-\kappa_1\varrho)A_{d,2},\quad A_{r,3}=\frac{\varphi-(1-\theta)\phi}{1-\phi}-(1-\kappa_1\theta)A_{d,3}=0. \tag{55}$$

Expected return over the next τ period is:

$$\sum_{s=0}^{\tau-1}r_{t+s+1}=sA_{r,0}+\left(\sum_{s=1}^{\tau-1}A_{r,1}\rho^s\right)x_t+\left(\sum_{s=1}^{\tau-1}A_{r,2}\varrho^s\right)\sigma_t^2+\left(\sum_{s=1}^{\tau-1}A_{r,2}(1-\varrho^s)\right)\sigma_c^2. \tag{56}$$

Finally, the risk free rate can be written as:

$$r_{f,t+1}=A_{f,0}+A_{f,1}x_t+A_{f,2}\sigma_t^2, \tag{57}$$

where:

$$A_{f,0} = -\zeta \log(\delta) - (\zeta - 1 - \frac{\zeta}{\psi})\mu_c - (\zeta - 1)(g_0 + (g_1 - 1)C_0 + g_1(1-\varrho)C_2\sigma_c^2$$
$$- \frac{1}{2}\left((\zeta-1)g_1C_1\right)^2\sigma_x^2 - \frac{1}{2}\left((\zeta-1)g_1C_2\sigma_\varsigma\right)^2.$$
$$A_{f,1} = -(\zeta - 1 - \frac{\zeta}{\psi})\gamma + (\zeta-1)(1-g_1\rho)C_1,\ A_{f,2} = (\zeta-1)(1-g_1\varrho)C_2 - \frac{1}{2}(\zeta - 1 - \frac{\zeta}{\psi})^2. \quad (58)$$

A.5.2 The Alternative Long Run Risks 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 (33), and investors preferences in (22). Our derivation closely follows the steps in Bansal and Yaron (2004). The log stochastic discount factor is given as:

$$m_{t+1} = \zeta \log(\delta) - \frac{\zeta}{\psi}\Delta\tilde{c}_{t+1} + (\zeta-1)\tilde{s}_{t+1}. \quad (59)$$

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

$$\tilde{s}_{t+1} = g_0 + g_1 z_{c,t+1} - z_{c,t} + \Delta\tilde{c}_{t+1}. \quad (60)$$

The log-linearizing constants are:

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

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

$$z_{c,t} = A_{c,0} + A_{c,1}x_t. \quad (61)$$

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

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

$$var_t\left(m_{t+1} + \tilde{s}_{t+1}\right) = \zeta^2\left(1 - \frac{1}{\psi}\right)^2\sigma_c^2 + \zeta^2\left(g_1 A_{c,1}\sigma_x\right)^2. \quad (62)$$

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

$$
\begin{aligned}
A_{c,0} &= \frac{\log(\delta)+(1-\frac{1}{\psi})\mu_c+g_0+\frac{1}{2}\zeta(1-\frac{1}{\psi})^2\sigma_c^2+\frac{1}{2}\zeta g_1^2(A_{c,1}^2\sigma_x^2+A_{c,2}^2\sigma_\varsigma^2)}{1-g_1},\\
A_{c,1} &= \frac{\left(1-\frac{1}{\psi}\right)\gamma}{1-g_1\rho}.
\end{aligned}
\tag{63}
$$

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

$$
\begin{aligned}
r_{t+1} &= \kappa_0+\kappa_1 z_{d,t+1}-z_{d,t}+\Delta d_{t+1},\\
\tilde{r}_{t+1} &= \kappa_0+\kappa_1 z_{d,t+1}-z_{d,t}+\Delta\tilde{d}_{t+1}.
\end{aligned}
\tag{64}
$$

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

$$
z_{d,t}=A_{d,0}+A_{d,1}x_t+A_{d,2}\lambda_t+A_{d,3}(q_t-\mu_q). \tag{65}
$$

Then note that:

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

$$
\begin{aligned}
var_t\left(m_{t+1}+\tilde{r}_{t+1}\right) =& \left(\zeta-1-\frac{\zeta}{\psi}\right)^2\sigma_c^2+\left(\frac{1}{1-\phi}\right)^2\sigma_d^2+\left((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1}\right)^2\sigma_x^2\\
&+\left(\kappa_1A_2\right)^2\sigma_\varsigma^2+\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)^2\sigma_q^2+2\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\left(\zeta-1-\frac{\zeta}{\psi}\right)\sigma_x\sigma_q\lambda_t\\
&+2\left((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1}\right)\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{xq}\sigma_x\sigma_q\\
&+\frac{2}{1-\phi}\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{dq}\sigma_d\sigma_q+\frac{2}{1-\phi}\left((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1}\right)\lambda_{dx}\sigma_d\sigma_x.
\end{aligned}
\tag{66}
$$

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

$$
\begin{aligned}
A_{d,0} &= \frac{\left(\begin{array}{c}
\zeta\log(\delta)+(\zeta-1)g_0+(\zeta-1)A_{c,0}(g_1-1) \\
+\left(\zeta-\frac{\zeta}{\psi}-1\right)\mu_c+\kappa_0+\mu_c+\frac{1}{2}(\frac{1}{1-\phi})^2\sigma_d^2+\frac{1}{2}((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1})^2\sigma_x^2 \\
+\frac{1}{2}(\kappa_1A_{d,2})^2\sigma_\lambda^2+\frac{1}{2}(\kappa_1A_{d,3}+\frac{\phi}{1-\phi})^2\sigma_q^2 \\
+((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1})\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{xq}\sigma_x\sigma_q \\
+\frac{1}{1-\phi}\left(\kappa_1A_{d,3}+\frac{\phi}{1-\phi}\right)\lambda_{dq}\sigma_d\sigma_q+\frac{1}{1-\phi}\left((\zeta-1)g_1A_{c,1}+\kappa_1A_{d,1}\right)\lambda_{dx}\sigma_d\sigma_x
\end{array}\right)}{1-\kappa_1}, \\
A_{d,1} &= \frac{\left(\zeta-1-\frac{\zeta}{\psi}\right)\gamma+(\zeta-1)A_{c,1}(g_1\rho-1)+\frac{1}{1-\phi}}{1-\kappa_1\rho}, \\
A_{d,2} &= \frac{\frac{1}{1-\phi}\left(\zeta-1-\frac{\zeta}{\psi}\right)\sigma_c\sigma_d}{1-\kappa_1\varrho},\ \ A_{d,3}=\frac{\psi-(1-\theta)\phi}{(1-\kappa_1\theta)(1-\phi)}.
\end{aligned}
\tag{67}
$$

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

$$E_t[r_{t+1}]=A_{r,0}+A_{r,1}x_t+A_{r,2}\lambda_t+A_{r,3}(q_t-\mu_q), \tag{68}$$

where:

$$
\begin{aligned}
&A_{r,0}=\kappa_0-(1-\kappa_1)A_{d,0}+\mu_d, \quad A_{r,1}=\frac{1}{1-\phi}-(1-\kappa_1\rho)A_{d,1}, \\
&A_{r,2}=-(1-\kappa_1\varrho)A_{d,2}, \quad A_{r,3}=\frac{\varphi-(1-\theta)\phi}{1-\phi}-(1-\kappa_1\theta)A_{d,3}=0.
\end{aligned}
\tag{69}
$$

Expected return over the next τ period is:

$$\sum_{s=0}^{\tau-1}r_{t+s+1}=sA_{r,0}+\left(\sum_{s=1}^{\tau-1}A_{r,1}\rho^s\right)x_t+\left(\sum_{s=1}^{\tau-1}A_{r,2}\varrho^s\right)\lambda_t+\left(\sum_{s=1}^{\tau-1}A_{r,2}(1-\varrho^s)\right)\mu_\lambda. \tag{70}$$

Finally, the risk free rate can be written as:

$$r_{f,t+1}=A_{f,0}+A_{f,1}x_t, \tag{71}$$

where:

$$
\begin{aligned}
&A_{f,0}=-\zeta\log(\delta)-(\zeta-1-\frac{\zeta}{\psi})\mu_c-(\zeta-1)(g_0+(g_1-1)C_0-\frac{1}{2}\left(\zeta-1-\frac{\zeta}{\psi}\right)^2\sigma_c^2-\frac{1}{2}\left((\zeta-1)g_1C_1\right)^2\sigma_x^2. \\
&A_{f,1}=-(\zeta-1-\frac{\zeta}{\psi})\gamma+(\zeta-1)(1-g_1\rho)C_1.
\end{aligned}
\tag{72}
$$