NBER WORKING PAPER SERIES

DIVIDEND DYNAMICS, LEARNING, AND EXPECTED STOCK INDEX RETURNS

Ravi Jagannathan
Binying Liu

Dividend Dynamics, Learning, and Expected Stock Index Returns
Ravi Jagannathan and Binying Liu
NBER Working Paper No. 21557
September 2015, Revised December 2017
JEL No. G10,G11,G12

ABSTRACT

We present a latent variable model of dividends that predicts, out-of-sample, 39.5% to 41.3% of the variation in annual dividend growth rates between 1975 and 2016. Further, when learning about dividend dynamics is incorporated into a long-run risks model, the model predicts, out-of-sample, 25.3% to 27.1% of the variation in annual stock index returns over the same time horizon, and learning contributes approximately half of the predictability in returns. These findings support the view that both investors' aversion to long-run risks and their learning about these risks are important in determining the stock index prices and expected returns.

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

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

The average return on equities has been substantially higher than the average return on risk free bonds over long periods of time. Between 1946 and 2016, the S&P500 earned 66 basis points per month more than 30 days T-bills (i.e. over 7% annualized). Over the years, many dynamic equilibrium asset pricing models have been proposed in the literature to understand the nature of risks in equities that require such a large premium and why risk free rates are so low. A common feature in most of these models is that risk premium on equities does not remain constant over time, but varies in a systematic and stochastic manner. A large number of academic studies have found support for such predictable variation in equity premium.[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, these investors nevertheless have complete knowledge of the parameters describing the economy. For example, these models assume that they know the true model and model parameters governing consumption and dividend dynamics. However, as Hansen (2007) argues, "this assumption has been only a matter of analytical convenience" and is unrealistic in that it requires us to "burden the investors with some of the specification problems that challenge the econometrician". Motivated by this insight, a recent but growing literature

---

[1]See, among others, Campbell and Shiller (1988b), Breen, Glosten, and Jagannathan (1989), Fama and French (1993), Glosten, Jagannathan, and Runkle (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), Chen, Da, and Zhao (2013), Kelly and Pruitt (2013), van Binsbergen, Hueskes, Koijen, and Vrugt (2013), Li, Ng, and Swaminathan (2013), Da, Jagannathan, and Shen (2014), and Martin (2017).

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 (2016) propose a Markov-switching model for consumption dynamics and show that learning about the consumption process is reflected in asset prices. Croce, Lettau, and Ludvigson (2014) show that a bounded rationality limited information long-run risks model can generate a downward-sloping equity term structure. Collin-Dufresne, Johannes, and Lochstoer (2016) provide the theoretical foundation that parameter learning can be a source of long-run risks under Bayesian learning.[2] We add to this literature.

The main contributions of our paper are as follows. First, we present a model for aggregate dividends of the stock index, based on simple economic intuition, that explains large variation in annual dividend growth rates out-of-sample. Then, we show that, when learning about dividend dynamics is incorporated into a long-run risks model, the model predicts large variation in annual stock index returns out-of-sample. This not only addresses the Goyal and Welch (2008) critique and significantly revises upward the degree of return predictability in the existing literature, but also lends support that both investors' aversion to long-run risks and learning about these risks play important roles in determining asset prices and expected returns.[3][4]

To study how learning about dividend dynamics affects stock index prices and expected returns, we first need a dividend model that is able to realistically capture how investors form expectations about future dividends. Inspired by Lintner (1956) and Campbell and Shiller (1988b), we put forth a model of dividend growth rates that incorporates information in corporate payout policy into the latent variable model used in Cochrane

---

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

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

[4]Following the existing literature, we adopt the stock index as a proxy for the market portfolio.

(2008), van Binsbergen and Koijen (2010), and others. Our model predicts 42.4% to 46.4% of the variation in annual dividend growth rates between 1946 and 2016 in-sample and predicts 39.5% to 41.3% of the variation in annual dividend growth rates between 1976 and 2016 out-of-sample. Based on these results, we comfortably reject the null that expected dividend growth rates are constant and demonstrate that the superior performance of our dividend model over alternative models in predicting annual dividend growth rates is statistically significant and economically meaningful.

We document that uncertainties about parameters in our dividend model, especially parameters surrounding the persistent latent variable, are high and resolve slowly. That is, these uncertainties remain substantial even at the end of our 71 years data sample, suggesting that learning about dividend dynamics is a difficult and slow process. Further, when our dividend model is estimated at each point in time based on data available at the time, model parameter estimates fluctuate, some significantly, over time as more data become available. In other words, if investors estimate dividend dynamics using our model, we expect their beliefs about the parameters governing the dividend process to vary significantly over time. We then show that these changes in investors' beliefs can have large effects on their expectations of future dividends. Through this channel, changes in investors' beliefs about the parameters governing the dividend process can contribute significantly to the variation in stock prices and expected returns.

We provide evidence that investors behave as if they learn about dividend dynamics and price stocks using our model. First, we define stock yields as discount rates that equate the present value of expected future dividends to the current prices of the stock index. From the log-linearized present value relationship of Campbell and Shiller (1988), we write stock yields as a function of price-to-dividend ratios and long-run dividend growth expectations.[5] We show that, assuming that investors learn about dividend

---

[5]See Jagannathan, McGrattan, and Scherbina (2001) for the dynamic version of the Gordon (1959)

3

dynamics, these stock yields explain 18.7% of the variation in annual stock index returns between 1975 and 2016. In comparison, stock yields, assuming full information, predict a statistically significantly lower 13.0% of the same variation over the same horizon. Next, we embed our dividend model into an dynamic equilibrium asset pricing model that features Epstein and Zin (1989) preferences, which capture preferences for the early resolution of uncertainty, and consumption dynamics similar to the long-run risks model of Bansal and Yaron (2004). We refer to this model as our long-run risks model. We find that, assuming learning, our long-run risks model predicts 25.3% to 27.1% of the variation in annual stock index returns between 1975 and 2016. Learning accounts for approximately half of the predictability in returns. Both the model's forecasting performance and the incremental contribution of learning to this performance are statistically significant and economically meaningful.

Our results suggest that, aside from a common persistent component in consumption and dividend growth rates, the assumption that investors hold Epstein and Zin (1989) preferences with early resolution of uncertainty, a critical component of any long-run risks model, is essential to the model's strong performance in predicting annual stock index returns.[6] More specifically, we find that, replacing Epstein and Zin (1989) preferences with constant relative risk aversion (CRRA) preferences, R-square value for predicting annual stock index returns, between 1975 and 2016, drops from 13.3% to 11.8% assuming full information and drops from at least 25.3 percent to at most 15.1 percent after learning is incorporated. This substantial deterioration in forecasting performance is evidence that the assumption of early resolution of uncertainty, as modeled through Epstein and Zin

---

growth model that gives an expression for stock yield in levels. When expected dividend growth rates vary over time, according to the present value relationship, stock yield, that is, the long-run expected return on stocks, is the current dividend yield plus a weighted average of expected future one period dividend growth rates.

[6]Alternatively, as Hansen and Sargent (2010) and Bidder and Dew-Becker (2016) show, if investors are ambiguity averse, they may behave as if there is such a common persistent component even if the actual consumption and dividend processes are not persistent.

(1989) preferences, is potentially important for building an asset pricing model consistent with investors behavior.

We follow Cogley and Sargent (2008), Piazzesi and Schneider (2010), and Johannes, Lochstoer, and Mou (2016), 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.[7]

The rest of this paper is as follows. In Section 1, we introduce our dividend model and evaluate its performance in capturing dividend dynamics. In Section 2, we show that investors' beliefs about dividend model parameters can vary significantly over time as a result of Kreps' learning about dividend dynamics. In Sections 3, we show that learning accounts for a significant fraction of the variation in both long-run and short-run expected stock index returns. In Section 4, we first discuss how an asset pricing model's performance in predicting stock index returns can be used as a criterion to evaluate that model. Then, we demonstrate that, between 1975 and 2016, a model that incorporates Kreps' learning into a long-run risks model predicts 25.3% to 27.1% of the variation in annual stock index returns and explain why such a finding provides us insight into investor preferences and the role of learning in describing investors' behavior. In Section 5, we conclude.

I. The Dividend Model

In this section, we present a model for dividend growth rates that extends the latent variable model of Cochrane (2008), van Binsbergen and Koijen (2010), and others by

---

[7]Collin-Dufresne, Johannes, and Lochstoer (2016) provide the theoretical foundation for studying uncertainties about model parameters as priced risk factors.

incorporating information in corporate payout policy into the model. The inclusion of corporate payout policy 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.

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

$$\Delta d_{t+1} - \mu_d = x_t + \sigma_d \epsilon_{d,t+1}$$

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

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

where time-t is defined in years to control for potential seasonality in dividend payments. Following van Binsbergen and Koijen (2010), we fit our model to the nominal dividend process. As shown in Jagannathan, McGrattan, and Scherbina (2000) and Boudoukh, Michaely, Richardson, and Roberts (2007), equity issuance and repurchase tend to be more sporadic and random compared to cash dividends. For this reason, we focus on modeling the cash dividend process.[8] In (1), expected dividend growth rates are a function of the latent variable x_t, the unconditional mean μ_d of dividend growth rates, and the persistence coefficient ρ of the latent variable x_t:

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

---

[8]A firm's investment opportunities set includes repurchasing its own shares. It will, all else equal, lead to an increase in future earnings of the remaining shares, just as investment in any other productive assets would. This is another reason to focus on cash dividends in our study.

Before we introduce corporate payout policy into this model, we first recall the dividend model used in Campbell and Shiller (1988b). Define p_t as log nominal price of the stock index, e_t as log nominal earnings, 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:

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

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

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

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

Estimates of β_{10}, β_{11}, β_{12}, and β_{13} from (3), based on data between 1946 and 2016, are reported in the first row of Table I. We see that both price-to-dividend ratios and CAPE ratios have significant effects on future dividends, but in the opposite direction. That is, increases in price-to-dividend ratios predict increases in future dividend growth rates, but increases in CAPE ratios predict decreases in future dividend growth rates. Further, we note from Table I that $\beta_{12} + \beta_{13} = 0$ cannot be statistically rejected. For this reason, we restrict $\beta_{13} = -\beta_{12}$ and re-write (3) as:

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

Stock index price p_t does not appear in (5). Instead, future dividend growth rates are a function of some measure of retention ratios, that is, $\bar{e}_t - d_t$. Estimated coefficients from

(5) are in the second row of Table I. We see that the β_2 estimate is significant, suggesting that expected dividend growth rates respond to corporate payout policy. High earnings relative to dividends implies that firms have been retaining earnings in the past and so they are expected to pay more dividends in the future.

[Table I is about here.]

We extend (1) based on this insight that corporate payout policy contains 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, that is, retention ratio. We write our dividend model as the following system of equations:

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

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

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

$$\begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{q,t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N} \left(\mathbf{0}, \begin{pmatrix} 1 & \lambda_{dx} & \lambda_{dq} \\ \lambda_{dx} & 1 & \lambda_{xq} \\ \lambda_{dq} & \lambda_{xq} & 1 \end{pmatrix} \right). \tag{6}$$

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

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

This means that, aside from the latent variable x_t and retention ratios, expected dividend

8

growth rates are a function of the unconditional mean μ_d of dividend growth rates, the unconditional mean μ_q and persistence θ of retention ratios, the persistence ρ of the latent variable x_t, and coefficient ϕ that connects corporate payout policy to dividend dynamics. The earnings process is not modeled explicitly in (6). However, because earnings growth rates are, by definition, a function of dividend growth rates and retention ratios, that is,

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

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

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

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

A type of model commonly used to forecast macroeconomic variables is a Markov-switching model. We take the Markov-switching model describing consumption dynamics in Johannes, Lochstoer, and Mou (2016). It is conceivable that the same model can be applied to dividend growth rates:

$$\Delta d_{t+1} = \mu_d(s_t) + \sigma_d(s_t)\epsilon_{d,t+1}, \ s_t \in \{1,2,3\},$$

$$\mathbb{p}(s_{t+1} = i | s_t = j) = \phi_{ij},$$

$$\phi_{ij} \in [0,1] \ \forall i,j \in \{1,2,3\}, \ \sum_{i=1}^{3} \phi_{ij} = 1 \ \forall j \in \{1,2,3\}. \tag{10}$$

That is, s_t is the underlying state of the economy, $\mathbb{p}(s_{t+1} = i | s_t = j)$ is the probability that the economy transfers from state $j \in \{1,2,3\}$ to state $i \in \{1,2,3\}$, and $\mu_d(s_t)$ and $\sigma_d(s_t)$ are the mean and volatility of dividend growth rates in a particular state. A key feature of this model that is not present in dividend models discussed so far is that it

is able to incorporate, albeit in a restricted manner, both regime changes and stochastic volatility. We adopt (10) as another baseline to compare against our dividend model.

A. Estimation and Results

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 estimate model parameters:

$$\mathbf{\Theta} = \{\mu_d, \phi, \sigma_d, \rho, \sigma_x, \mu_q, \theta, \sigma_q\} \tag{11}$$

based on maximum-likelihood. For parameter reduction, we assume in our model that cross-correlations $\lambda(\cdot, \cdot)$ of different shocks to the dividends are zeros, that is, $\lambda(\epsilon_{d,t+1}, \epsilon_{q,t+1}) = 0$, $\lambda(\epsilon_{d,t+1}, \epsilon_{x,t+1}) = 0$, and $\lambda(\epsilon_{x,t+1}, \epsilon_{q,t+1}) = 0$. The log-likelihood function $l(\cdot)$ is then separable and maximizing it is equivalent to:

$$\max_{\mathbf{\Theta}} l\left(\Delta d_1, ..\Delta d_T, q_0, ...q_T | \mathbf{\Theta}\right) = \max_{\{\mu_q, \theta, \sigma_q\}} l_1 + \max_{\{\mu_d, \phi, \sigma_d, \rho, \sigma_x\}} l_2$$

$$l_1 = l\left(q_0, .., q_T | \{\mu_q, \theta, \sigma_q\}\right)$$

$$l_2 = l\left(\Delta d_1 | q_0, \{\mu_d, \phi, \sigma_d, \rho, \sigma_x\}\right) + \sum_{t=0}^{T-1} l\left(\Delta d_{t+1} | q_0, ..., q_t, \Delta d_1, .., \Delta d_t, \{\mu_d, \phi, \sigma_d, \rho, \sigma_x\}\right).$$

$$\tag{12}$$

Thus, we can separately estimate $\{\mu_q, \theta, \sigma_q\}$ from the AR[1] process of retention ratios by maximizing l_1 using least squares, and estimate $\{\mu_d, \phi, \sigma_d, \rho, \sigma_x\}$ from the rest of the dividend model by maximizing l_2 using Kalman filter (Hamilton (1994)). Appendix A describes Kalman filter. Table II reports model parameter estimates based on non-

10

overlapping annual data between 1946 and 2016.[9] Standard errors of parameter estimates are based on bootstrap simulation, as described in Appendix B. Previous works have suggested a regime shift in dividend dynamics before and after World War II. Fama and French (1988) note that dividends are more smoothed in the post-war period. Chen, Da, and Priestley (2012) argue that the lack of predictability in dividend growth rates by price-to-dividend ratios in the post-war period is attributable to this dividend smoothing behavior. So our sample is for the post-war period between 1946 and 2016. Consistent with our intuition, coefficient ϕ that connects corporate payout policy to dividend dynamics is estimated to be positive and significant. That is, high retention ratios imply high future dividend growth rates. The annual persistence of retention ratios is estimated to be 0.370. The latent variable x_t is estimated to be slightly more persistent at 0.469. So there is a moderate level of persistence in dividend growth rates between 1946 and 2016 based on estimates from our model.

[Table II is about here.]

In the first column of Table III, we report our dividend model's performance in predicting annual dividend growth rates. Between 1946 and 2016, our model predicts 46.4% of the variation in annual dividend growth rates, which is a significant improvement over the baseline models. Given these statistics are in-sample, we know that at least a part of this improved forecasting performance comes from adding more parameters to existing models and is thus mechanical. Thus, to address the concern that our model overfits the data, we also assess our model based on how it predicts annual dividend growth rates out-of-sample. That is, instead of estimating model parameters based on the full data sample, we predict dividend growth rates at each point in time using model parameters estimated based on data available at the time. Model $\mathcal{M}_i$'s out-of-sample performance is

---

[9]All annual statistics reported are based on year-end data, that is, from January to December. For all of the results in this paper, we have also replicated them using overlapping annual data, the findings are very similar.

then evaluated using out-of-sample R-square value in Goyal and Welch (2008):

$$\mathcal{R}_O^2(\mathcal{M}_i) = 1 - \frac{\sum_{t=T_0}^{T-1} \left(\Delta d_{t+1} - E_t[\Delta d_{t+1}|\mathcal{M}_i]\right)^2}{\sum_{t=T_0}^{T-1} \left(\Delta d_{t+1} - \hat{\mu}_{d,t}\right)^2}, \tag{13}$$

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

$$\hat{\mu}_{d,t} = \frac{1}{t} \sum_{s=0}^{t-1} \Delta d_{s+1}. \tag{14}$$

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. We use the data sample prior to 1975 as the training period and out-of-sample prediction is for the 42 years period between 1975 and 2016. In the second and third columns of Table III, we report out-of-sample R-square values for predicting annual dividend growth rates and the corresponding bootstrap simulated p-values. Results show that our model predicts 41.3% of the variation in annual dividend growth rates between 1975 and 2016 out-of-sample, which is economically meaningful improvement over the 16.1%, 25.6%, and -4.2% from the baseline models.

[Table III is about here.]

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

$$\mathcal{R}_I^2(\mathcal{M}_i, \mathcal{M}_j) = 1 - \frac{\sum_{t=T_0}^{T-1} \left(\Delta d_{t+1} - E_t[\Delta d_{t+1}|\mathcal{M}_i]\right)^2}{\sum_{t=T_0}^{T-1} \left(\Delta d_{t+1} - E_t[\Delta d_{t+1}|\mathcal{M}_j]\right)^2}, \tag{15}$$

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

12

B. Inflation and Real Rates

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

$$\Delta\pi_{t+1} - \mu_\pi = \eta\left(\Delta\pi_t - \mu_\pi\right) + \sigma_\pi\epsilon_{\pi,t+1}, \ \epsilon_{\pi,t+1} \sim \text{i.i.d. } \mathbb{N}(0,1). \tag{16}$$

Table IV reports parameter estimates of the inflation model based on non-overlapping annual data between 1946 and 2016. There is a moderate level of persistence in inflation rates. Based on the reported R-square value for predicting inflation rates, which is 44.8% in-sample between 1946 and 2016 and 54.0% out-of-sample between 1975 and 2016, we see that the AR[1] model does a reasonable job in describing the inflation process.

[Table IV is about here.]

For parameter reduction, we assume cross-correlations of different shocks to inflation and shocks to the dividend process, that is, $\lambda_{d\pi} = \lambda\left(\epsilon_{d,t+1}, \epsilon_{\pi,t+1}\right)$, $\lambda_{x\pi} = \lambda\left(\epsilon_{x,t+1}, \epsilon_{\pi,t+1}\right)$, and $\lambda_{q\pi} = \lambda\left(\epsilon_{q,t+1}, \epsilon_{\pi,t+1}\right)$, are zeros. This assumption also implies that estimating the inflation model separately from dividend dynamics using least squares is equivalent to a joint maximum likelihood estimation. Given this inflation model, we can then derive the expression for expected real dividend growth rates based on expected nominal rates and inflation as:

$$E_t[\Delta\tilde{d}_{t+s+1}] = (\mu_d - \mu_\pi) + \rho^s x_t + \phi\theta^s(q_t - \mu_q) - \eta^{s+1}\left(\Delta\pi_t - \mu_\pi\right), \ \forall s \geq 0. \tag{17}$$

where $\Delta\tilde{d}_t = \Delta d_t - \Delta\pi_t$ denotes real dividend growth rate.[10] To provide a more intuitive visualization of how various types of shocks to real dividend growth rates at a given time

---

[10]Throughout, we put $\sim$ on top of a variable to denote that the variable is defined in real terms.

affect investors' expectations of real dividends going forward, we consider an one unit change to shocks to the real dividend process, that is, $\epsilon_{d,t}$, $\epsilon_{x,t}$, $\epsilon_{q,t}$, and $\epsilon_{\pi,t}$, and show how such a change affects both real dividend growth rates immediately and expected real dividend growth rates up to 10 years into the future. We report these impulse response functions in Figure 1. We see that $\epsilon_{d,t}$ affects dividend growth rates instantly but its effect does not persist into the future, whereas $\epsilon_{x,t}$ and $\epsilon_{q,t}$ affect dividend growth rates with a one-period lag but their effects are persistent over time. Figure 1 shows that expected inflation 1-to-1 negatively affects expected real dividend growth rates. This is a result of our choice to fit our dividend model to the nominal dividend process and extract real dividend growth rate expectations by subtracting from it expected inflation rates. The underlying assumption behind this choice is that firms do not adjust for inflation in paying out dividends to investors, so we expect real rates to fall as inflation expectations rise. This is supported in the data and is consistent with the economic rationale that expected inflation is negatively related to expected growth in real activity.[11]

[Figure 1 is about here.]

In Table V, we report the in-sample and out-of-sample R-square values for predicting real, rather than nominal, annual dividend growth rates using either our model or one of the baseline models. We find that our model also outperforms the baseline models in forecasting real annual dividend growth rates. It predicts 42.4% of the variation in real annual dividend growth rates between 1946 and 2016 in-sample and 39.5% of the variation in real rates between 1975 and 2016 out-of-sample.

[Table V is about here.]

---

[11]See, among others, Fama (1981) and Piazzesi and Schneider (2006).

II. 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 II are estimated using data up to 2016, so they reflect investors' knowledge of dividend dynamics at the end of 2016. So if investors were to estimate our dividend model at an earlier date, they would have estimated a set of parameter values different from those reported in Table II. This is a result of investors' knowledge of dividend dynamics evolving as more data become available. We call this learning. That is, we use learning to refer to investors estimating model parameters at each point in time based on data available at the time. In this section, we summarize how learning affects investors' beliefs about the parameters governing the dividend process, assuming that investors behave as if they learn about dividend dynamics using our model. We then show evidences supporting that learning about dividend dynamics can have significant asset pricing implications.

To better reflect investors' information set, from this point onwards, we assume investors have access to earnings information 6 months after fiscal quarter or year end. The choice of 6 months is reasonably conservative and 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.[12] We report, in Figure 2, model parameters estimated based on non-overlapping annual data up to time-τ, for τ between 1975 and 2016. 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 retention ratios. This suggests that firms have been paying a smaller fraction of earnings as cash dividends in recent decades. Second, there are gradual downward drifts in investors' beliefs about ϕ that connects

---

[12]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. However, in our research we find that a small percentage of firms do miss these deadlines.

corporate payout policy to dividend dynamics. This means that dividends have become more smoothed over time. The decline in the impact of retained earnings on future dividends is consistenet with declining investment opportunities and more of the retained earnings being used for share repurchases. Third, a sharp drop in investors' beliefs about the persistence θ of retention ratios towards the end of our data sample is due to the abnormally low earnings reported around the time of the 2009 recession and the strong stock market recovery that followed. The changes in the volatility of shocks to dividends and retention ratios are products of these trends.

[Figure 2 is about here.]

Figure 2 shows that the persistence ρ of the latent variable x_t appears to be the parameter hardest to learn and least stable over time. Investors' belief about ρ fluctuates significantly over the sample period. For example, there are three times when investors' beliefs about ρ sharply drops during our sample. The first is around the time of the 2001 recession. The second is at the start of what is sometimes referred to as the Dot-Com bubble. The third is around the time of the 2009 recession. This is a standard feature of a latent variable model. That is, when a large and unexpected shock hits, in our context either in the form of a recession or what is sometimes referred to as a bubble, our model assigns some positive probability that such a shock belongs to the persistent process and revises ρ downward.

We can infer, from standard errors reported in Table II, that learning about dividend dynamics is a slow process. That is, even with 71 years of data, there are still significant uncertainties surrounding the estimates of some model parameters. To quantify the speed of learning, following Johannes, Lochstoer, and Mou (2016), for a parameter in our dividend model, we construct a measure that is one minus the inverse ratio between the bootstrap simulated standard error assuming that the parameter is estimated based on data up to 2016 and the bootstrap simulated standard error assuming that the parameter

is estimated based on 10 more years of data, that is, if the parameter were estimated using data up to 2026. In other words, this ratio reports how much an estimated parameter's standard error would have reduced in 10 years. So the closer this ratio is to zero, the more difficult it is for investors to learn about that parameter. In Table VI, we report this measure for each of the eight model parameters. Overall, 10 years of additional data would decrease the standard errors of parameter estimates by between approximately 3% and 8%. Further, consistent with results in Figure 2 and in Table I, reducing uncertainties surrounding ρ is the most difficult among these parameters.

[Table VI is about here.]

A. Parameter Uncertainty and Expectations for the Long-Run

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

$$p_t - d_t = \frac{\kappa_0}{1 - \kappa_1} + \sum_{s=0}^{\infty} \kappa_1^s \left(E_t[\Delta d_{t+s+1}] - E_t[R_{t+s+1}] \right), \tag{18}$$

where κ_0 and κ_1 are log-linearizing constants and R_{t+1} is the stock index's log return.[13] The expression is a mathematical identity that connects price-to-dividend ratios, expected dividend growth rates, and discount rates, that is, expected returns. We define stock yields as discount rates that equate the present value of expected future dividends to the current price of the stock index. That is, rearranging (18), we can write stock yields as:

$$sy_t \equiv (1 - \kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta R_{t+s+1}]$$

$$= \kappa_0 - (1 - \kappa_1)(p_t - d_t) + (1 - \kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta d_{t+s+1}]. \tag{19}$$

---

[13]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 3.474.

We define long-run dividend growth expectations as:

$$\partial_t \equiv (1 - \kappa_1) \sum_{s=0}^{\infty} \kappa_1^s E_t[\Delta d_{t+s+1}]. \tag{20}$$

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

$$\partial_t = (1 - \kappa_1) \sum_{s=0}^{\infty} \kappa_1^s \left(\mu_d + \rho^s x_t + \phi \theta^s (q_t - \mu_q) \right). \tag{21}$$

If investors instead use a different dividend model, their expectations of long-run dividend growth rates will also be different. For example, if we assume that dividend growth rates follow a white noise process centered around μ_d, we can rewrite (21) instead as $\partial_t = \mu_d$. Further, because long-run dividend growth expectations are functions of dividend model parameters, it is also affected by whether these parameters are estimated once based on the full data sample, or estimated at each point in time based on data available at the time. The first case corresponds to investors having complete knowledge of the parameters describing the dividend process. The second case corresponds to investors having to learn about dividend dynamics. In Figure 3, we plot our model's long-run dividend growth expectations, either assuming learning or assuming full information. The plot shows that learning can have a considerable effect on investors' long-run dividend growth expectations.

[Figure 3 is about here.]

In Figure 4, we plot stock yields, either assuming learning or assuming full information,

18

computed by substituting (21) into (19):

$$sy_t = \kappa_0 - (1 - \kappa_1)(p_t - d_t) + \mu_d + (1 - \kappa_1)\left(\frac{1}{1 - \kappa_1\rho}x_t + \frac{\phi}{1 - \kappa_1\theta}(q_t - \mu_q)\right). \qquad (22)$$

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

[Figure 4 is about here.]

III. Learning about Dividends and the Time Variation in Discount Rates

So far in our analysis, we have focused on how learning affects the econometrician. From this point onwards, we examine the asset pricing implications of assuming that investors have to learn about dividend dynamics in a manner similar to learning by the econometrician. While the econometrician does not price assets, investors do. So, assuming investors behave as if they learn about dividend dynamic using our model, we expect such behavior to affect stock index prices and returns. This assumption is not unreasonable. Because our dividend model outperforms alternative models in forecasting dividends, it is natural to assume that investors behave as if they use a model that is more similar to ours in their investment decisions, at least among the choices examined.

19

In this section, we present evidence that are consistent with this assumption.

First, we show that stock yields, assuming learning, predict annual 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 yields can be simplified to:

$$sy_t = \kappa_0 - (1 - \kappa_1)(p_t - d_t) + \mu_d. \tag{23}$$

That is, under the white noise assumption, stock yields are just scaled price-to-dividend ratios. So, we regress future annual stock index returns on price-to-dividend ratios, based on non-overlapping annual data between 1975 and 2016. Statistics are in the first column of Table VII. Results from Table VII show that, between 1975 and 2016, price-to-dividend ratios predict 13.6% of the variation in annual stock index returns.

We then regress future annual stock index returns on stock yields in (22), assuming learning. We report regression statistics in the second column of Table VII. We see that R-square value from this regression is 18.7%. We note that the only difference between this regression and the baseline regression is the assumption on the dividend process. That is, here we assume that investors behave as if they learn about dividend dynamics using our model, whereas in the baseline regression we assume that expected dividend growth rates are constant. This means that we can attribute the increase in R-square value from 13.6% to 18.7% to our modeling of learning about dividend dynamics. To emphasize the importance of learning, we regress future annual stock index returns on stock yields in (22), assuming full information. Statistics are in the third column of Table VII. Results show that stock yields, assuming full information, perform roughly as well as price-to-dividend ratios in predicting annual stock index returns. This is consistent with results in Figure 4, which show that there is very little difference between the time series of price-to-dividend ratios and stock yields, assuming full information. To show the superior predictive power of stock yields, assuming learning, is significant, we run

bi-variate regressions of future annual stock index returns on both stock yields, assuming learning, and either price-to-dividend ratios or stock yields, assuming full information. Statistics are in the fourth and fifth columns of Table VII. Results show that stock yields, assuming learning, significantly dominate both price-to-dividend ratios and stock yields, assuming full information, in predicting annual stock index returns.

[Table VII is about here.]

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

We can also demonstrate the relevance of our dividend model by showing that stock index prices respond to contemporaneous changes in long-run dividend growth rates from our model better than the alternatives. That is, if investors behave as if they price the stock index using our model, we expect that, all else equal, when dividend expectations rise according to our model, so should prices, and vice versa. We regress annual stock index returns on contemporaneous changes in long-run dividend expectations, assuming investors behave as if they learn using our dividend model. We report regression statistics in the first column of Table VIII. Results confirm that increases in expectations about future dividends is accompanied by more positive stock index returns, and vice versa. In fact, contemporaneous changes to expected dividends account for a statistically significant 10.6% of the variation in annual stock index returns. As points of reference, we also run regressions of annual stock index returns on contemporaneous changes in long-run dividend expectations, either based on our model but assuming full information, or based

21

on assuming learning using one of the alternative models. These results are reported in the second to fifth columns of Table VIII. We note that, under any other cases considered, the relationship between annual stock index returns and contemporaneous changes in expected dividends is negative. Taken as a whole, our findings suggest that the absence of a relationship between dividend expectations and stock index pricing documented in the existing literature may be due to a failure to simultaneously account for corporate payout policy and the role of learning in pricing.[14]

[Table VIII is about here.]

IV. Learning about Dividends in a Dynamic Equilibrium Model

Although long-run discount rates can be uniquely pinned down based on the price-to-dividend ratios and expectations of long-run dividend growth rates, the present value relationship cannot fully capture how discount rates over short horizons vary over time. In other words, the variation of expected long-run returns and expected returns over the short-run are not necessarily perfectly correlated with each other. In this section, we search for a dynamic equilibrium asset pricing model that is able to quantitatively capture the possible role of learning in determining short-run expected returns. That is, a model that, after incorporating parameter uncertainty, is able to show strong performance in predicting short horizon stock index returns that is consistent with the data.

For the rest of this section, we first argue that an asset pricing model's performance in predicting stock index returns can be used to assess that model. Then, we incorporate learning into a long-run risks model and show that 25.3% to 27.1% of the variation in annual stock index returns can be predicted using such a model.

---

[14]See, for example, Cochrane (2008).

A. Return Predictability and Assessing Asset Pricing Models

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

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

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

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

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

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

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

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

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

B. The Long-Run Risks Model

We propose a long-run risks model that combines our dividend model, Epstein and Zin (1989) investor preferences, and persistent consumption growth rates similar to Bansal and Yaron (2004) and show that such a model predicts 25.3% to 27.1% of the variation in annual stock index returns.

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

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

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 $\zeta > 0$.[15] Log of the intertemporal marginal rate of substitution (IMRS) is:

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

---

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

where $\tilde{c} = \log(\tilde{C})$ and $\tilde{R}^c_{t+1}$ denotes the real return of the representative agent's wealth portfolio. For quarterly calibration, we set $\psi = 1.5$ to be consistent with preferences for the early resolution of uncertainty, and set $\alpha = 5$ and $\delta = 0.975$, all of which are within the range of parameter choices commonly made by the existing literature.

Similar to Bansal and Yaron (2004), we assume that consumption and dividend growth rates carry the same persistent latent component x_t. That is, we describe real consumption growth rates as:

$$\Delta \tilde{c}_{t+1} - (\mu_d - \mu_\pi) = \frac{1}{\gamma} (x_t + \sigma_d \epsilon_{c,t+1}). \tag{27}$$

Following Bansal and Yaron (2004), we set the unconditional mean of consumption growth rates to equal to that of dividend growth rates. The parameter γ is the leverage of the equity market. 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.[16] This criticism serves as the rationale for why we expect learning to be important in this context.

Unfortunately, we cannot adopt the Bansal and Yaron (2004) model in its exact form because our dividend model does not feature stochastic volatility, which is a key component of Bansal and Yaron (2004). However, our long-run risks model still needs the additional degree of freedom from a second latent variable to be able to simultaneously capture the time series of dividends and price-to-dividend ratios in the data. So, instead of stochastic volatility, our long-run risks model assumes stochastic correlation between shocks to consumption and shocks to dividend and earnings processes. That is, we assume that the correlations between shocks $\epsilon_{c,t+1}$ to real consumption growth rates and shocks $\epsilon_{d,t+1}$ and $\epsilon_{e,t+1}$ to dividend and earnings growth rates are equal, denoted $\lambda_t = \lambda(\epsilon_{c,t+1}, \epsilon_{d,t+1}) = \lambda(\epsilon_{c,t+1}, \epsilon_{e,t+1})$, and follow an AR[1] process centered around zero:

$$\lambda_{t+1} = \omega \lambda_t + \sigma_\lambda \epsilon_{\lambda,t+1}, \quad \epsilon_{\lambda,t+1} \sim \text{i.i.d. } \mathbb{N}(0,1). \tag{28}$$

---

[16]See Beeler and Campbell (2012), Jagannathan and Marakani (2015).

It can then be derived that the correlation between consumption and retention ratios is:

$$\lambda\left(\epsilon_{c,t+1}, \epsilon_{q,t+1}\right) = \frac{\sqrt{\sigma_d^2 + \sigma_q^2} - \sigma_d}{\sigma_q}\lambda_t \tag{29}$$

For those other cross-correlations of different shocks that we cannot identify, we set them all to zeros. So to summarize, the correlation matrix of shocks to consumption, dividends, and retention ratios is:

$$
\begin{pmatrix}
\epsilon_{c,t+1} \\
\epsilon_{d,t+1} \\
\epsilon_{x,t+1} \\
\epsilon_{\lambda,t+1} \\
\epsilon_{q,t+1} \\
\epsilon_{\pi,t+1}
\end{pmatrix}
\sim \text{ i.i.d. } \mathbb{N}\left(\mathbf{0},
\begin{pmatrix}
1 & \lambda_t & 0 & 0 & \frac{\sqrt{\sigma_d^2+\sigma_q^2}-\sigma_d}{\sigma_q}\lambda_t & 0 \\
\lambda_t & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
\frac{\sqrt{\sigma_d^2+\sigma_q^2}-\sigma_d}{\sigma_q}\lambda_t & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\right). \tag{30}
$$

C. Estimation and Results

We solve our long-run risk model in Appendix D. In solving this model, we closely follow the steps in Bansal and Yaron (2004). The model consists of four state variables: latent variables x_t and λ_t, retention ratios, and inflation rates. We can solve for price-to-dividend ratio as a linear function of these four state variables:

$$p_t - d_t = A_{d,0} + A_{d,1}x_t + A_{d,2}\lambda_t + A_{d,3}\left(q_t - \mu_q\right) + A_{d,4}(\Delta\pi_t - \mu_\pi). \tag{31}$$

Expectation of stock return over the next period is:

$$E_t[R_{t+1}] = A_{r,0} + A_{r,1}x_t + A_{r,2}\lambda_t + A_{r,4}(\Delta\pi_t - \mu_\pi), \tag{32}$$

The coefficients $A_{d,\cdot}$ and $A_{r,\cdot}$, derived in Appendix D, are functions of the parameters that describe investor preferences, and the joint processes of consumption and dividends. We note that, substituting (31) into (32), we can avoid estimating the latent variable λ_t directly from macroeconomic data and instead write expected future returns as a function of price-to-dividend ratios and the other three state variables:

$$E_t[R_{t+1}] = A_0 + A_1 x_t + A_2 q_t + A_3 \Delta\pi_t + A_4(p_t - d_t),$$
$$A_0 = \frac{A_{r,0}A_{d,2} - A_{r,2}A_{d,0}}{A_{d,2}} - \frac{A_{r,2}}{A_{d,2}}\mu_q + \left(\frac{A_{r,2}A_{d,3}}{A_{d,2}} - A_{r,4}\right)\mu_\pi,$$
$$A_1 = \frac{A_{r,1}A_{d,2} - A_{r,2}A_{d,1}}{A_{d,2}}, \quad A_2 = \frac{A_{r,2}}{A_{d,2}}, \quad A_3 = -\frac{A_{r,2}A_{d,3}}{A_{d,2}},$$
$$A_4 = \frac{A_{r,4}A_{d,2} - A_{r,2}A_{d,4}}{A_{d,2}}. \tag{33}$$

Price-to-dividend ratios, earnings-to-dividend ratios, and inflation rates are directly observable. Aside from those in investors preferences, all but three parameters in our long-run risks model, as well as the latent variable x_t, appears in either (6) or (16) and can thus be estimated from dividend dynamics. We follow Krep's learning and use these parameters and state variable estimates as if they were their true values. There are multiple ways through which we can estimate the remaining three parameters, that is, ω, σ_λ, and γ, that are not part of dividends or preferences. One approach is to estimate them from consumption data. However, as discussed in Savov (2011), consumption is measured with significant noise and the right measure of consumption itself is still up for debate. So instead, we estimate them from price-to-dividend ratios. That is, fix a set of parameters ω, σ_λ, and γ, then for each time-t, by substituting in price-to-dividend ratios, retention ratios, inflation rates, and dividend model estimates, we can back out λ_t as:

$$\lambda_t = \frac{(p_t - d_t) - A_{d,0} - A_{d,1}x_t - A_{d,3}\left(q_t - \mu_q\right) - A_{d,4}\left(\Delta\pi_t - \mu_\pi\right)}{A_{d,2}} \tag{34}$$

From (28), we know that the latent variable λ_t should be an AR[1] process with an unconditional distribution of $\mathbb{N}(0, \frac{\sigma_\lambda}{\sqrt{1-\omega^2}})$. So we can choose the set of parameters ω, σ_λ, and γ to best fit these distributional characteristics λ_t. In other words, we solve for ω, σ_λ, and γ using Generalized Method of Moments, fitting the three parameters to the three moments:

$$E[\lambda_t] = 0$$

$$E[\lambda_t^2 - E[\lambda_t]^2] - \frac{\sigma_\lambda^2}{1 - \omega^2} = 0$$

$$E[(\lambda_t - E[\lambda_t])((\lambda_{t+1} - \omega\lambda_t) - E[\lambda_{t+1} - \omega\lambda_t])] = 0 \tag{35}$$

Under the assumption that our long-run risks model holds, exactly three independent moment conditions, as in (35), are required to identify the three parameters ω, σ_λ, and γ. Our choice of the three moment conditions is standard. First, we choose the three parameters so that the sample mean of the latent variable λ_t is set to zero. Second, the sample variance of the latent variable λ_t is set to equal the variance specified in our model. Third, the sample first-order serial covariance of the latent variable λ_t is made to match the covariance specified in our model. Standard errors of parameter estimates are based on bootstrap simulation, as described in Appendix B.

Our choice to estimate ω, σ_λ, and γ from price-to-dividend ratios is consistent with the existing literature on learning from prices.[17] Still, the fact that our model feature consumption but our estimation of the model does not is a drawback of our approach. Including clean consumption data in our estimation, if such data were avaliable, would mean having extra independent observations for estimating state variables and parameters. However, as simulation results in Table VI suggest, the gain in efficiency as a result of having this extra data would be rather limited.[18]

---

[17]For example, the literature on Rational Expectations Equilibrium models.
[18]Also, we need high frequency consumption data to reasonably fit a model with time-varying

To focus on the role of learning about dividends on asset pricing and differentiate ourselves from Johannes, Lochstoer, and Mou (2016), we first run (35) by fitting the three moments of the latent variable λ_t based on the entire data sample between 1946 and 2015. This is a senario where learning is restricted to parameters in the dividend model, that is, learning about dividends. However, because the entire data sample is used, a forward-looking bias may be introduced. To overcome this concern, we also run (35) at each point in time using only data avaliable at the time. That is, in this scenario, learning is applied to parameters both in and beyond the dividend model, that is, full learning.

In Figure 5, we report parameters in our long-run risks model not in the dividend process, estimated based on non-overlapping annual data up to time-τ, for τ between 1975 and 2016. In the same figure, we also report coefficients $A_.$ that relate price-to-dividend ratios and state variables to expected returns in (33), assuming full information, learning about dividends, or full learning. We see that coefficients $A_.$ fluctuates significantly over time. In other words, as learning is introduced, these coefficients become additional model specific state variables in determining stock index expected returns. This observation is consistent with the findings of Collin-Dufresne, Johannes, and Lochstoer (2016). Unsurprisingly, the fluctuations of coefficients $A_.$ are more intensive for full learning than learning about dividends.

[Figure 5 is about here.]

In Figure 6, we report the evolution of the four state variables of our long-run risks model, that is, latent variables x_t and λ_t, earnings-to-dividend ratios, and inflation rates, as well as expected excess stock index returns returns and risk free rate, over time, assuming full information, learning about dividends, or full learning. We see that, consistent with data, most of the variation in expected stock index returns is attributable

_____

correlations in dividends and consumption shocks.

to variation in expected excess returns. Unsurprisingly, learning increases the volatility of both expected excess returns and risk free rate. Interestingly, Figure 6 suggests that, around the time of the 2001 recession, expectations of one-year excess returns are negative. This is a result of model implied correlations between shocks to dividends and consumption being highly negative, that is, the stock index temporarily serving as a hedge to consumption, during this period and is an equilibrium outcome, based on parameters estimated from data, of our model.

[Figure 6 is about here.]

We examine how our long-run risks model, assuming either learning about dividends or full learning, that is, our learning models, perform in predicting annual stock index returns. We then measure forecasting performance using out-of-sample R-square value:

$$\mathcal{R}_{\mathcal{O}}^2(\mathcal{L}) = 1 - \frac{\sum_{t=T_0}^{T-1} \left(R_{t+1} - E_t \left[R_{t+1} | \mathcal{L} \right] \right)^2}{\sum_{t=T_0}^{T-1} \left(R_{t+1} - \hat{\mu}_{r,t} \right)^2}. \tag{36}$$

where $\mathcal{L}$ stands for learning. We use data since 1946 as the training period and compute the out-of-sample R-square value using non-overlapping annual data between 1975 and 2016. In the second and third rows of Table IX, we report out-of-sample R-square value for predicting annual stock index returns using our learning models. We find that, between 1975 and 2016, our learning models predict 25.3% to 27.1% of the variation in annual stock index returns out-of-sample.

To better quantify the incremental contribution of learning to the model's performance in predicting annual stock index returns, we compute expected returns in (32) using dividend model parameters estimated based on the entire data sample between 1975 and 2016, that is, our full information model. We report out-of-sample R-square value for predicting stock index returns using our full information model in the third row of Table IX. From learning to full information, R-square value reduces from at least 25.3% to 13.3%. So learning acounts for approximately half of the return predictability

documented. To examine the significance of this difference, we report, also in Table IX, incremental R-square value of our learning models over our full information model:

$$\mathcal{R}_I^2(\mathcal{L}, \mathcal{F}) = 1 - \frac{\sum_{t=T_0}^{T-1} \left(R_{t+1} - E_t[R_{t+1}|\mathcal{L}]\right)^2}{\sum_{t=T_0}^{T-1} \left(R_{t+1} - E_t[R_{t+1}|\mathcal{F}]\right)^2}. \tag{37}$$

Results from Table IX show that there is a statistically significant gain in forecasting performance from modeling investors' learning about dividend dynamics.[19] Also, performance for predicting annual stock index returns is slightly better for full learning than learning about dividends. However, we do not have enough statistical power to conclude that learning about parameters beyond the dividend model plays a statistically significant role.

[Table IX is about here.]

For additional details on how our learning models' forecasting performance evolves over time, at each time-τ, we follow Goyal and Welch (2008) and define the cumulative sum of squared errors difference (SSED) between predicting annual stock index return using our learning models and using the historical mean of returns as:

$$\mathcal{D}_\tau(\mathcal{L}) = \sum_{t=T_0}^{\tau-1} \left(R_{t+1} - E_t[R_{t+1}|\mathcal{L}]\right)^2 - \sum_{t=T_0}^{\tau-1} \left(R_{t+1} - \hat{\mu}_{r,\tau}\right)^2. \tag{39}$$

The SSED for our learning models are plotted in Figure 7. If the forecasting performance of our learning model is stable and robust over time, we should observe a steady but constant and consistent 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. A flat SSED suggests that our model neither adds or destroys

---

[19]We note that $\mathcal{R}_I^2(\mathcal{L}, \mathcal{F})$, $\mathcal{R}^2(\mathcal{L})$ and out-of-sample R-square value of our full information model, that is, $\mathcal{R}^2(\mathcal{F})$, are related through the following equation:

$$\mathcal{R}_I^2(\mathcal{L}, \mathcal{F}) = 1 - \frac{1 - \mathcal{R}^2(\mathcal{L})}{1 - \mathcal{R}^2(\mathcal{F})}. \tag{38}$$

forecasting performance. We note that our model's forecasting performance is positive through the majority of the data sample. Overall, as shown in Figure 7, most of the forecasting performance can be attributed to the early three-fourth of the sample, while performance is relatively flat during the most recent decade.

[Figure 7 is about here.]

To see the incremental contribution of learning to SSED over time, we plot, in Figure 8, the incremental SSED defined as the difference in SSED between our learning model and our full information model:

$$\mathcal{D}_\tau(\mathcal{L}) - \mathcal{D}_\tau(\mathcal{F}) = \sum_{t=T_0}^{\tau-1} (R_{s+1} - E_t[R_{s+1}|\mathcal{L}])^2 - \sum_{s=T_0}^{t-1} (R_{s+1} - E_t[R_{s+1}|\mathcal{F}])^2. \qquad (40)$$

Similar to what we observed earlier, the incremental gain in forecasting performance from learning is large and reasonably consistent, but is mostly concentrated in the early three-fourth of the sample.

[Figure 8 is about here.]

1. Long-Run Risks Model and Other Return Forecasts

Goyal and Welch (2008) document that empirical forecasts of stock index returns overwhelmingly lack out-of-sample predictive power. However, subsequent literature, such as Kelly and Pruitt (2013) and Li, Ng, and Swaminathan (2013), overcomes the Goyal and Welch (2008) critique and finds out-of-sample return predictability. We compare the out-of-sample forecasting performance of our long-run risks model, assuming either learning about dividends or full learning, with these more successful empirical proxies of stock index returns in the existing literature.

To avoid the concern of selection bias in our reporting of results, we compare our learning models with these emprical proxies of returns based on the sample period used

32

by the corresponding original authors. For Kelly and Pruitt (2013), this is between 1981 and 2009. For Li, Ng, and Swaminathan (2013), this is between 1995 and 2013. To evaluate performance, we report, in Table X, incremental R-square values for predicting stock index returns using our learning models over these alternative expected return proxies for the selected periods. Results show that our learning models outperform these empirical proxies of expected stock index returns.[20]

[Table X is about here.]

2. Recession versus Expansion

Figure 7 suggests that the times around the 2001 recession plays an especially important role in the return predictability results. It appears, from Figure 7, that the learning models' performance in most of the other recessions are positive as well. To make more clear on how return predictability differs between periods of expansions and periods of recessions, we divide our data sample between 1975 and 2016 into expansion versus recession periods and report separate forecasting performance results. We define a year to be in recession if any of its months overlap with the NBER recession dates. There are six such years in our 42 years data sample. Results are reported in Table XI. We find that the forecasting performance of our learning models is much stronger during recessions than expansions, but performance during expansions is nevertheless robust, with an R-square value of 19.1% to 19.6% assuming learning versus 13.2% assuming full information. The finding that predictability is strongest during market downturns is not surprising and is consistent with the existing literature. For example, Golez and Koudijs (2017) find, based on four centuries of stock market data, that most of the predictability of future stock returns using price-to-dividend ratios stems from recessions. However, in both expansions and recessions, our learning models outperform our full information

---

[20]Although when compared to Li, Ng, and Swaminathan (2013), the outperformance is not statistically significant due to the much shorter data sample.

model, suggesting that learning plays an important role regardless of economic conditions.

[Table XI is about here.]

3. The Role of Epstein and Zin (1989) Preferences

To emphasize that Epstein and Zin (1989) preferences are critical to our return predictability results, we build a model where we replace Epstein and Zin (1989) preference with Constant Relative Risk Aversion (CRRA) preferences:

$$U_t = \sum_{t=0}^{\infty} \delta^t \frac{\tilde{C}_t^{1-\alpha}}{1-\alpha} \tag{41}$$

and set $\alpha = 5$ and $\delta = 0.975$. While estimates of parameters in the dividend model do not change with preferences, the three remaining parameters, ω, σ_λ, and γ, need to be reset. That is, we estimate parameters ω, σ_λ, and γ of our model using General Method of Moments by fitting the same set of moments in (35) under CRRA preferences and the chosen preference parameters. We then derive expected returns under CRRA preferences. We report, in Table XII, R-square values for predicting annual stock index returns using the CRRA model, assuming learning about dividends, full learning, or full information. We see that, assuming learning, R-square values for predicting annual stock index returns reduce from at least 25.3% for Epstein and Zin (1989) preferences to at most 15.1% for CRRA preferences, and the lack of the incremental contribution of learning to R-square value accounts for most of this reduction. These results suggest that modeling investors behavior using CRRA preferences cannot fully capture the effect of learning on expected stock index returns.

[Table XII is about here.]

34

V. Conclusion

In this paper, we develop a time series model for dividend growth rates that is inspired by both the latent variable model of Cochrane (2008), van Binsbergen and Koijen (2010), and others and the vector-autoregressive model of Campbell and Shiller (1988b). The model shows strong performance in predicting annual 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 evaluate the economic and statistical significance of learning about parameters in the dividend process in determining asset prices and expected 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 our long run risks model, featuring Epstein and Zin (1989) preferences and persistence shocks to dividends and consumption, assuming learning, explains 25.3% to 27.1% of the variation in annual stock index returns, shutting down learning reduces the R-square value to 13.3%. This drop in R-square value is statistically significant and economically meaningful. We also show that we cannot replicate our learning results under CRRA preferences. Our findings highlight the joint importance of investors aversion to long run risks and investors learning about these risks in understanding asset pricing dynamics.

REFERENCES

ALBUQUERQUE, R., M. S. EICHENBAUM, AND S. REBELO (2015): "Valuation Risk and Asset Pricing," *Working Paper.*

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

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

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

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

BIDDER, R., AND I. DEW-BECKER (2016): "Long-Run Risk is the Worst-Case Scenario," *The American Economic Review*, 106.

BOUDOUKH, J., R. MICHAELY, M. RICHARDSON, AND M. R. ROBERTS (2007): "On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing," *The Journal of Finance*, 62, 877–915.

BREEN, W., L. R. GLOSTEN, AND R. JAGANNATHAN (1989): "Economic Significance of Predictable Variations in Stock Index Returns," *Journal of Finance*, 44, 1177–1189.

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

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

CAMPBELL, J. Y., AND T. VUOLTEENAHO (2004): "Bad Beta, Good Beta," *Journal of Finance*, 55, 2219–2257.

CHEN, L., Z. DA, AND R. PRIESTLEY (2012): "Dividend Smoothing and Predictability," *Management Science*, 58, 1834–1853.

CHEN, L., Z. DA, AND X. ZHAO (2013): "What Drives Stock Price Movements," *Review of financial studies*, 26, 841–876.

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

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

COLLIN-DUFRESNE, P., M. JOHANNES, AND L. A. LOCHSTOER (2016): "Parameter Learning in General Equilibrium: The Asset Pricing Implications," *The American Economic Review*, 106, 664–698.

CROCE, M. M., M. LETTAU, AND S. C. LUDVIGSON (2014): "Investor Information, Long-Run Risk, and the Term Structure of Equity," *Working Paper*.

DA, Z., R. JAGANNATHAN, AND J. SHEN (2014): "Growth Expectations, Dividend Yields, and Future Stock Returns," *Working Paper*.

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

FAMA, E. F. (1981): "Stock Returns, Real Activity, Inflation, and Money," *The American Economic Review*, 71, 545–565.

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 Expeted Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

GOLEZ, B., AND P. KOUDIJS (2017): "Four Centuries of Return Predictability," *Journal of Financial Economics*, Forthcoming.

GORDON, M. J. (1959): "Dividends, Earnings, and Stock Prices," *Review of Economics and Statistics*, 41, 99–105.

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

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

HANSEN, L. P. (2007): "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," *The American Economic Review*, 97, 1–30.

HANSEN, L. P., AND T. J. SARGENT (2010): "Fragile Beliefs and the Price of Uncertainty," *Quantitative Economics*, 1, 129–162.

JAGANNATHAN, R., AND S. MARAKANI (2015): "Price-Dividend Ratio Factor Proxies for Long-Run Risks," *Review of Asset Pricing Studies*, 5, 1–47.

JAGANNATHAN, R., E. MCGRATTAN, AND A. SCHERBINA (2000): "The Declining US Equity Premium," *Quarterly Review*, pp. 3–19.

JAGANNATHAN, R., E. R. MCGRATTAN, AND A. SCHERBINA (2001): "The Declining US Equity Premium," *Working Paper*.

JOHANNES, M., L. A. LOCHSTOER, AND Y. MOU (2016): "Learning about Consumption Dynamics," *The Journal of Finance*, pp. 551–600.

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.

LINTNER, J. (1956): "Distribution of Incomes of Corporations Among Dividends, Retained Earnings, and Taxes," *American Economic Review*, 46, 97–113.

MARTIN, I. (2017): "What is the Expected Return on the Market?," *The Quarterly Journal of Economics*, 132, 367–433.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

PIAZZESI, M., AND M. SCHNEIDER (2006): "Equilibrium Yield Curves," *NBER Macroeconomics Annual 2006*, 21.

——— (2010): "Interest Rate Risk in Credit Markets," *The American Economic Review*, 100, 579–584.

POLK, C., S. THOMPSON, AND T. VUOLTEENAHO (2006): "Cross-Sectional Forecasts of the Equity Premium," *Journal of Finance*, 55, 2219–2257.

SAVOV, A. (2011): "Asset Pricing with Garbage," *The Journal of Finance*, 66, 177–201.

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. Estimation of Parameters in Our Dividend Model

We estimate parameters of the following system of equations that jointly describe the dividend, earnings, and inflation processes (see (6), (9), and (16)).

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

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

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

$$\Delta \pi_{t+1} - \mu_\pi = \eta \left(\Delta \pi_t - \mu_\pi \right) + \sigma_\pi \epsilon_{\pi,t+1},$$

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

To estimate parameters in the third equation of (42), we run an autoregression on retention ratios:

$$q_{t+1} - \mu_q = \theta \left(q_t - \mu_q \right) + \sigma_q \epsilon_{q,t+1}, \quad \epsilon_{q,t+1} \sim \text{i.i.d. } \mathbb{N}(0,1). \tag{43}$$

To estimate parameters in the fourth equation of (42), we run an autoregression on inflation rates:

$$\Delta \pi_{t+1} - \mu_\pi = \eta \left(\Delta \pi_t - \mu_q \right) + \sigma_\pi \epsilon_{\pi,t+1}, \quad \epsilon_{\pi,t+1} \sim \text{i.i.d. } \mathbb{N}(0,1). \tag{44}$$

For the remaining parameters in the first to second equations of (42), we note that dividend growth rates and contemporaneous earnings are related, as shocks to dividends also impact contemporaneous earnings in (9), and vice versa. So we estimate the process

of dividends and earnings through the following system of equations:

$$\Delta d_{t+1} = a_1 + y_{t+1} + b_1 \Delta e_{t+1} + b_2 q_t + \nu_{d,t+1}$$

$$y_{t+1} = b_3 y_t + \nu_{y,t+1}$$

$$\begin{pmatrix} \nu_{d,t+1} \\ \nu_{y,t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N} \left(\mathbf{0}, \begin{pmatrix} \varsigma_d & 0 \\ 0 & \varsigma_y \end{pmatrix} \right). \tag{45}$$

To apply the Kalman filter, let $\hat{y}_{t|s}$ denote the time-s expectation of the latent variable y_t and $P_{t|s}$ denote the variance of y_t conditioning on information in time-s. Set initial conditions $\hat{y}_{0|0} = 0$ and $P_{0|0} = \frac{\sigma_y^2}{1-b_3^2}$. We can then iterate the following system of equations:

$$\hat{y}_{t+1|t} = b_3 \hat{y}_{t|t}, \quad P_{t+1|t} = b_3^2 P_{t|t} + \varsigma_y^2,$$

$$\epsilon_{t+1} = \Delta d_{t+1} - a_1 - \hat{y}_{t+1|t} - b_1 \Delta e_{t+1} - b_2 q_t,$$

$$\hat{y}_{t+1|t+1} = \hat{y}_{t+1|t} + \frac{P_{t+1|t}}{P_{t+1|t} + \varsigma_d^2} \epsilon_{t+1}, \quad P_{t+1|t+1} = P_{t+1|t} - \frac{P_{t+1|t}^2}{P_{t+1|t} + \varsigma_d^2}. \tag{46}$$

At each time τ, to estimate parameters in (45), we maximize the log-likelihood function:

$$l = -\sum_{t=0}^{\tau-1} \left(\log \left(P_{t+1|t} + \varsigma_d^2 \right) + \frac{\epsilon_{t+1}^2}{P_{t+1|t} + \varsigma_d^2} \right). \tag{47}$$

Throughout, we apply Kalman filter based on non-overlapping annual data. To then map parameters in (45) to those in the first and second equation of (42), we substitute $\Delta e_{t+1} = q_{t+1} - q_t + \Delta d_{t+1}$ and (43) into (45) and re-arrange (45) into:

$$\Delta d_{t+1} - \frac{a_1 + b_2 \mu_q}{1 - b_1} = x_t + \frac{b_1(\theta - 1) + b_2}{1 - b_1} \left(q_t - \mu_q \right) + \frac{b_1}{1 - b_1} \sigma_q \epsilon_{q,t+1} + \frac{1}{1 - b_1} \nu_{d,t+1}$$

$$x_{t+1} = b_3 x_t + \nu_{x,t+1}, \quad x_t = \frac{1}{1 - b_1} y_t,$$

$$\begin{pmatrix} \nu_{d,t+1} \\ \nu_{x,t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N} \left(\mathbf{0}, \begin{pmatrix} \varsigma_d & 0 \\ 0 & \frac{1}{1-b_1} \varsigma_y \end{pmatrix} \right). \tag{48}$$

Thus, the mapping is:

$$\mu_d = \frac{a_1 + b_2 \mu_q}{1 - b_1}, \quad \phi = \frac{b_1(\theta - 1) + b_2}{1 - b_1}, \quad \rho = b_3,$$

$$\sigma_x = \frac{1}{1 - b_1}\varsigma_y, \quad \sigma_d = \sqrt{\left(\frac{b_1}{1 - b_1}\right)^2 \sigma_q^2 + \left(\frac{1}{1 - b_1}\right)^2 \varsigma_d^2}. \tag{49}$$

B. Bootstrap Simulation

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

$$\begin{pmatrix} \epsilon_{d,t+1} \\ \epsilon_{x,t+1} \\ \epsilon_{q,t+1} \end{pmatrix} \sim \text{ i.i.d. } \mathbb{N} \left(\mathbf{0}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right). \tag{50}$$

Dividend model parameters used for simulations are those reported in Table II, which are estimated based on the full data sample between 1946 and 2015. In our simulations, we use these estimates as if they were the true parameter values. From these innovations, we can simulate the latent variable x_t and retention ratios iteratively as:

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

$$q_{t+1} - \mu_q = \theta \left(q_t - \mu_q \right) + \sigma_q \epsilon_{q,t+1}. \tag{51}$$

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

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

$$\Delta e_{t+1} = \Delta q_{t+1} + \Delta d_{t+1}. \tag{52}$$

43

To simulate price-to-dividend ratios, we use (31), which is derived from our long-run risks model.

C. Proof of Proposition 1

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

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

$$= E\left[(R_{t+1} - E_t[R_{t+1}|\mathcal{M}_i])^2\right] + E\left[\epsilon_{t+1}^2\right] - 2 \cdot E\left[(R_{t+1} - \mathbb{E}_t[R_{t+1}|\mathcal{M}_i])\,\epsilon_{t+1}\right]$$

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

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

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. $E\left[\epsilon_{t+1}^2\right]$ and $E\left[R_{t+1}\epsilon_{t+1}\right]$ are independent of the model $\mathcal{M}_i$ and so:

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

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

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

D. Derivation of Price-Dividend Ratios and Expected Returns in 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 (3), consumption dynamics in (27), and investors preferences in (25). Our model differs from Bansal and Yaron (2004), as discussed in the body of the paper. Nevertheless, we can still use the methodology in Bansal and Yaron (2004) to solve for prices and expected returns in our model. The log stochastic discount factor is given as:

$$m_{t+1} = \zeta \log(\delta) - \frac{\zeta}{\psi} \Delta \tilde{c}_{t+1} + (\zeta - 1) \tilde{R}_{t+1}^c. \tag{53}$$

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

$$\tilde{R}_{t+1}^c = g_0 + g_1 z_{c,t+1} - z_{c,t} + \Delta \tilde{c}_{t+1}. \tag{54}$$

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:

$$z_{c,t} = A_{c,0} + A_{c,1} x_t. \tag{55}$$

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

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

$$var_t \left(m_{t+1} + \tilde{R}_{t+1}^c \right) = \zeta^2 \left(1 - \frac{1}{\psi} \right)^2 \sigma_c^2 + \zeta^2 (g_1 A_{c,1})^2 \sigma_x^2. \tag{56}$$

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

$$A_{c,0} = \frac{\log(\delta) + (1 - \frac{1}{\psi})\mu_c + g_0 + \frac{1}{2}\zeta(1 - \frac{1}{\psi})^2\sigma_c^2 + \frac{1}{2}\zeta(g_1 A_{c,1})^2\sigma_x^2}{1 - g_1},$$

$$A_{c,1} = \frac{\left(1 - \frac{1}{\psi}\right)\frac{1}{\gamma}}{1 - g_1\rho}. \tag{57}$$

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

$$\tilde{R}_{t+1} = \kappa_0 + \kappa_1 z_{d,t+1} - z_{d,t} + \Delta\tilde{d}_{t+1}. \tag{58}$$

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

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

$$z_{d,t} = A_{d,0} + A_{d,1}x_t + A_{d,2}\lambda_t + A_{d,3}(q_t - \mu_q) + A_{d,4}(\Delta\pi_t - \mu_\pi). \tag{59}$$

Then note that:

$$
\begin{aligned}
E_t\left[m_{t+1} + \tilde{R}_{t+1}\right] &= \zeta\log(\delta) + (\zeta - 1)(g_1 - 1)A_{c,0} + (\zeta - 1)(g_1\rho - 1)A_{c,1}x_t \\
&\quad + \left(\zeta - \frac{\zeta}{\psi} - 1\right)(\mu_c + \gamma x_t) + (\zeta - 1)g_0 + \kappa_0 + (\kappa_1 - 1)A_{d,0} + (\kappa_1\rho - 1)A_{d,1}x_t \\
&\quad + (\kappa_1\omega - 1)A_{d,2}\lambda_t + (\kappa_1\theta - 1)A_{d,3}(q_t - \mu_q) + (\kappa_1\eta - 1)A_{d,4}(\Delta\pi_t - \mu_\pi) \\
&\quad + \mu_c + x_t + \phi(q_t - \mu_q) - \eta(\Delta\pi_t - \mu_\pi). \\
var_t\left(m_{t+1} + \tilde{R}_{t+1}\right) &= \left(\zeta - 1 - \frac{\zeta}{\psi}\right)^2\sigma_c^2 + \sigma_d^2 + ((\zeta - 1)g_1 A_{c,1} + \kappa_1 A_{d,1})^2\sigma_x^2 \\
&\quad + (\kappa_1 A_{d,2})^2\sigma_\lambda^2 + (\kappa_1 A_{d,3})^2\sigma_q^2 + (\kappa_1 A_{d,4})^2\sigma_\pi^2 + 2\left(\zeta - 1 - \frac{\zeta}{\psi}\right)\sigma_c\sigma_d\lambda_t \\
&\quad + 2\left(\zeta - 1 - \frac{\zeta}{\psi}\right)(\kappa_1 A_{d,3})\left(\sqrt{\sigma_d^2 + \sigma_q^2} - \sigma_d\right)\sigma_c\lambda_t. \tag{60}
\end{aligned}
$$

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

$$A_{d,0} = \frac{\begin{pmatrix} \zeta \log(\delta) + (\zeta - 1)g_0 + (\zeta - 1)(g_1 - 1)A_{c,0} \\ + \left(\zeta - \frac{\zeta}{\psi} - 1\right)\mu_c + \kappa_0 + \mu_c + \frac{1}{2}\sigma_d^2 + \frac{1}{2}((\zeta - 1)g_1 A_{c,1} + \kappa_1 A_{d,1})^2\sigma_x^2 \\ + \frac{1}{2}(\kappa_1 A_{d,2})^2\sigma_\lambda^2 + \frac{1}{2}(\kappa_1 A_{d,3})^2\sigma_q^2 + \frac{1}{2}(\kappa_1 A_{d,4})^2\sigma_\pi^2 \\ + \frac{1}{2}\left(\zeta - \zeta - \frac{\zeta}{\psi}\right)^2\sigma_c^2 \end{pmatrix}}{1 - \kappa_1},$$

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

$$A_{d,2} = \frac{\left(\zeta - 1 - \frac{\zeta}{\psi}\right)\left((\kappa_1 A_{d,3})\left(\sqrt{\sigma_d^2 + \sigma_q^2} - \sigma_d\right) + \sigma_d\right)\sigma_c}{1 - \kappa_1\omega},$$

$$A_{d,3} = \frac{\phi}{1 - \kappa_1\theta}, \quad A_{d,4} = \frac{-\eta}{1 - \kappa_1\eta}. \tag{61}$$

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

$$E_t[\tilde{R}_{t+1}] = A_{r,0} + A_{r,1}x_t + A_{r,2}\lambda_t + A_{r,3}(q_t - \mu_q) + A_{r,4}(\Delta\pi_t - \mu_\pi), \tag{62}$$

where:

$$A_{r,0} = \kappa_0 - (1 - \kappa_1)A_{d,0} + \mu_d, \quad A_{r,1} = 1 - (1 - \kappa_1\rho)A_{d,1},$$

$$A_{r,2} = -(1 - \kappa_1\omega)A_{d,2}, \quad A_{r,3} = \phi - (1 - \kappa_1\theta)A_{d,3}, \quad A_{r,4} = -\eta - (1 - \kappa_1\eta)A_{d,4}. \tag{63}$$

Expected real return over the next τ period is:

$$\sum_{s=0}^{\tau-1}\tilde{R}_{t+s+1} = \tau A_{r,0} + \left(\sum_{s=0}^{\tau-1}A_{r,1}\rho^s\right)x_t + \left(\sum_{s=0}^{\tau-1}A_{r,2}\omega^s\right)\lambda_t + \left(\sum_{s=0}^{\tau-1}A_{r,3}\theta^s\right)(q_t - \mu_q)$$

$$+ \left(\sum_{s=0}^{\tau-1}A_{r,4}\eta^s\right)(\Delta\pi_t - \mu_\pi) \tag{64}$$

For nominal returns, add expected inflation based on the AR[1] model.

47

E. Estimating ω, σ_λ, and γ by the Generalized Method of Moments

The three parameters of interest in our General Method of Moments (GMM) estimation are the AR[1] coefficient ω, volatility σ_λ, which should be strictly above 0, and leverage γ. The three moment conditions are described in (35). First, note that we can use the third moment condition in (35) to write ω as:

$$\omega = \frac{\mu\left((\lambda_{t+1} - \mu(\lambda_{t+1}))(\lambda_t - \mu(\lambda_t))\right)}{\mu\left(\lambda_t^2 - \mu(\lambda_t)^2\right)} \tag{65}$$

where $\mu(\cdot)$ is the sample mean function. Substituting in (65), we can then use the second moment condition in (35) to write σ_λ as:

$$\sigma_\lambda = \sqrt{(1-\omega^2)\,\mu\left(\lambda_t^2 - \mu(\lambda_t)^2\right)}$$
$$= \sqrt{\left(1 - \left(\frac{\mu\left((\lambda_{t+1} - \mu(\lambda_{t+1}))(\lambda_t - \mu(\lambda_t))\right)}{\mu\left(\lambda_t^2 - \mu(\lambda_t)^2\right)}\right)^2\right)\mu\left(\lambda_t^2 - \mu(\lambda_t)^2\right)} \tag{66}$$

So for a given γ, we can solve for the corresponding equilibrium ω and σ_λ, denoted ω^* and σ_λ^*, by finding ω and σ_λ that are the fixed points to the system of expressions in (34), (65), and (66). In other words, by finding these fixed points in equilibrium, the second and third moment conditions in (35) are automatically satisfied.

We can then solve for γ based on the first moment condition in (35), while satisfying the equilibrium $\omega = \omega^*$ and $\sigma_\lambda = \sigma_\lambda^*$. We can do this numerically by maximizing $l = -\mu(\lambda_t)^2$.

[Table AI is about here.]

[Table AII is about here.]

[Table AIII is about here.]

[Table AIV is about here.]

48

[Table AV is about here.]

[Table AVI is about here.]

[Table AVII is about here.]

[Figure AI is about here.]

β_{10}	β_{11}	β_{12}	β_{13}
0.045	0.425***	0.184***	-0.217***
(0.054)	(0.061)	(0.064)	(0.077)
β_0	β_1	β_2	
-0.037	0.455***	0.147***	
(0.024)	(0.070)	(0.052)	

Table I: **Campbell and Shiller (1988b) Betas for Predicting Dividend Growth Rates**: This table reports coefficients from predicting dividend growth rates using the vector-autoregression specification in Campbell and Shiller (1988b). Statistics are based on non-overlapping annual data between 1946 and 2016. Reported in parentice are Newey and West (1987) standard errors that account for up to 10 years of serial correlations. Estimates significant at 90%, 95%, and 99% confidence levels are highlighted using $*$, $**$, and $***$.

μ_d	ϕ	σ_d
0.064	0.140	0.015
(0.016)	(0.021)	(0.017)
	ρ	σ_x
	0.469	0.048
	(0.168)	(0.011)
μ_q	θ	σ_q
0.729	0.370	0.251
(0.065)	(0.120)	(0.026)

Table II: **Dividend Model Parameter Estimates**: This table reports estimated parameters from our dividend model. Dividends are estimated based on non-overlapping annual data since 1946. Reported in parentice are bootstrap simulated standard errors.

| | In-Sample | Out-of-Sample | |
	$\mathcal{R}^2$	$\mathcal{R}^2_{\mathcal{O}}$	p-val.
Our Model	0.464	0.413	0.000
van Binsbergen and Koijen (2010)	0.174	0.161	0.008
Campbell and Shiller (1988b)	0.278	0.256	0.001
Johannes, Lochstoer, and Mou (2016)	0.137	-0.042	1.000

| | Out-of-Sample | |
	$\mathcal{R}^2_{\mathcal{I}}$	p-val.
van Binsbergen and Koijen (2010)	0.301	0.000
Campbell and Shiller (1988b)	0.212	0.002
Johannes, Lochstoer, and Mou (2016)	0.437	0.000

Table III: **Dividend Growth Rates and Expected Growth Rates**. The table on the top reports R-square values for predicting dividend growth rates using our dividend model, the latent variable model in van Binsbergen and Koijen (2010), the VAR model in Campbell and Shiller (1988b), or the Markov-switching model in Johannes, Lochstoer, and Mou (2016). The first column reports in-sample R-square values. The second and third columns report out-of-sample R-square values and the corresponding bootstrap simulated p-values. The table on the bottom reports incremental R-square values for predicting dividend growth rates using our model over one of the baseline models. Dividends are estimated based on non-overlapping annual data since 1946. In-sample (out-of-sample) statistics are based on non-overlapping annual data between 1946 and 2016 (1975 and 2016).

	μ_π	η	σ_π
	0.036	0.557	0.027
	(0.014)	(0.111)	(0.018)

	In-Sample	Out-of-Sample	
	$\mathcal{R}^2$	$\mathcal{R}^2_{\mathcal{O}}$	p-val.
Our Model	0.448	0.540	0.000

Table IV: **Inflation Model Parameter Estimates and Inflation Predictability**: The table on the top reports estimated parameters from our inflation model, based on non-overlapping data between 1946 and 2016. Reported in parentice are bootstrap simulated standard errors. The table on the bottom reports R-square values for predicting inflation rates using our inflation model. The first column reports in-sample R-square value. The second and third columns report the out-of-sample R-square value and the corresponding bootstrap simulated p-value. In-sample (out-of-sample) statistics are based on non-overlapping annual data between 1946 and 2016 (1975 and 2016).

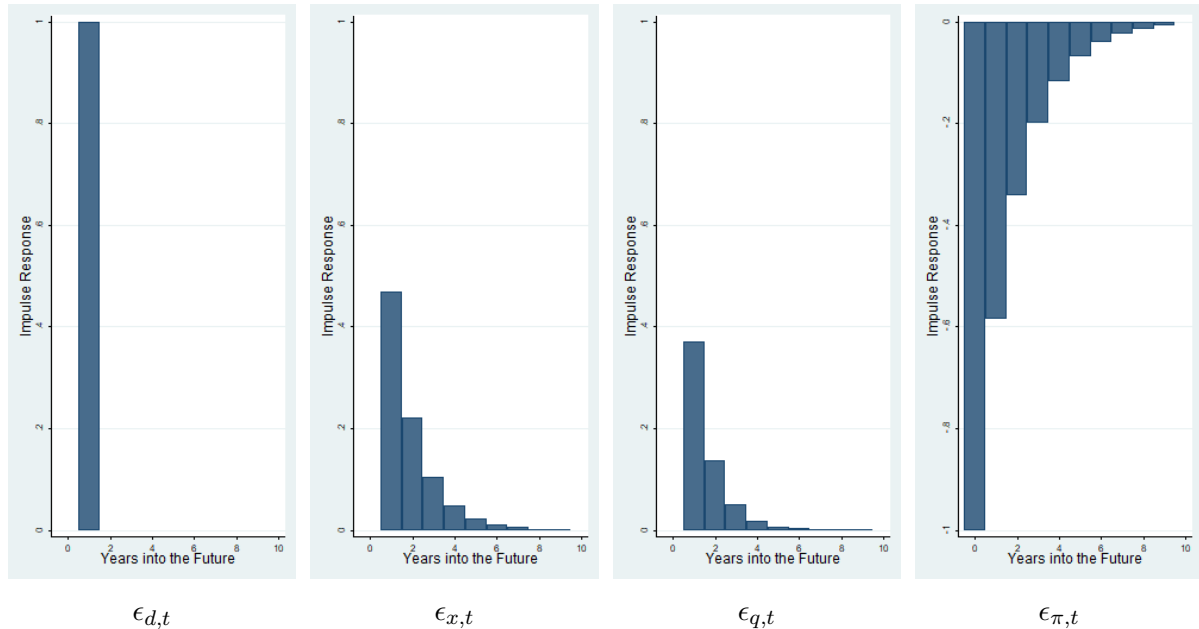$\epsilon_{d,t}$ $\epsilon_{x,t}$ $\epsilon_{q,t}$ $\epsilon_{\pi,t}$

Figure 1: **Impulse Response Functions of Dividend Shocks**. This figure plots the changes to real annual dividend growth rates immediately and expected real dividend growth rates over the next 10 years as a result of a unit change in shocks to the dividend process, that is, $\epsilon_{d,t}$, $\epsilon_{x,t}$, $\epsilon_{q,t}$, and $\epsilon_{\pi,t}$.

	In-Sample	Out-of-Sample	
	$\mathcal{R}^2$	$\mathcal{R}^2_{\mathcal{O}}$	p-val.
Our Model	0.424	0.395	0.000
van Binsbergen and Koijen (2010)	0.160	0.146	0.012
Campbell and Shiller (1988b)	0.259	0.234	0.001
Johannes, Lochstoer, and Mou (2016)	0.172	-0.058	1.000

	Out-of-Sample	
	$\mathcal{R}^2_{\mathcal{I}}$	p-val.
van Binsbergen and Koijen (2010)	0.292	0.000
Campbell and Shiller (1988b)	0.210	0.002
Johannes, Lochstoer, and Mou (2016)	0.428	0.000

Table V: **Dividend Growth Rates and Expected Growth Rates (Real Rates)**. The table on the top reports R-square values for predicting (real) dividend growth rates using our dividend model, the latent variable model in van Binsbergen and Koijen (2010), the VAR model in Campbell and Shiller (1988b), or the Markov-switching model in Johannes, Lochstoer, and Mou (2016). The first column reports in-sample R-square values. The second and third columns report out-of-sample R-square values and the corresponding bootstrap simulated p-values. The table on the bottom reports incremental R-square values for predicting dividend growth rates using our model over one of the baseline models. Dividends are estimated based on non-overlapping annual data since 1946. In-sample (out-of-sample) statistics are based on non-overlapping annual data between 1946 and 2016 (1975 and 2016).

μ_d	ϕ	σ_d	ρ	σ_x	μ_q	θ	σ_q
0.078	0.076	0.067	0.036	0.041	0.061	0.068	0.080

Table VI: **Speed of Learning about Dividend Model Parameters**: This table reports the speed of learning for the eight parameters in our dividend model. Speed of learning is defined as one minus the inverse ratio between the bootstrap simulated standard errors assuming that parameters are estimated based on data between 1946 and 2016 and the bootstrap simulated standard errors assuming that parameters are estimated based on 10 more years of data.

		Our Model			Baseline Model		
					vBK	CS	JLM
$p_t - d_t$	-0.130***		0.014				
	(0.035)		(0.078)				
$sy_t(\mathcal{L})$		4.399***	4.748**	5.423**	3.379***	4.160***	1.965**
		(0.775)	(2.137)	(2.024)	(0.850)	(1.216)	(0.843)
$sy_t(\mathcal{F})$		4.097***		-1.282			
		(1.036)		(2.100)			
$\mathcal{R}^2$	0.136	0.187	0.130	0.187	0.190	0.114	0.114	0.054

Table VII: **Stock Index Returns and Stock Yields**: This table reports the coefficient estimates and R-square values from regressing future stock index returns on log price-to-dividend ratios and stock yields, computed using our dividend model, the latent variable model in van Binsbergen and Koijen (2010) (vBK), the VAR model in Campbell and Shiller (1988b) (CS), or the Markov-switching model in Johannes, Lochstoer, and Mou (2016) (JLM), and assuming investors either learn, that is, $sy_t(\mathcal{L})$, or do not learn, that is, $sy_t(\mathcal{F})$, about dividends. Dividends are estimated based on non-overlapping annual data since 1946. Regressions are based on non-overlapping annual data between 1975 and 2016. Reported in parentice are Newey and West (1987) standard errors that account for up to 10 years of serial correlation. Estimates significant at 90%, 95%, and 99% confidence levels are highlighted using $*$, $**$, and $***$.

| | Our Model | | Baseline Model | | |
			vBK	CS	JLM
$\Delta\partial_{t+1}(\mathcal{L})$	8.324***		-1.800	0.229	-4.183**
	(2.702)		(4.147)	(6.781)	(1.787)
$\Delta\partial_{t+1}(\mathcal{F})$		-4.203			
		(5.411)			
$\mathcal{R}^2$	0.106	0.004	0.002	0.000	0.030

Table VIII: **Stock Index Returns and Contemporaneous Shocks to Dividend Expectations**: This table reports the coefficient estimates and R-square values from regressing stock index returns on contemporaneous shocks to long-run dividend growth rate expectations, computed using our dividend model, the latent variable model in van Binsbergen and Koijen (2010) (vBK), the VAR model in Campbell and Shiller (1988b) (CS), or the Markov-switching model in Johannes, Lochstoer, and Mou (2016), and assuming investors either learn, that is, $\Delta\partial_{t+1}(\mathcal{L})$, or do not learn, that is, $\Delta\partial_{t+1}(\mathcal{F})$, about dividends. Dividends are estimated based on non-overlapping annual data since 1946. Regressions are based on non-overlapping annual data between 1975 and 2016. Reported in parentice are Newey and West (1987) standard errors that account for up to 10 years of serial correlations. Estimates significant at 90%, 95%, and 99% confidence levels are highlighted using $*$, $**$, and $***$.

| | $\mathcal{R}^2$ | p-val. | Incremental | |
			$\mathcal{R}^2_{\mathcal{I}}$	p-val.
Full Info.	0.133	0.017		
Learning about Dividends	0.253	0.001	0.138	0.015
Full Learning	0.271	0.000	0.159	0.009

Table IX: **Stock Index Returns and Expected Returns under Epstein and Zin (1989) Preferences**. This table reports out-of-sample R-square values for predicting stock index returns using our long-run risks model, assuming investors have full information, learn about dividends, or learn about all parameters in our long-run risks model, that is, full learning, and the corresponding bootstrap simulated p-values. Also reported are incremental out-of-sample R-square values for predicting stock index returns assuming learning over assuming full information. Dividends are estimated based on non-overlapping annual data since 1946. Statistics are based on non-overlapping annual data between 1975 and 2016.

| | Incremental | | | |
| | Learning about Dividends | | Full Learning | |
	$\mathcal{R}^2_{\mathcal{I}}$	p-val.	$\mathcal{R}^2_{\mathcal{I}}$	p-val.
Kelly and Pruitt (2013) (1981 - 2009)	0.140	0.045	0.152	0.036
Li, Ng, and Swaminathan (2013) (1995 - 2013)	0.083	0.230	0.069	0.276

Table X: **Long-Run Risks Model and Empirical Proxies of Expected Returns**. This table reports incremental R-square values for predicting stock index returns using our long-run risks model, assuming learning, over the proxies of expected returns in Kelly and Pruitt (2013) or Li, Ng, and Swaminathan (2013), and the corresponding bootstrap simulated p-values. Dividends are estimated based on non-overlapping annual data since 1946. Statistics are based on non-overlapping annual data between 1981 and 2009 for Kelly and Pruitt (2013) and between 1995 and 2013 for Li, Ng, and Swaminathan (2013).

| | Boom | | | | Recession | | | |
| | | | Incremental | | | | Incremental | |
	$\mathcal{R}^2$	p-val.	$\mathcal{R}^2_{\mathcal{I}}$	p-val.	$\mathcal{R}^2$	p-val.	$\mathcal{R}^2_{\mathcal{I}}$	p-val.
Full Info.	0.132	0.029			0.138	0.455		
Learning about Dividends	0.196	0.007	0.074	0.109	0.516	0.085	0.438	0.128
Full Learning	0.191	0.008	0.068	0.024	0.641	0.037	0.584	0.056

Table XI: **Stock Index Returns and Expected Returns under Epstein and Zin (1989) Preferences (Expansion versus Recession)**. This table reports out-of-sample R-square values for predicting stock index returns using our long-run risks model, assuming investors have full information, learn about dividends, or learn about all parameters in our long-run risks model, that is, full learning, and the corresponding bootstrap simulated p-values. Also reported are incremental R-square values for predicting stock index returns assuming learning over assuming full information. Statistics are based on non-overlapping annual data between 1975 and 2016 and are separately reported for expansions versus recessions. A year is in recession if any of its months overlap with NBER recession dates.

	$\mathcal{R}^2$	p-val.	Incremental $\mathcal{R}_{\mathcal{I}}^2$	p-val.
Full Info.	0.118	0.026		
Learning about Dividends	0.144	0.013	0.030	0.276
Full Learning	0.151	0.011	0.037	0.219

Table XII: **Stock Index Returns and Expected Returns under CRRA Preferences**. This table reports out-of-sample R-square values for predicting stock index returns using our CRRA model, assuming investors have full information, learn about dividends, or learn about all parameters in our CRRA model, that is, full learning, and the corresponding bootstrap simulated p-values. Also reported are incremental R-square values for predicting stock index returns assuming learning over assuming full information. Dividends are estimated based on non-overlapping annual data since 1946. Statistics are based on non-overlapping annual data between 1975 and 2016.

	Full Info.	Learning about Dividends	Full Learning
Constant	0.016	0.003	0.008
	(0.034)	(0.013)	(0.015)
Slope	0.902***	1.203***	1.174***
	(0.246)	(0.083)	(0.140)
$\mathcal{R}^2$	0.135	0.254	0.262

	Full Info.	Learning about Dividends	Full Learning
$sy_t(\mathcal{L})$		3.614***	3.895***
		(0.772)	(0.670)
$sy_t(\mathcal{F})$	4.410***		
	(0.103)		
$\mathcal{R}^2$	0.972	0.708	0.661

Table AI: **Stock Index Returns, Long-Run Risks Expected Returns, and Stock Yields**: The table on the top reports estimated coefficients from regressing future stock index returns on expected returns from our long-run risks model, assuming investors have full information, learn about dividends, or learn about all parameters in our long-run risks model, that is, full learning. The table on the bottom reports estimated coefficients from regressing expected returns from our long-run risks model, assuming investors have full information, learn about dividends, or learn about all parameters in our long-run risks model, that is, full learning, on stock yields, assuming investors either learn, that is, $sy_t(\mathcal{L})$, or do not learn, that is, $sy_t(\mathcal{F})$, about dividends. Dividends are estimated based on non-overlapping annual data since 1946. Reported in parentice are Newey and West (1987) standard errors that account for up to 10 years of serial correlations. Estimates significant at 90%, 95%, and 99% confidence levels are highlighted using $*$, $**$, and $***$.

	Quarterly		Semi-Annual		Bi-Annual	
	$\mathcal{R}_{\mathcal{O}}^2$	p-val.	$\mathcal{R}_{\mathcal{O}}^2$	p-val.	$\mathcal{R}_{\mathcal{O}}^2$	p-val.
Our Model	0.502	0.000	0.515	0.000	0.379	0.003
van Binsbergen and Koijen (2010)	0.337	0.000	0.267	0.000	0.151	0.080
Campbell and Shiller (1988b)	0.328	0.000	0.298	0.000	0.262	0.017
Johannes, Lochstoer, and Mou (2016)	0.075	0.079	0.001	0.855	-0.034	1.000
	$\mathcal{R}_{\mathcal{I}}^2$	p-val.	$\mathcal{R}_{\mathcal{I}}^2$	p-val.	$\mathcal{R}_{\mathcal{I}}^2$	p-val.
van Binsbergen and Koijen (2010)	0.212	0.002	0.311	0.000	0.268	0.015
Campbell and Shiller (1988b)	0.222	0.002	0.281	0.000	0.158	0.073
Johannes, Lochstoer, and Mou (2016)	0.435	0.000	0.495	0.000	0.399	0.002

Table AII: **Dividend Growth Rates and Expected Growth Rates (Quarterly, Semi-Annual, and Bi-Annual Rates).** The table on the top reports out-of-sample R-square values for predicting dividend growth rates (quarterly, semi-annual, or bi-annual rates) using our dividend model, the latent variable model in van Binsbergen and Koijen (2010), the VAR model in Campbell and Shiller (1988b), or the Markov-switching model in Johannes, Lochstoer, and Mou (2016), and the corresponding bootstrap simulated p-values. The table on the bottom reports incremental R-square values for predicting dividend growth rates using our model over one of the baseline models. Dividends are estimated based on non-overlapping annual data since 1946. Out-of-sample statistics are based on non-overlapping quarterly, semi-annual, or bi-annual data between 1975 and 2016.

| | | Our Model | | | Baseline Model | |
	$p_t - d_t$	$sy_t(\mathcal{L})$	$sy_t(\mathcal{F})$	vBK	CS	JLM
Quarterly	-0.033***	0.993***	1.101***	0.857***	1.183***	0.384
	(0.009)	(0.201)	(0.284)	(0.232)	(0.327)	(0.239)
$\mathcal{R}^2$	0.032	0.035	0.032	0.026	0.029	0.007
Semi-Annual	-0.065***	1.894***	2.137***	1.686***	2.215***	0.912
	(0.016)	(0.409)	(0.531)	(0.415)	(0.615)	(0.480)
$\mathcal{R}^2$	0.069	0.081	0.072	0.058	0.061	0.022
Bi-Annual	-0.281***	9.312***	9.311***	7.786***	9.063***	4.868**
	(0.073)	(2.216)	(2.341)	(1.986)	(2.663)	(1.884)
$\mathcal{R}^2$	0.258	0.299	0.263	0.231	0.208	0.120

Table AIII: **Stock Index Returns and Stock Yields (Quarterly, Semi-Annual, and Bi-Annual Returns)**: This table reports the coefficient estimates and R-square values from regressing future stock index returns (quarterly, semi-annual, or bi-annual returns) on log price-to-dividend ratios and stock yields, computed using our dividend model, the latent variable model in van Binsbergen and Koijen (2010) (vBK), the VAR model in Campbell and Shiller (1988b) (CS), or the Markov-switching model in Johannes, Lochstoer, and Mou (2016) (JLM), and assuming investors either learn, that is, $sy_t(\mathcal{L})$, or do not learn, that is, $sy_t(\mathcal{F})$, about dividends. Dividends are estimated based on non-overlapping annual data since 1930. Regressions are based on non-overlapping quarterly, semi-annual, or bi-annual data between 1975 and 2016. Reported in parentice are Newey and West (1987) standard errors that account for up to 10 years of serial correlation. Estimates significant at 90%, 95%, and 99% confidence levels are highlighted using $*$, $**$, and $***$.

Quarterly	$\mathcal{R}^2$	p-val.	Incremental $\mathcal{R}_{\mathcal{I}}^2$	p-val.
Full Info.	0.034	0.017		
Learning about Dividends	0.051	0.003	0.018	0.079
Full Learning	0.053	0.003	0.020	0.067
Semi-Annual				
Full Info.	0.070	0.015		
Learning about Dividends	0.155	0.000	0.092	0.005
Full Learning	0.166	0.000	0.103	0.003
Bi-Annual				
Full Info.	0.203	0.039		
Learning about Dividends	0.329	0.006	0.159	0.072
Full Learning	0.405	0.002	0.253	0.019

Table AIV: **Stock Index Returns and Expected Returns under Epstein and Zin (1989) Preferences (Quarterly, Semi-Annual, and Bi-Annual Returns).** This table reports out-of-sample R-square values for predicting stock index returns (quarterly, semi-annual, or bi-annual returns) using our long-run risks model, assuming investors have full information, learn about dividends, or learn about all parameters in our long-run risks model, that is, full learning, and the corresponding bootstrap simulated p-values. Also reported are incremental R-square values for predicting stock index returns assuming learning over assuming full information. Dividends are estimated based on non-overlapping annual data since 1946. Statistics are based on non-overlapping quarterly, semi-annual, or bi-annual data between 1975 and 2016.

| | In-Sample | Out-of-Sample | |
	$\mathcal{R}^2$	$\mathcal{R}^2_{\mathcal{O}}$	p-val.
Our Model	0.319	0.258	0.001
van Binsbergen and Koijen (2010)	0.133	0.102	0.039
Campbell and Shiller (1988b)	0.151	0.137	0.016
Johannes, Lochstoer, and Mou (2016)	0.063	-0.033	1.000

| | Out-of-Sample | |
	$\mathcal{R}^2_{\mathcal{I}}$	p-val.
van Binsbergen and Koijen (2010)	0.174	0.006
Campbell and Shiller (1988b)	0.140	0.014
Johannes, Lochstoer, and Mou (2016)	0.282	0.000

Table AV: **Dividend Growth Rates and Expected Growth Rates (Extended Data Sample).** The table on the top reports R-square values for predicting dividend growth rates using our dividend model, the latent variable model in van Binsbergen and Koijen (2010), the VAR model in Campbell and Shiller (1988b), or the Markov-switching model in Johannes, Lochstoer, and Mou (2016). The first column reports in-sample R-square values. The second and third columns report out-of-sample R-square values and the corresponding bootstrap simulated p-values. The table on the bottom reports incremental R-square values for predicting dividend growth rates using our model over one of the baseline models. Dividends are estimated based on non-overlapping annual data post Great Depression. Out-of-sample statistics are based on non-overlapping annual data between 1975 and 2016.

		Our Model				Baseline 1	Baseline 2	Baseline 3
$p_t - d_t$	-0.130*** (0.035)			-0.008 (0.103)				
$sy_t(\mathcal{L})$		3.676*** (0.633)		2.953 (2.071)	3.928* (2.128)	2.819*** (0.799)	2.842** (1.519)	1.828* (1.116)
$sy_t(\mathcal{F})$			3.014*** (1.127)		-0.666 (1.975)			
R^2	0.136	0.173	0.133	0.154	0.153	0.103	0.083	0.049

Table AVI: **Stock Index Returns and Stock Yields (Extended Data Sample)**: This table reports the coefficient estimates and R-square values from regressing future stock index returns on log price-to-dividend ratios and stock yields, computed using our dividend model, the latent variable model in van Binsbergen and Koijen (2010) (vBK), the VAR model in Campbell and Shiller (1988b) (CS), or the Markov-switching model in Johannes, Lochstoer, and Mou (2016) (JLM), and assuming investors either learn, that is, $sy_t(\mathcal{L})$, or do not learn, that is, $sy_t(\mathcal{F})$, about dividends. Dividends are estimated based on non-overlapping annual data post Great Depression. Regressions are based on non-overlapping annual data between 1975 and 2016. Reported in parentice are Newey and West (1987) standard errors that account for up to 10 years of serial correlation. Estimates significant at 90, 95, and 99 percent confidence levels are highlighted using $*$, $**$, and $***$.

	$\mathcal{R}^2$	p-val.	Incremental $\mathcal{R}^2_{\mathcal{I}}$	p-val.
Full Info.	0.122	0.023		
Learning about Dividends	0.198	0.003	0.086	0.058
Full Learning	0.219	0.002	0.110	0.031

Table AVII: **Stock Index Returns and Expected Returns under Epstein and Zin (1989) Preferences (Extended Data Sample)**. This table reports out-of-sample R-square values for predicting stock index returns using our long-run risks model, assuming investors have full information, learn about dividends, or learn about all parameters in our long-run risks model, that is, full learning, and the corresponding bootstrap simulated p-values. Also reported are incremental R-square values for predicting stock index returns assuming learning over assuming full information. Dividends are estimated based on non-overlapping annual data post Great Depression. Statistics are based on non-overlapping annual data between 1975 and 2016.
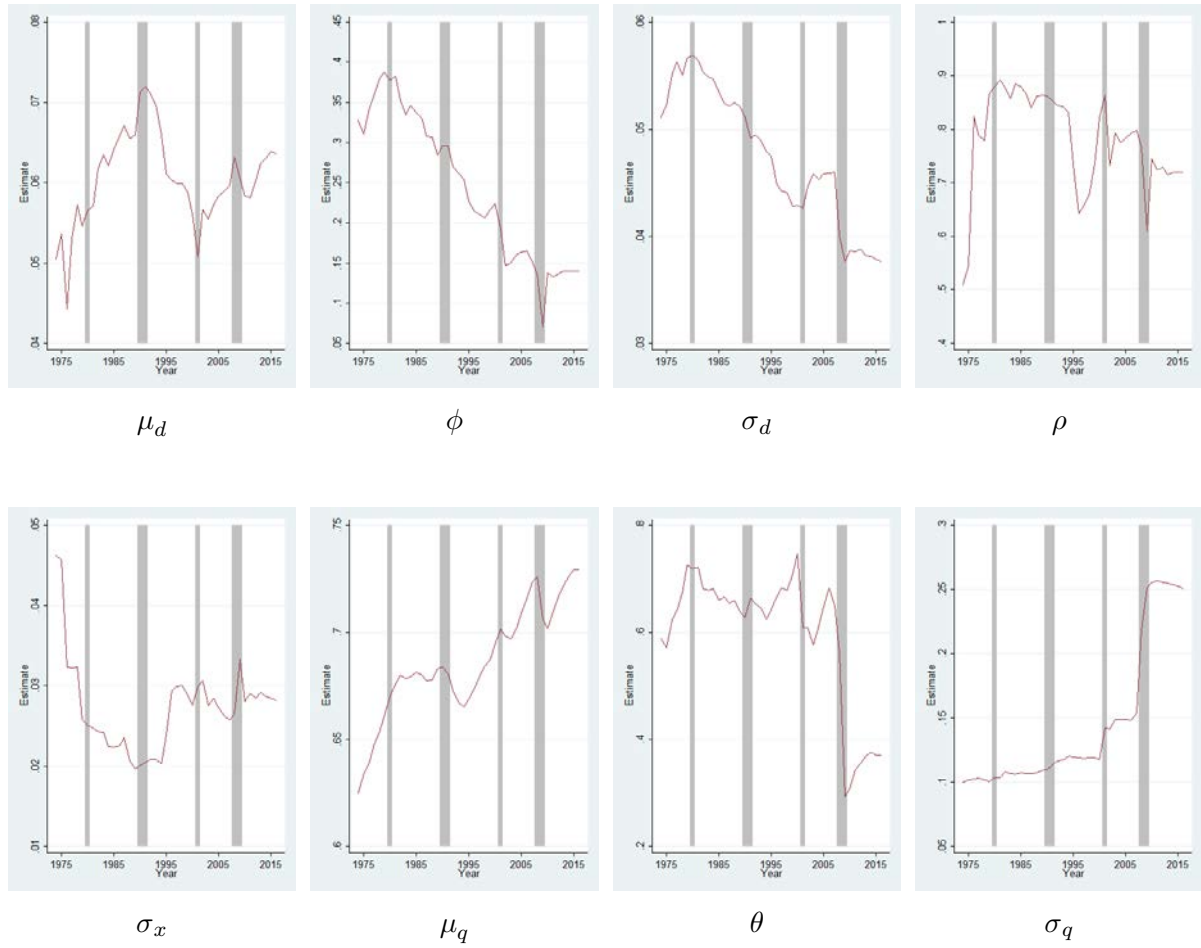
Figure 2: **Evolution of Dividend Model Parameter Estimates Over Time**. This figure plots estimates of the eight parameters in our dividend model, assuming that these parameters are estimated based on data up to time-τ for τ between 1975 and 2016. The shaded regions are recessions. A year is in recession if any of its months overlap with NBER recession dates.
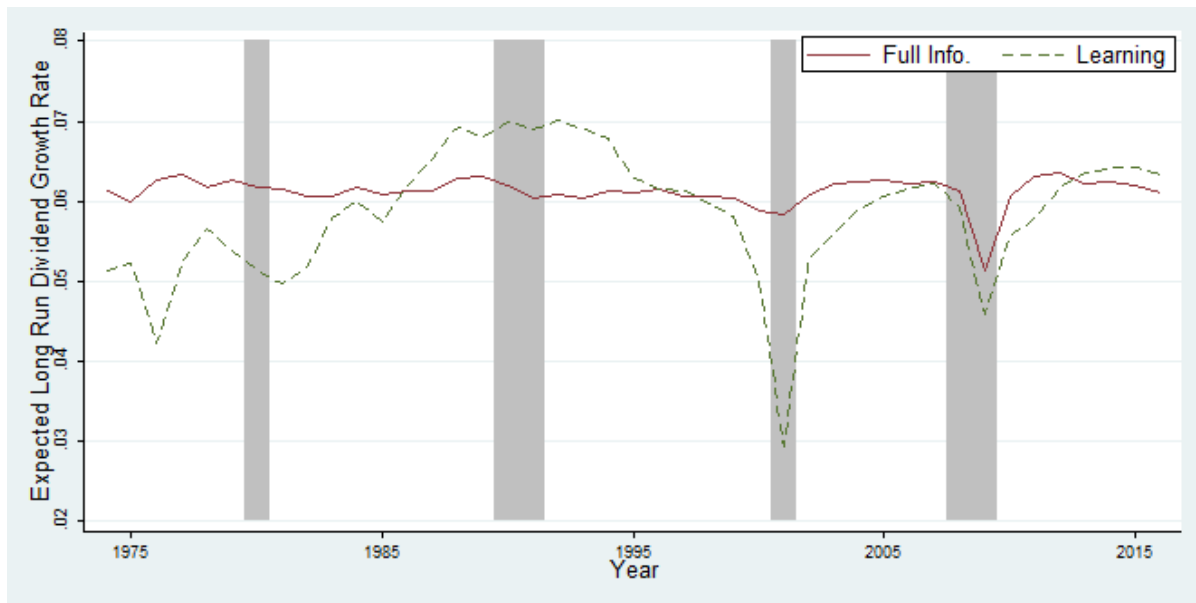
Figure 3: **Expected Long-Run Dividend Growth Rates**. This figure plots long-run dividend growth expectations, computed using our dividend model, for the period between 1975 and 2016. Dividends are estimated based on non-overlapping annual data since 1946. Assuming full information, parameters are estimated once based on the full data sample. Assuming learning, parameters are estimated at each point in time based on data available at the time. The shaded regions are recessions. A year is in recession if any of its months overlap with NBER recession dates.

65
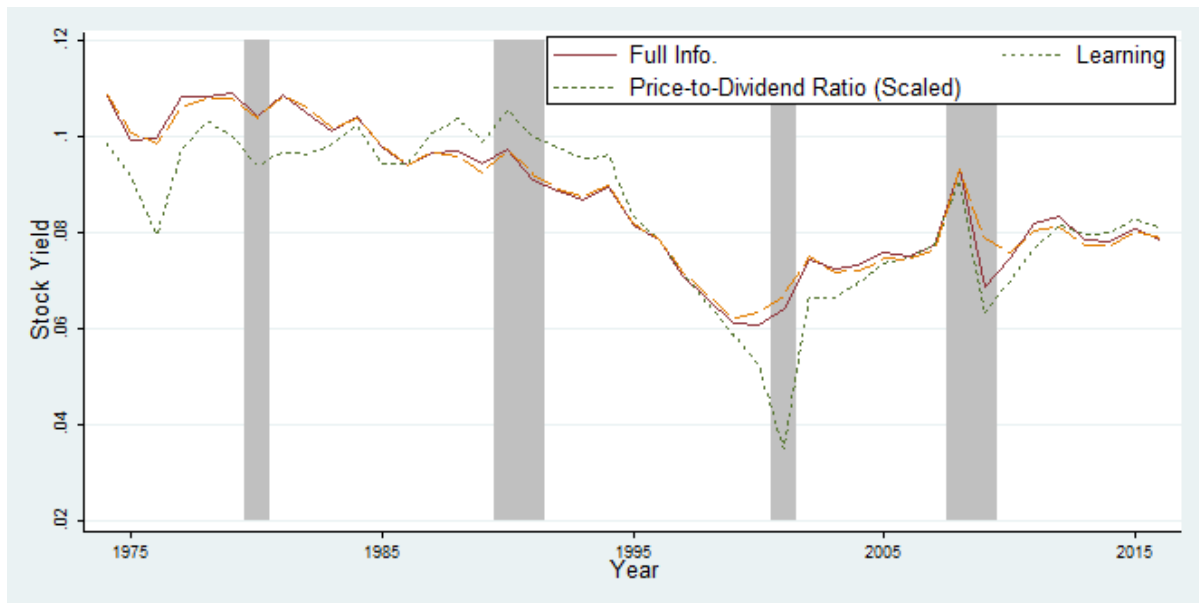
Figure 4: **Stock Yields**. This figure plots stock yields sy_t, computed using our dividend model, and log price-to-dividend ratios (scaled) for the period between 1975 and 2016. Dividends are estimated based on non-overlapping annual data since 1946. Assuming full information, parameters are estimated once based on the full data sample. Assuming learning, parameters are estimated at each point in time based on data available at the time. The shaded regions are recessions. A year is in recession if any of its months overlap with NBER recession dates.
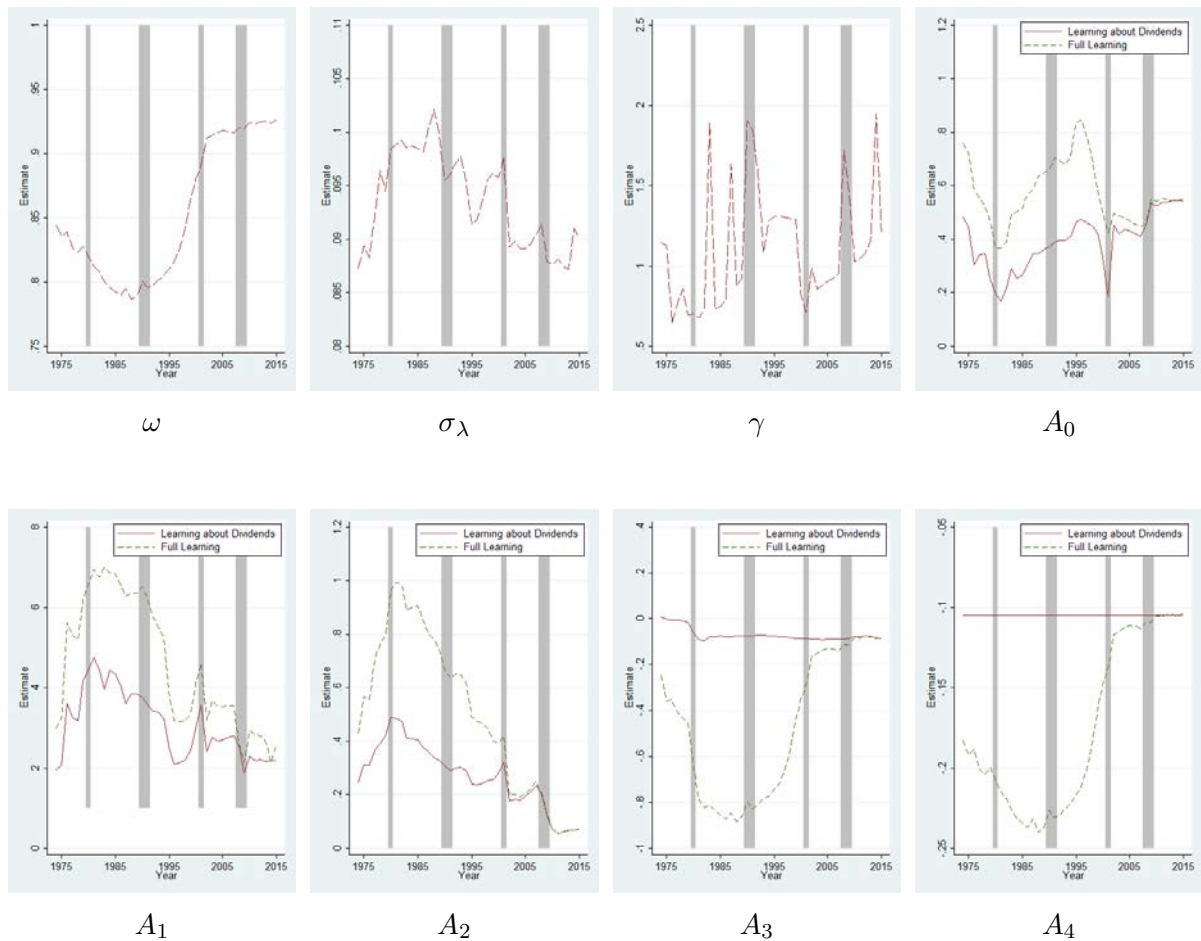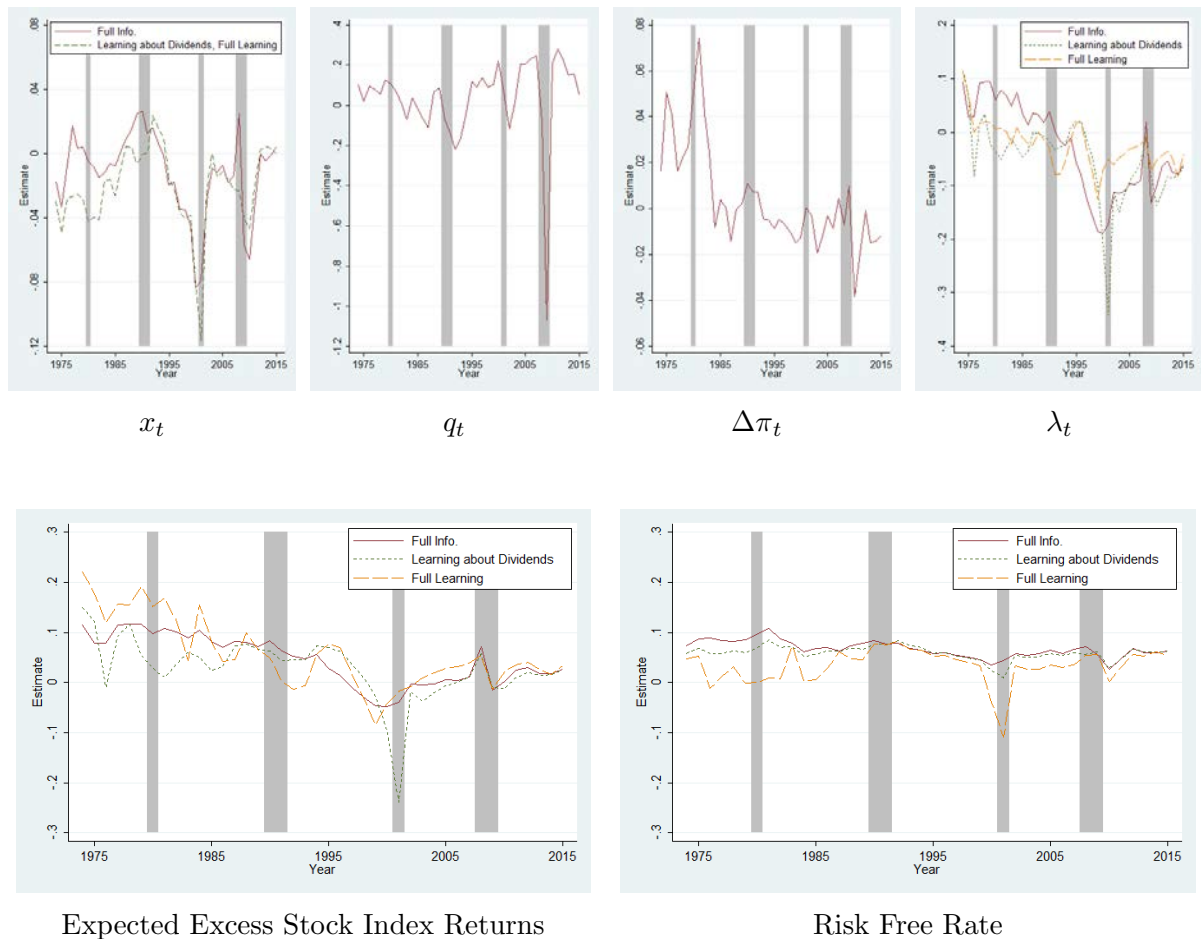
Figure 5: **Evolution of Long-Run Risks Model Parameter and Coefficient Estimates Over Time**. This figure plots estimates of the parameters in our long-run risks model, aside from those in the dividend process, and coefficients $A.$ that relate price-to-dividend ratios, the latent variable x_t, retention ratios, and inflation rates to expected returns, assuming that these parameters are estimated based on data up to time-τ for τ between 1975 and 2016. The shaded regions are recessions. A year is in recession if any of its months overlap with NBER recession dates.

x_t q_t $\Delta \pi_t$ λ_t

Expected Excess Stock Index Returns   Risk Free Rate

Figure 6: **Evolution of Long-Run Risks Model State Variables, Expected Excess Stock Index Returns, and Risk Free Rate Over Time**. This figure plots estimates of the state variables of our long-run risks model, as well as expected excess returns and risk free rate from our model, assuming full information, learning about dividends, or learning about all parameters in our long-run risks model, that is, full learning, between 1975 and 2016. The shaded regions are recessions. A year is in recession if any of its months overlap with NBER recession dates.
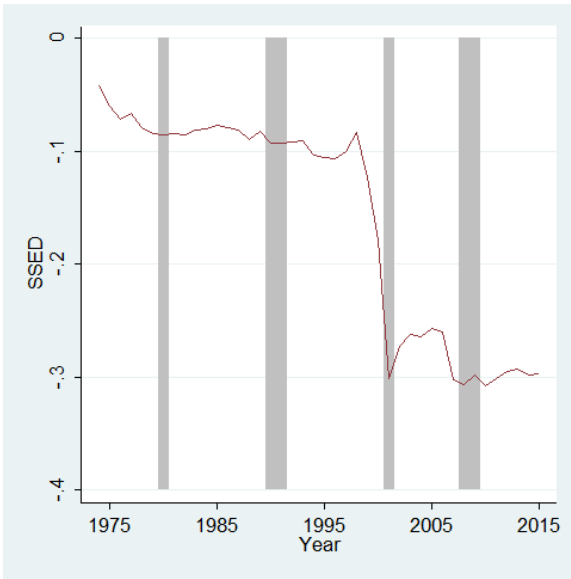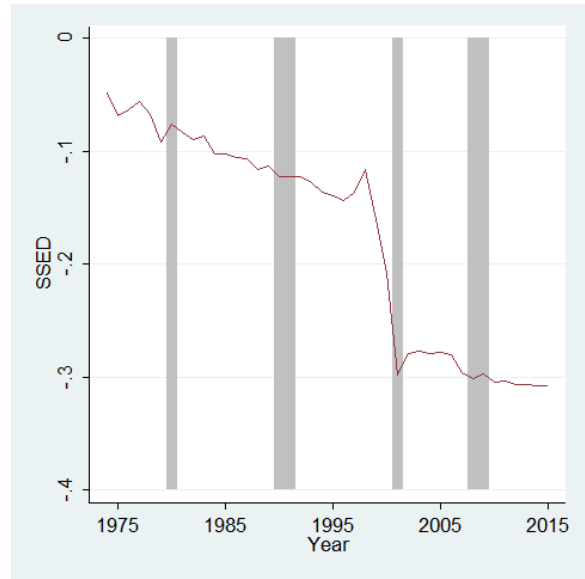
Learning about Dividends   Full Learning

Figure 7: **Cumulative Sum of Squared Errors Difference**. The figure on the left plots the cumulative sum of squared errors difference (SSED) of our long-run risks model, assuming learning about dividends, in predicting stock index returns. The figure on the right plots the SSED of our long-run risks model, assuming learning about all parameters in our long-run risks model, that is, full learning. Dividends are estimated based on non-overlapping annual data since 1946. Statistics are based on non-overlapping annual data between 1975 and 2016. The shaded regions are recessions. A year is in recession if any of its months overlap with NBER recession dates.
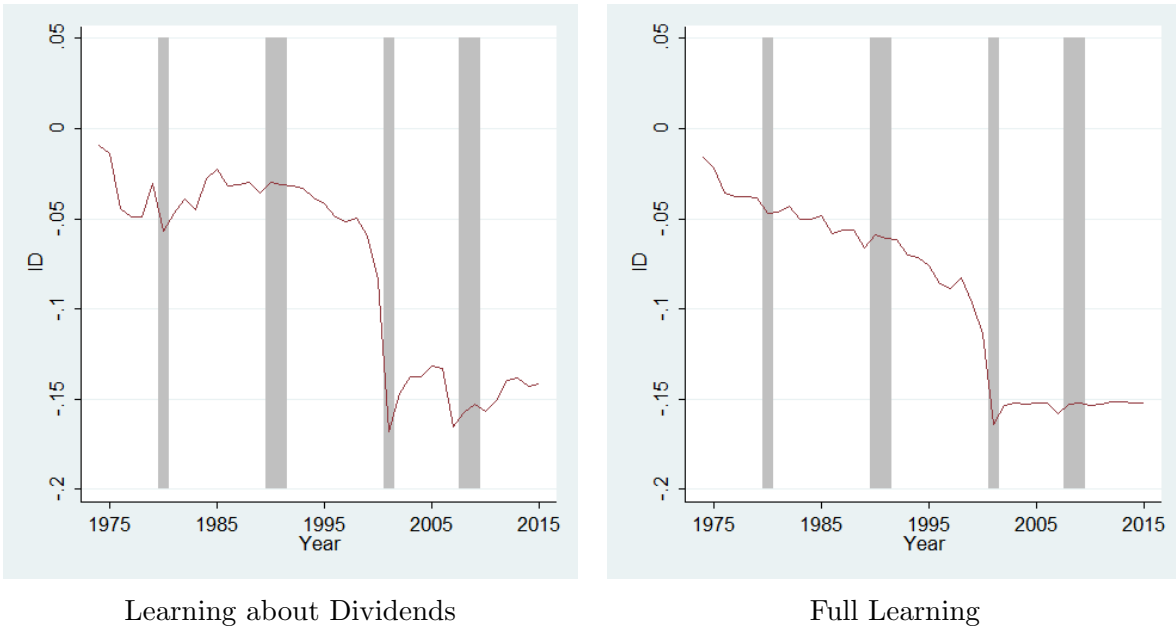
| Learning about Dividends | Full Learning |

Figure 8: **Incremental Gain in Cumulative Sum of Squared Errors Difference from Learning**. The figure on the left plots the incremental gain in the cumulative sum of squared errors difference (SSED) of our long-run risks model, assuming learning about dividends versus full information. The figure on the right plots the incremental gain in SSED of our long-run risks model, assuming learning about all parameters in our long-run risks model, that is, full learning, versus full information. Dividends are estimated based on non-overlapping annual data since 1946. Statistics are based on non-overlapping annual data between 1975 and 2016. The shaded regions are recessions. A year is in recession if any of its months overlap with NBER recession dates.
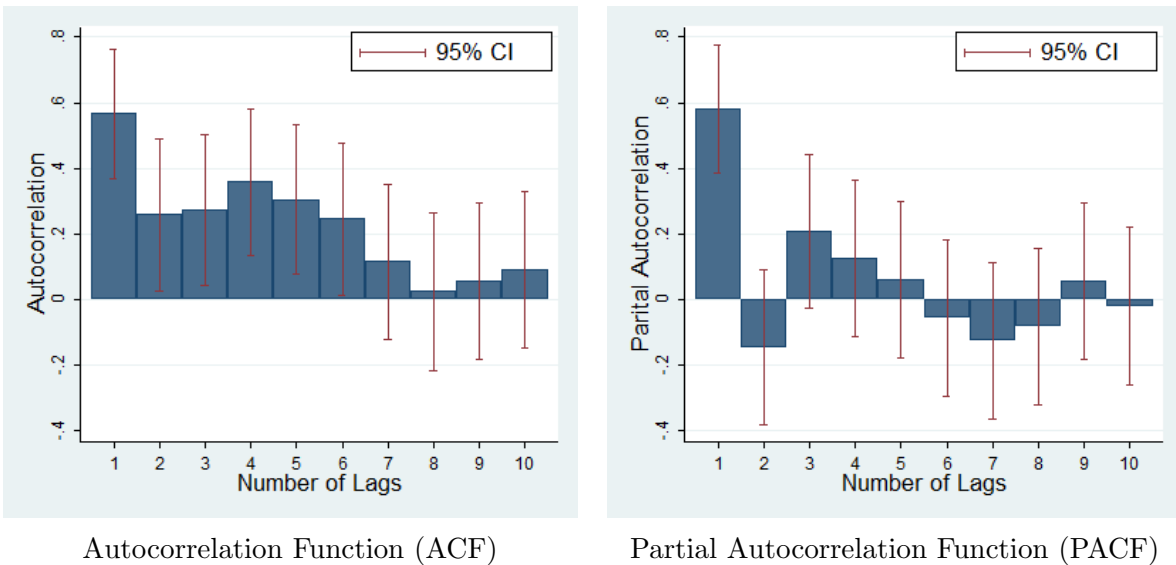

| Autocorrelation Function (ACF) | Partial Autocorrelation Function (PACF) |

Figure A1: **Autocorrelation Function and Partial Autocorrelation Function of Inflation Rate**. This figure plots the autocorrelation function and partial autocorrelation function of inflation rates, up to 10 years lag. Correlations are estimated based on data between 1946 and 2015.

70