Supplemental Appendix

for

Dividend Growth Predictability and the Price-Dividend Ratio

Ilaria Piatti and Fabio Trojani

February 2018

A Relation between State-space model and predictive regressions

In this appendix we first connect the state-space model presented in Section 2 of the paper with its observables vector autoregressive (VAR) representation, as in Cochrane (2008), and then we consider the special cases of constant expected dividend growth and of equal autoregressive coefficients, i.e. $\delta_1 = \gamma_1$.

A.1 General case

In the present-value model there are three potentially independent shocks $\{\varepsilon^g, \varepsilon^\mu, \varepsilon^d\}$. The observable version of this model is equivalent to the ARMA Wold representation for $\{pd_t, \Delta d_t\}$ (or $\{pd_t, r_t\}$), so the shocks have to be linear functions of only the two observable shocks $\{\varepsilon_t^{pd}, \varepsilon_t^d\}$ (see Cochrane (2008)):[1]

$$\begin{aligned} \varepsilon_{t+1}^g &= \alpha_{g,pd}\varepsilon_{t+1}^{pd} + \alpha_{g,d}\varepsilon_{t+1}^d, \\ \varepsilon_{t+1}^\mu &= \alpha_{\mu,pd}\varepsilon_{t+1}^{pd} + \alpha_{\mu,d}\varepsilon_{t+1}^d, \end{aligned} \tag{1}$$

Moreover, we know that the shocks must obey

$$\varepsilon_{t+1}^{pd} = -B_1\varepsilon_{t+1}^\mu + B_2\varepsilon_{t+1}^g. \tag{2}$$

[1]Remind that given the price-dividend ratio, returns or dividend growth can be dropped from the analysis. We can for example predict $\{pd_t, \Delta d_t\}$ and then recover returns from the present-value constraints.

Hence, the weights in (1) have to be such that

$$\begin{aligned} 1 &= -B_1\alpha_{\mu,pd} + B_2\alpha_{g,pd}, \\ 0 &= -B_1\alpha_{\mu,d} + B_2\alpha_{g,d}. \end{aligned} \tag{3}$$

Imposing the constraints, we can parametrize the α weights by

$$\begin{aligned} \varepsilon^g_{t+1} &= \frac{1+\alpha_{pd}}{B2}\varepsilon^{pd}_{t+1} + \frac{\alpha_d}{B2}\varepsilon^d_{t+1}, \\ \varepsilon^\mu_{t+1} &= \frac{\alpha_{pd}}{B1}\varepsilon^{pd}_{t+1} + \frac{\alpha_d}{B1}\varepsilon^d_{t+1}. \end{aligned} \tag{4}$$

To derive VAR representations, we first express the latent state variables in as a function of the observable shocks:

$$\begin{aligned} \hat{\mu}_{t+1} &= \frac{1}{1-\delta_1 L}\varepsilon^\mu_{t+1} = \frac{1}{1-\delta_1 L}\left[\frac{\alpha_{pd}}{B1}\varepsilon^{pd}_{t+1} + \frac{\alpha_d}{B1}\varepsilon^d_{t+1}\right], \\ \hat{g}_{t+1} &= \frac{1}{1-\gamma_1 L}\varepsilon^g_{t+1} = \frac{1}{1-\gamma_1 L}\left[\frac{1+\alpha_{pd}}{B2}\varepsilon^{pd}_{t+1} + \frac{\alpha_d}{B2}\varepsilon^d_{t+1}\right], \end{aligned} \tag{5}$$

where L is the lag operator. Substituting (5) in the expressions for the price-dividend ratio and the realized return we get:

$$\begin{aligned} pd_{t+1} &= A - \frac{1}{1-\delta_1 L}\left[\alpha_{pd}\varepsilon^{pd}_{t+1} + \alpha_d\varepsilon^d_{t+1}\right] + \frac{1}{1-\gamma_1 L}\left[(1+\alpha_{pd})\varepsilon^{pd}_{t+1} + \alpha_d\varepsilon^d_{t+1}\right] \\ &= A + \left(\frac{1+\alpha_{pd}}{1-\gamma_1 L} - \frac{\alpha_{pd}}{1-\delta_1 L}\right)\varepsilon^{pd}_{t+1} + \left(\frac{\alpha_d}{1-\gamma_1 L} - \frac{\alpha_d}{1-\delta_1 L}\right)\varepsilon^d_{t+1}, \end{aligned} \tag{6}$$

and

$$\begin{aligned} \Delta d_{t+1} &= \gamma_0 - \frac{L}{1-\delta_1 L}\left[\frac{1+\alpha_{pd}}{B_2}\varepsilon^{pd}_{t+1} + \frac{\alpha_d}{B_2}\varepsilon^d_{t+1}\right] + \varepsilon^d_{t+1} \\ &= \gamma_0 + \frac{L}{1-\gamma_1 L}\frac{1+\alpha_{pd}}{B_2}\varepsilon^{pd}_{t+1} + \left(1 + \frac{L}{1-\gamma_1 L}\frac{\alpha_d}{B_2}\right)\varepsilon^d_{t+1}, \end{aligned} \tag{7}$$

respectively.

Hence we have the MA representation for the demeaned price-dividend ratio, $\widetilde{pd}_t = pd_t - A$, and dividend growth, $\widetilde{\Delta d}_t = \Delta d_t - \gamma_0$:

$$\begin{aligned} \widetilde{pd}_{t+1} &= \frac{1-[\delta_1 - \alpha_{pd}(\gamma_1-\delta_1)]L}{(1-\delta_1 L)(1-\gamma_1 L)}\varepsilon^{pd}_{t+1} + \alpha_d\frac{(\gamma_1-\delta_1)L}{(1-\delta_1 L)(1-\gamma_1 L)}\varepsilon^d_{t+1}, \\ \widetilde{\Delta d}_{t+1} &= \frac{L}{1-\gamma_1 L}\frac{1+\alpha_{pd}}{B_2}\varepsilon^{pd}_{t+1} + \frac{1-\left(\gamma_1 - \frac{\alpha_D}{B2}\right)L}{1-\gamma_1 L}\varepsilon^d_{t+1}, \end{aligned} \tag{8}$$

or, in vector format:

$$\begin{bmatrix} \widetilde{pd}_{t+1} \\ \widetilde{\Delta d}_{t+1} \end{bmatrix} = \begin{bmatrix} \frac{1-[\delta_1-\alpha_{pd}(\gamma_1-\delta_1)]L}{(1-\delta_1 L)(1-\gamma_1 L)} & \alpha_d\frac{(\gamma_1-\delta_1)L}{(1-\delta_1 L)(1-\gamma_1 L)} \\ \frac{L}{1-\gamma_1 L}\frac{1+\alpha_{pd}}{B_2} & \frac{1-\left(\gamma_1-\frac{\alpha_D}{B2}\right)L}{1-\gamma_1 L} \end{bmatrix} \begin{bmatrix} \varepsilon^{pd}_{t+1}. \\ \varepsilon^d_{t+1} \end{bmatrix}. \tag{9}$$

Multiplying out the denominator lag polynomials, we get a VARMA(2,1) representation:

$$\begin{bmatrix} (1-\delta_1 L)(1-\gamma_1 L) & 0 \\ 0 & (1-\gamma_1 L) \end{bmatrix} \begin{bmatrix} \widetilde{pd}_{t+1} \\ \widetilde{\Delta d}_{t+1} \end{bmatrix} = \left(I - \begin{bmatrix} \delta_1 - \alpha_{pd}(\gamma_1 - \delta_1) & \alpha_d(\gamma_1 - \delta_1) \\ -\frac{1+\alpha_{pd}}{B_2} & \gamma_1 - \frac{\alpha_D}{B2} \end{bmatrix} L \right) \begin{bmatrix} \varepsilon^{pd}_{t+1}. \\ \varepsilon^{d}_{t+1} \end{bmatrix}. \tag{10}$$

Finally, we can invert the MA lag polynomial to find the VAR representation, which describes forecasting regressions.

$$\left(I - \begin{bmatrix} \delta_1 - \alpha_{pd}(\gamma_1 - \delta_1) & \alpha_d(\gamma_1 - \delta_1) \\ -\frac{1+\alpha_{pd}}{B_2} & \gamma_1 - \frac{\alpha_D}{B2} \end{bmatrix} L \right)^{-1} \begin{bmatrix} (1-\delta_1 L)(1-\gamma_1 L) & 0 \\ 0 & (1-\gamma_1 L) \end{bmatrix} \begin{bmatrix} \widetilde{pd}_{t+1} \\ \widetilde{\Delta d}_{t+1} \end{bmatrix} = \begin{bmatrix} \varepsilon^{pd}_{t+1}. \\ \varepsilon^{d}_{t+1} \end{bmatrix}. \tag{11}$$

Let us define

$$\Phi \equiv \begin{bmatrix} \delta_1 - \alpha_{pd}(\gamma_1 - \delta_1) & \alpha_d(\gamma_1 - \delta_1) \\ -\frac{1+\alpha_{pd}}{B_2} & \gamma_1 - \frac{\alpha_D}{B2} \end{bmatrix}.$$

The inverse of $I - \Phi L$, at the left-hand side of (11), is equal to $\sum_{i=0}^{\infty} \Phi^i L^i$. Therefore, the VAR representation of the present-value model has infinite lags. The single coefficients can be solved explicitly and are functions of the structural parameters,[2] but they do not lead to easily interpreted quantities, unless we consider special cases.

A.2 Special cases

Here we consider the special cases of constant expected dividend growth and of equal expectation persistences.

In the case of constant expected dividend growth, the model-implied dividend growth predictability is trivially zero, and the present-value model collapses to a standard linear regression of returns on lagged price-dividend ratio, which follows an AR(1) process:

$$\begin{aligned} \widetilde{pd}_{t+1} &= \delta_1 \widetilde{pd}_t + \varepsilon^{pd}_{t+1} \\ \widetilde{r}_{t+1} &= -\frac{1}{B_1}\widetilde{pd}_t + \varepsilon^{r}_{t+1}, \end{aligned}$$

In the special case of equal autoregressive coefficients, which we denote $\phi \equiv \delta_1 = \gamma_1$, the price-dividend reduces to

$$pd_t = A - B(\hat{\mu}_t - \hat{g}_t), \tag{12}$$

where $B = \frac{1}{1-\rho\phi}$. Substituting the dynamics of expected returns and dividend growth in Equation (12) we find that price-dividend ratio follows an AR(1) process:

$$pd_{t+1} = (1-\phi)A + \phi pd_t + \varepsilon^{pd}_{t+1} \tag{13}$$

[2] Also the α coefficients are functions of the parameters of the structural model which can be found by matching second moments.

and, using (11), demeaned dividend growth, $\tilde{\Delta d} = \Delta d - \gamma_0$, are given by:

$$\widetilde{\Delta d}_{t+1} = \left(\frac{1}{B} + \alpha_{pd}\right)\widetilde{pd}_t + \alpha_d \sum_{i=0}^{\infty} (\phi - \alpha_d)^i (\widetilde{\Delta d}_{t-i} - \left(\frac{1}{B} + \alpha_{pd}\right)\widetilde{pd}_{t-i-1}) + \varepsilon^d_{t+1}, \tag{14}$$

where $\widetilde{pd} = pd - A$ denotes the demeaned price-dividend ratio. The expression for demeaned returns follows from Campbell and Shiller (1988) approximation:

$$\begin{aligned}\tilde{r}_{t+1} &= \rho(\phi\widetilde{pd}_t + \varepsilon^{pd}_{t+1}) + \left(\frac{1}{B} + \alpha_{pd}\right)\widetilde{pd}_t + \alpha_d \sum_{i=0}^{\infty} (\phi - \alpha_d)^i \left[\widetilde{\Delta d}_{t-i} - \left(\frac{1}{B} + \alpha_{pd}\right)\widetilde{pd}_{t-i-1}\right] + \varepsilon^d_{t+1} - \widetilde{pd}_t \\ &= \alpha_{pd}\widetilde{pd}_t + \alpha_d \sum_{i=0}^{\infty} (\phi - \alpha_d)^i \left[\widetilde{\Delta d}_{t-i} - \left(\frac{1}{B} + \alpha_{pd}\right)\widetilde{pd}_{t-i-1}\right] + \varepsilon^r_{t+1}. \end{aligned} \tag{15}$$

where α coefficients are defined as

$$\begin{aligned}\varepsilon^g_{t+1} &= \left(\frac{1}{B} + \alpha_{pd}\right)\varepsilon^{pd}_{t+1} + \alpha_d \varepsilon^d_{t+1}, \\ \varepsilon^\mu_{t+1} &= \alpha_{pd}\varepsilon^{pd}_{t+1} + \alpha_d \varepsilon^d_{t+1},\end{aligned}$$

and are functions of the model parameters which can be found by matching second moments. These can be thought as standard predictive regressions with the addition of a long moving average of errors from the dividend growth regression. Typically, the variation in ex-post dividend growth is much larger than its expected value, thus (15) and (14) can be well approximated (see Cochrane (2008)) by:

$$\begin{aligned}\widetilde{\Delta d}_{t+1} &= \left(\frac{1}{B} + \alpha_{pd}\right)\widetilde{pd}_t + \alpha_d \sum_{i=0}^{\infty} (\phi - \alpha_d)^i \widetilde{\Delta d}_{t-i} + \varepsilon^d_{t+1}, \\ \tilde{r}_{t+1} &= \alpha_{pd}\widetilde{pd}_t + \alpha_d \sum_{i=0}^{\infty} (\phi - \alpha_d)^i \widetilde{\Delta d}_{t-i} + \varepsilon^r_{t+1}.\end{aligned}$$

Thus, estimates of the structural parameters of the model for the case $\delta_1 = \gamma_1$ can be recovered from regressions of returns and dividend growth on lagged price-dividend ratio and on a moving average of past dividend growth.[3] Predictability implied by these adjusted predictive regressions are very similar to what we find using standard predictive regression, with $R^2_{Ret} = 8.7\%$ and $R^2_{Div} = 0.78\%$, compared to $R^2_{Ret} = 8.61\%$ and $R^2_{Div} = 0.01\%$, for the same sample period. In particular, the R-squared for dividend growth is almost zero. Moreover, as expected returns and dividend growth display equal persistence, the EIV-induced asymptotic bias in standard predictive regressions is very moderate (see Figure III). Therefore, the constraint $\delta_1 = \gamma_1$ can be seen as a way of obtaining low dividend predictability and small EIV bias in a setting where expected dividend growth is time-varying and stationary.

[3]Using a 10 year moving average, we find results which are very similar to those we obtain using constrained Kalman filter estimation.

B Consistency with standard predictive regressions

A good specification of the present-value model has to produce predictability features consistent with the empirical evidence of standard OLS regressions for dividend growth and returns.

$$r_{t+1} = a_r + b_r pd_t + \varepsilon^r_{t+1}, \quad (16)$$

$$\Delta d_{t+1} = a_d + b_d pd_t + \varepsilon^d_{t+1}, \quad (17)$$

This feature is essential in order to avoid a model misspecification along some important predictability dimension, which would weaken the interpretation of predictability structures estimated by latent variables approach.

We use a nonparametric bootstrap approach to obtain the distribution of the predictive coefficients b_r and b_d, of their t-statistics and of returns and dividends predictive regression R-squared, implied by the estimated present-value model, both unconstrained and under the constraint of constant expected dividend growth and of equal expectation persistences. The distributions of the t-statistics in the three cases (see Figure I) are almost identical and are all well consistent with results of standard predictive regressions on observed data. The only difference we can note is that the t-statistics of the dividend growth predictive coefficient implied by the present-value model estimated under the constraint of equal persistence parameters is slightly more in line with observed data, since the observed value is included in the interquartile range of the bootstrap distribution. The predictive regression R-squared implied by the estimated models are also consistent with the observed values (see Figure II), even if the true model-implied R-squared can be very different.

The main message is that constrained and unconstrained models cannot be discriminated based on their consistency with standard predictive regression results, even if the estimated dividend growth predictability in the unconstrained present-value setting is very large.

C Expectation persistence and EIV-problem

This appendix studies the link between the relative persistence of the two expectation processes, the severity of the EIV problem and the degree of estimated return and dividend growth predictability.

Consistent with the literature, e.g., Campbell and Shiller (1988), Campbell (1991) and Cochrane (1992), the largest estimated fraction of price-dividend ratio variation is generated by expected return shocks. The loadings B_1 and B_2 of expected dividend growth and expected returns in the price-dividend ratio expression are in a close relation to their persistence features. Therefore, the relative persistence of dividend and return expectations is a key parameter for quantifying the potential EIV problem in the present-value model.

Figure III reproduces graphically this link, by plotting the asymptotic EIV-induced bias for dividend and return predictive regressions, as a function of different hypotheses about the relative persistence, $\delta_1 - \gamma_1$, of dividend and return expectations.[4] While the bias for the return predictive regression coefficient (top panel) is moderate and less than 20% across all values of $\delta_1 - \gamma_1$, the one for the dividend predictive regression coefficient (bottom panel) is very sensitive to differences in the persistence of the two expectations.

The predictability features implied by a present-value model are also strongly dependent on the relative persistence of the unobservable expected returns and expected dividend growth processes. Figure IV shows the R-squared of returns and dividend growth implied by the present-value model described in Section 2 of the paper as a function of the difference between the autoregressive coefficients in the expected returns and cash flow growth dynamics. For low values of the difference between the persistence parameters, the predictability of growth rates implied by the model is almost zero, contrary to what we obtain in Section 3 of the paper. Note that under the null hypothesis of equal expectation persistences:[5]

$$H_0 : \gamma_1 = \delta_1 \ , \tag{18}$$

the model implies an identical sensitivity of price-dividend ratios to return and dividend expectation shocks. Under this constraint, the EIV-induced asymptotic bias for both dividend and return predictive regressions is very small.[6]

Based on asymptotic critical values from a χ^2_r distribution ($r = 1$), the null hypothesis (18) is clearly rejected with a p-value of 0.05%, while the bootstrap p-value is much larger, at 2.4%. Thus, null hypothesis (18) cannot be rejected at a 1% significance level, but it is significantly rejected at the 5% level by our bootstrap testing procedure. Overall, when considering also

[4]The asymptotic EIV-induced bias for dividend and return predictive regressions in the benchmark present-value model can be computed explicitly, as a function of the model parameters; see Supplemental Appendix D.

[5]At the estimated parameters, the difference between the autoregressive parameters in the expected returns and dividend growth dynamics is equal to 0.623. The parameter constraint $\gamma_1 = \delta_1$ was also imposed, e.g., for the present-value models in Cochrane (2008) and Lettau and Van Niewerburgh (2008).

[6]Under null hypothesis (18), pd_t follows a standard AR(1) process; see also Stambaugh (1999) and Lewellen (2004), among others, and Section A of this Supplemental Appendix.

the slightly too liberal behaviour of bootstrap tests in our Monte Carlo simulations,[7] the evidence against a similar persistence of expected returns and expected dividend growth is more ambiguous than under the conventional asymptotic tests.

D Asymptotic EIV Bias in Standard Predictive Regressions

Standard predictive regressions of either returns or dividend growth rates on the lagged log price-dividend ratio suffer from an error-in-variables (EIV) problem, which does not disappear as the sample size increases. Indeed, the true model for aggregate stock returns is:

$$r_{t+1} = \delta_0 + \hat{\mu}_t + \varepsilon^r_{t+1}, \tag{19}$$

but we wrongly assume the following model to hold:

$$r_{t+1} = a_r + b_r pd_t + \tilde{\varepsilon}^r_{t+1}, \tag{20}$$

where $pd_t = A - B_1\hat{\mu}_t + B_2\hat{g}_t$, and we try to estimate the true parameter $b_r = -1/B_1$ from (20). The p-limit of the OLS slope coefficient is the following:[8]

$$\hat{b}_r \longrightarrow \frac{Cov(pd_t, r_{t+1})}{Var(pd_t)}, \tag{21}$$

where

$$\begin{aligned}
Cov(pd_t, r_{t+1}) &= Cov(A - B_1\hat{\mu}_t + B_2\hat{g}_t, \delta_0 + \hat{\mu}_t + \tilde{\varepsilon}^r_{t+1}) \\
&= -B_1 Var(\hat{\mu}_t) + B_2 Cov(\hat{g}_t, \hat{\mu}_t) \\
Var(pd_t) &= B_1^2 Var(\hat{\mu}_t) + B_2^2 Var(\hat{g}_t) - 2B_1B_2Cov(\hat{g}_t, \hat{\mu}_t)
\end{aligned}$$

so that

$$\hat{b}_r \longrightarrow \frac{1}{-B_1 + \frac{B_2^2 Var(\hat{g}_t) - B_1B_2Cov(\hat{g}_t,\hat{\mu}_t)}{B_2Cov(\hat{g}_t,\hat{\mu}_t) - B_1Var(\hat{\mu}_t)}}, \tag{22}$$

and the unconditional variances and covariance of demeaned expected return and dividend growth are the following:

$$Var(\hat{\mu}_t) = \frac{\sigma_\mu^2}{1-\delta_1^2},$$

[7]For null hypothesis (18), we obtain an empirical rejection frequency $\alpha_T^* = 7.5\%$ for the bootstrap test, which is again clearly lower than the rejection frequency $\alpha_T = 26.4\%$ of the asymptotic test. Also in this case, the bootstrap test corrects the asymptotic critical values in the correct direction, even though it is again slightly too liberal.

[8]Note that here we denote with $\hat{b}_r$ the OLS estimate of the slope coefficient b_r in (20).

$$
\begin{aligned}
Var(\hat{g}_t) &= \frac{\sigma_g^2}{1-\gamma_1^2}, \\
Cov(\hat{g}_t, \hat{\mu}_t) &= \frac{\sigma_{g\mu}}{1-\gamma_1\delta_1}.
\end{aligned}
$$

Thus, the OLS slope coefficient in the regression of returns on lagged price-dividend ratio is biased. However, at the estimated parameters the bias is small due to the relative persistence of expected dividend growth and returns.

The model for aggregate log dividend growth is:

$$
\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \varepsilon^d_{t+1}, \tag{23}
$$

while the wrong model is:

$$
\Delta d_{t+1} = a_d + b_d pd_t + \tilde{\varepsilon}^d_{t+1}, \tag{24}
$$

and we try to estimate the true parameter $b_d = 1/B_2$ from (24). The p-limit of the OLS slope is the following:

$$
\hat{b}_d \longrightarrow \frac{Cov(pd_t, \Delta d_{t+1})}{Var(pd_t)}, \tag{25}
$$

where

$$
\begin{aligned}
Cov(pd_t, \Delta d_{t+1}) &= Cov(A - B_1\hat{\mu}_t + B_2\hat{g}_t, \gamma_0 + \hat{g}_t + \tilde{\varepsilon}^d_{t+1}) \\
&= B_2 Var(\hat{g}_t) - B_1 Cov(\hat{g}_t, \hat{\mu}_t)
\end{aligned}
$$

so that

$$
\hat{b}_d \longrightarrow \frac{1}{B_2 + \frac{B_1^2 Var(\hat{\mu}_t) - B_1 B_2 Cov(\hat{g}_t, \hat{\mu}_t)}{B_2 Var(\hat{g}_t) - B_1 Cov(\hat{g}_t, \hat{\mu}_t)}}. \tag{26}
$$

Therefore, the OLS slope coefficient in the regression of dividend growth on lagged price-dividend ratio is also biased. This bias is negative and, at the estimated parameters, much more significant than the one for standard return regressions.

E Reliability of ML estimation

In this appendix we analyse the finite-sample properties of the ML estimator of the present-value model under the null hypotheses of constant expected return (Table I), constant expected dividend growth (Table II) and of equal persistence coefficients (Table III). We generate 1000 samples from the model using nonparametric bootstrap, with the same number of observations as in the data (65 years), starting from the unconditional mean of the state variables and using the point estimates obtained under the null hypothesis considered as the true parameters in the simulations. We subsequently estimate the model for each of the bootstrap samples.[9] We report the true parameters along with the mean, standard deviation and quantiles of the distribution of parameter estimates, obtained from the bootstrap samples. In all cases, the estimation of the persistence parameters and of the volatility of the shocks can be highly biased, due to the very limited size of the sample and to the fact that different parameter structures can lead to very similar properties of the model-implied dividend growth and price-dividend ratio.

For example, in the case of equal persistence parameters (Table III), the autoregressive coefficient of expected returns, δ_1, is slightly downward biased, which corresponds to an upward bias in σ_μ, due to the high negative correlation between these two parameters. The distribution of δ_1 is also left skewed but does not have an excessive standard deviation. On the other hand, the persistence of expected dividend growth, γ_1 is heavily downward biased and has a highly left skewed distribution if the true autoregressive parameters are similar. Due to the high negative correlation between the estimated γ_1 and σ_g, the volatility of the expected dividend growth shock is strongly upward biased. We can also note a very high negative correlation between, σ_g and σ_d, which is intuitive, since the variance of realized dividend growth is divided between expected and unexpected component. Furthermore, the correlation parameters appear to be quite poorly estimated in finite samples. In fact, the confidence intervals for $\rho_{g\mu}$ and $\rho_{\mu d}$ are very broad.[10]

This simple study of the finite-sample properties of the estimated parameters gives some intuition for the poor finite-sample properties of the asymptotic LR test outlined in Section 4.4 of the paper. In fact, the probability of estimating a γ_1 which is equal or lower than the one obtained in the unconstrained estimation of the benchmark model, i.e. 0.304, is about 30% even if the true parameter is equal to 0.926. This potential bias in the estimation of the persistence parameter of expected dividend growth has repercussions on the amount of dividend growth predictability we may find from model estimation.

[9] The same exercise is run in the Internet Appendix to Binsbergen and Koijen (2010) for the unconstrained estimation results. Unreported findings show that we get results similar to what they report starting from our unconstrained parameter estimates.

[10] To make sure that the bias is due to finite-sample issues and not to convergence problems of the optimization algorithm, we also repeat the same exercise simulating longer samples of observations and we show that already with $T = 200$ the median of the distribution of estimated parameters is almost identical to the true parameters, even if the mean is still slightly different due to the skewness of the distribution.

F Estimation Methodology for the Extended Model

Appendix B in the main text describes the estimation methodology for the benchmark model. Here we describe the modifications required to account for the extended version of the model discussed in Appendix F of the paper, which includes additional observable predictive variables.

In the case of the extended model in Appendix, the transition dynamics are the following:

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \gamma_2 \hat{z}_t + \varepsilon^g_{t+1}, \tag{27}$$

$$\hat{\mu}_{t+1} = \delta_1 \hat{\mu}_t + \delta_2 \hat{z}_t + \varepsilon^\mu_{t+1}. \tag{28}$$

The observable variables are dividend growth Δd_{t+1}, the price-dividend ratio pd_{t+1} and an additional observable predictor variable, z_t. Since the equation for the pd ratio contains no error term, as for the benchmark model we can reduce the number of transition equations and we arrive at a final system with one transition equation, (27), and three measurement equations:

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \varepsilon^d_{t+1}. \tag{29}$$

$$\begin{aligned} pd_{t+1} = {} & (1-\delta_1)A + B_2(\gamma_1 - \delta_1)\hat{g}_t + [\gamma_2 B_2 + (\xi_1 - \delta_1)(B_4 - B_3) - \delta_2 B_1]\hat{z}_t + \\ & + \delta_1 pd_t - B_1 \varepsilon^\mu_{t+1} + B_2 \varepsilon^g_{t+1} + (B_4 - B_3)\varepsilon^z_{t+1}. \end{aligned} \tag{30}$$

We use the Kalman filter to derive the likelihood of the model and we estimate it using ML. The parameters to be estimated are the following:

$$\theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_g, \sigma_\mu, \sigma_d, \rho_{g\mu}, \rho_{\mu d}, \rho_{gd}, \xi_0, \xi_1, \rho_{gz}, \rho_{\mu z}, \rho_{dz}, \sigma_z, \delta_2, \gamma_2).$$

For identification purposes, we impose the constraint $\rho_{gd} = 0$, as in Binsbergen and Koijen (2010) and Yun (2012). Overall the model implies 17 free parameters to estimate. The estimation procedure is the following: We first define an expanded 5-dimensional state vector by the concatenation of the original state variable $\hat{g}$ and the process and observation noise random variables:

$$X_t = \begin{pmatrix} \hat{g}_{t-1} \\ \varepsilon^g_t \\ \varepsilon^\mu_t \\ \varepsilon^d_t \\ \varepsilon^z_t \end{pmatrix},$$

which satisfies:

$$X_{t+1} = F X_t + B u_{t+1} + \Gamma \varepsilon^X_{t+1},$$

where $u_t = z_{t-1} - \xi_0$ and

$$\varepsilon_{t+1}^X = \begin{pmatrix} \varepsilon_{t+1}^g \\ \varepsilon_{t+1}^\mu \\ \varepsilon_{t+1}^d \\ \varepsilon_{t+1}^z \end{pmatrix},$$

with conditional variance

$$\Sigma = \begin{bmatrix} \sigma_g^2 & \sigma_{g\mu} & \sigma_{gd} & \sigma_{gz} \\ \sigma_{g\mu} & \sigma_\mu^2 & \sigma_{\mu d} & \sigma_{\mu z} \\ \sigma_{gd} & \sigma_{\mu d} & \sigma_d^2 & \sigma_{dz} \\ \sigma_{gz} & \sigma_{\mu z} & \sigma_{dz} & \sigma_z^2 \end{bmatrix}. \tag{31}$$

Moreover,

$$F = \begin{bmatrix} \gamma_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad B = [\gamma_2 \quad 0_{1\times4}]' \quad \text{and} \quad \Gamma = \begin{bmatrix} 0_{1\times4} \\ I_4 \end{bmatrix},$$

The measurement equation,

$$Y_t = \begin{pmatrix} \Delta d_t \\ pd_t \\ z_t \end{pmatrix},$$

is of the form

$$Y_t = M_0 + M_1 Y_{t-1} + M_2 X_t,$$

where

$$M_0 = \begin{bmatrix} \gamma_0 \\ (1-\delta_1)A \\ \xi_0(1-\xi_1) \end{bmatrix}, \quad M_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \delta_1 & \xi_2 \\ 0 & 0 & \xi_1 \end{bmatrix},$$

$$\xi_2 = \gamma_2 B_2 + (\xi_1 - \delta_1)(B_4 - B_3) - \delta_2 B_1,$$

and

$$M_2 = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ B_2(\gamma_1 - \delta_1) & B_2 & -B_1 & 0 & B_4 - B_3 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The steps of the filter algorithm are exactly as for the benchmark model (see Appendix B in the paper) apart from a slight change in the time-update equation for the state, which becomes:

$$X_{t,t-1} = FX_{t-1,t-1} + Bu_t. \tag{32}$$

G Tables and Figures

Table I: Finite-sample properties of the ML estimator of the present-value model. We generate 1000 samples using nonparametric bootstrap with the same number of observations as in the data (65 years), starting at the unconditional mean of the state variables and using estimated parameters under the constraint of constant expected returns. The model is then estimated for each simulated sample. We report true parameters as well as average, standard deviation and quantiles of the bootstrap distribution of 1000 parameter estimates.

	True	Mean	St.dev	Q10	Q25	Q50	Q75	Q90
γ_0	0.072	0.074	0.018	0049	0.064	0.077	0.082	0.095
δ_0	0.079	0.085	0.017	0.062	0.076	0.086	0.095	0.106
γ_1	0.996	0.950	0.125	0.918	0.956	0.980	0.993	0.999
δ_1	0	0.122	0.248	0	0	0.001	0.093	0.461
σ_g	0.002	0.007	0.009	0.002	0.003	0.005	0.008	0.014
σ_μ	0	0.036	0.020	0.014	0.024	0.034	0.045	0.058
σ_d	0.069	0.066	0.010	0.054	0.060	0.067	0.072	0.077
$\rho_{g\mu}$	0	-0.165	0.585	-0.912	-0.714	-0.200	0.329	0.668
$\rho_{\mu d}$	0	0.626	0.397	0.121	0.433	0.752	0.938	0.988

Table II: Finite-sample properties of the ML estimator of the present-value model. We generate 1000 samples using nonparametric bootstrap with the same number of observations as in the data (65 years), starting at the unconditional mean of the state variables and using estimated parameters under the constraint of constant expected dividend growth. The model is then estimated for each simulated sample. We report true parameters as well as average, standard deviation and quantiles of the bootstrap distribution of 1000 parameter estimates.

	True	Mean	St.dev	Q10	Q25	Q50	Q75	Q90
γ_0	0.055	0.055	0.010	0044	0.048	0.056	0.062	0.067
δ_0	0.082	0.088	0.027	0.066	0.074	0.084	0.094	0.104
γ_1	0	0.268	0.264	0	0.042	0.197	0.383	0.699
δ_1	0.903	0.854	0.087	0.6749	0.811	0.863	0.907	0.954
σ_g	0	0.042	0.025	0.005	0.018	0.049	0.064	0.071
σ_μ	0.028	0.028	0.013	0.014	0.020	0.026	0.034	0.044
σ_d	0.068	0.037	0.025	0.003	0.012	0.041	0.060	0.068
$\rho_{g\mu}$	0	0.081	0.456	-0.703	-0.138	0.191	0.350	0.523
$\rho_{\mu d}$	0.357	0.228	0.422	-0.413	0.085	0.307	0.470	0.697

Table III: Finite-sample properties of the ML estimator of the present-value model. We generate 1000 samples using residual bootstrap with the same number of observations as in the data (65 years), starting at the unconditional mean of the state variables and using estimated parameters under the constraint of equal persistence parameters. The model is then estimated for each simulated sample. We report true parameters as well as average, standard deviation and quantiles of the bootstrap distribution of 1000 parameter estimates.

	True	Mean	St.dev	Q10	Q25	Q50	Q75	Q90
γ_0	0.054	0.054	0.010	0041	0.047	0.054	0.062	0.068
δ_0	0.081	0.083	0.017	0.061	0.070	0.083	0.094	0.106
γ_1	0.926	0.598	0.342	0.107	0.261	0.745	0.917	0.949
δ_1	0.926	0.865	0.068	0.775	0.829	0.876	0.915	0.940
σ_g	0.004	0.026	0.025	0.003	0.005	0.013	0.049	0.066
σ_μ	0.021	0.028	0.011	0.016	0.021	0.027	0.034	0.043
σ_d	0.068	0.049	0.025	0.006	0.036	0.058	0.067	0.073
$\rho_{g\mu}$	0.950	0.351	0.664	-0.875	0.056	0.441	0.946	0.971
$\rho_{\mu d}$	0.312	0.259	0.382	-0.138	0.208	0.309	0.410	0.578

Figure I: Box plots of t-statistics of dividend growth (upper panel) and return (lower panel) standard predictive regression coefficients implied by the estimated present-value model, both unconstrained (1) and under the null hypotheses of constant expected dividend growth (2) and equal expectation persistences (3). Horizontal dotted red line represents the value of the t-statistic on the observed data.

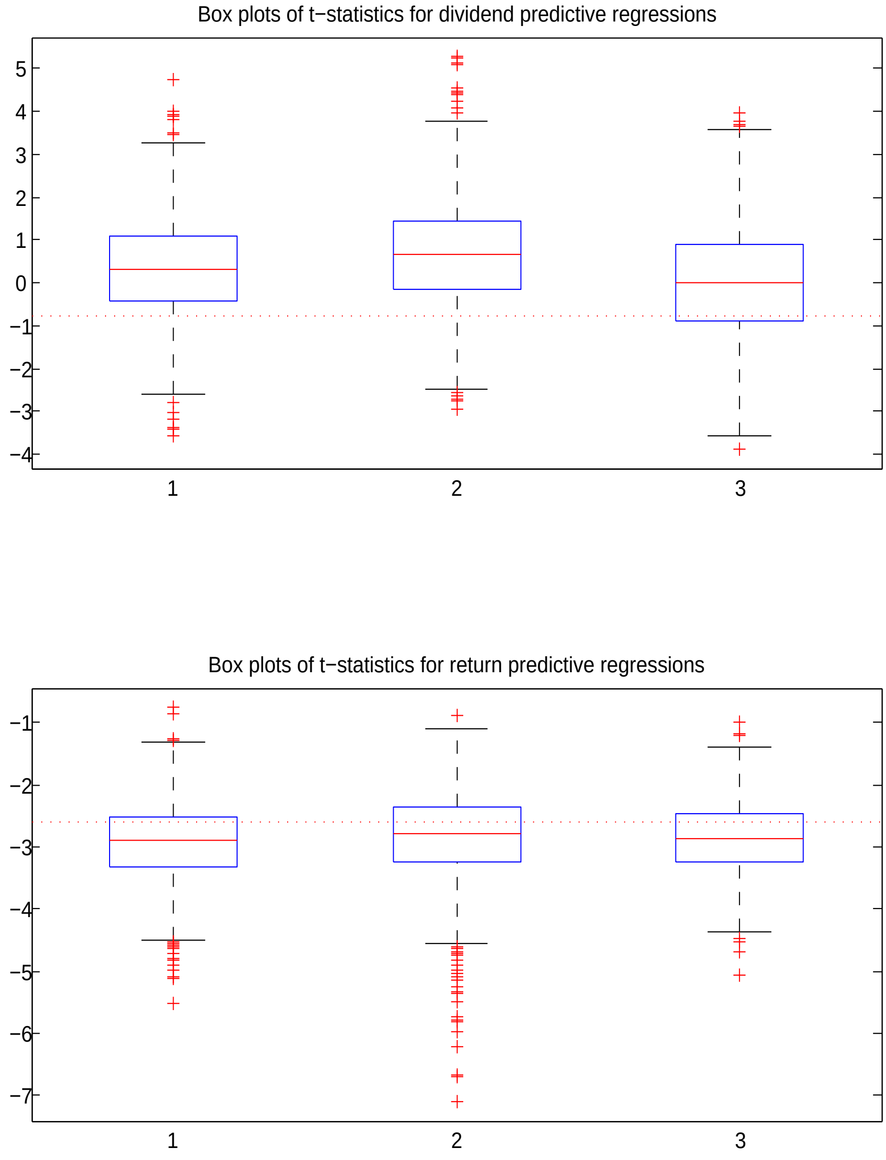


Figure II: The upper (lower) panel represents Box plots of R-squared of dividend growth (return) standard predictive regression implied by the estimated present-value model, both unconstrained (1) and under the null hypotheses of constant expected dividend growth (2) and equal expectation persistences (3). Horizontal dotted red line represents the value of predictive regression R-squared on the observed data.

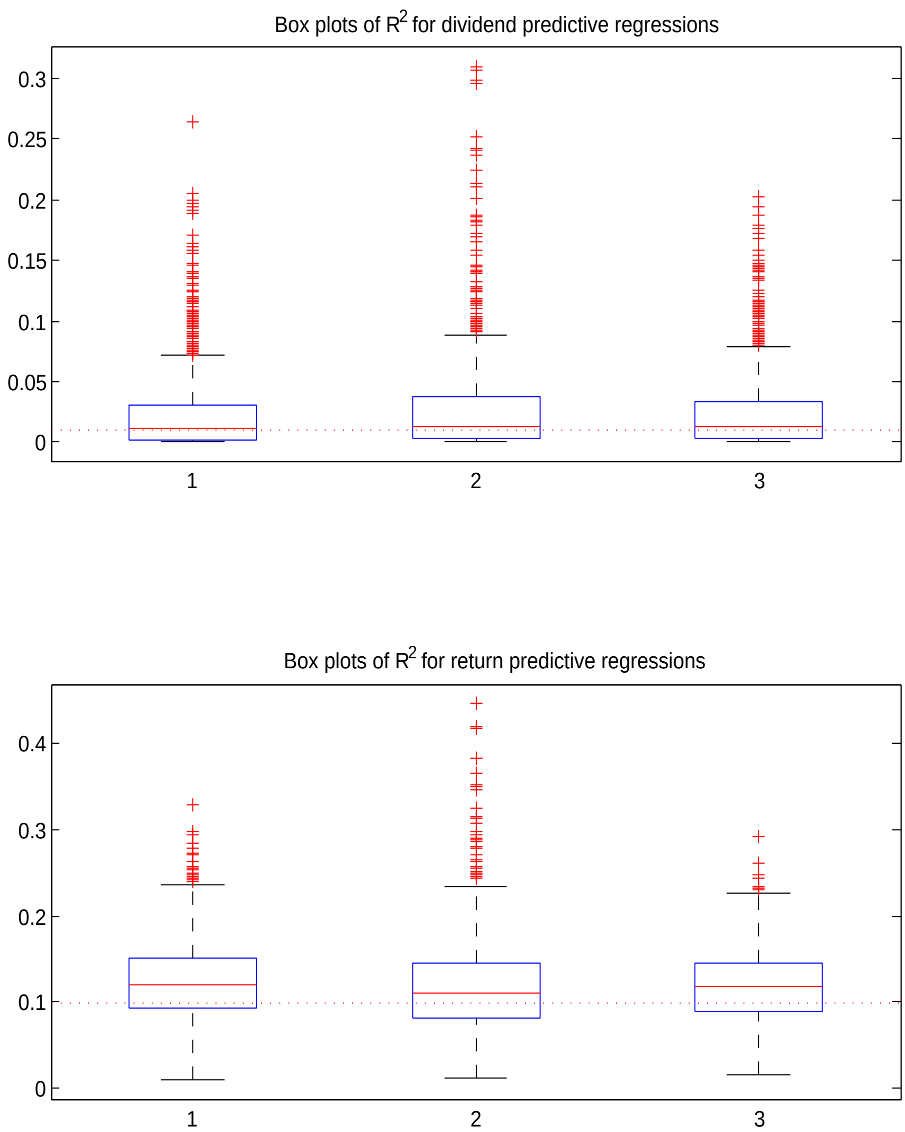


Figure III: Model-implied EIV bias for standard predictive regressions for returns (upper panel) and dividend growth (lower panel) as a function of the difference between the autoregressive coefficients in the dynamics of expected returns (δ_1) and expected dividend growth (γ_1). Solid blue lines denote the true model-implied value of the regression coefficients, $b_r = -1/B_1$ and $b_d = 1/B_2$, while dashed red lines denote the limit of the OLS estimator of b_r and b_d.

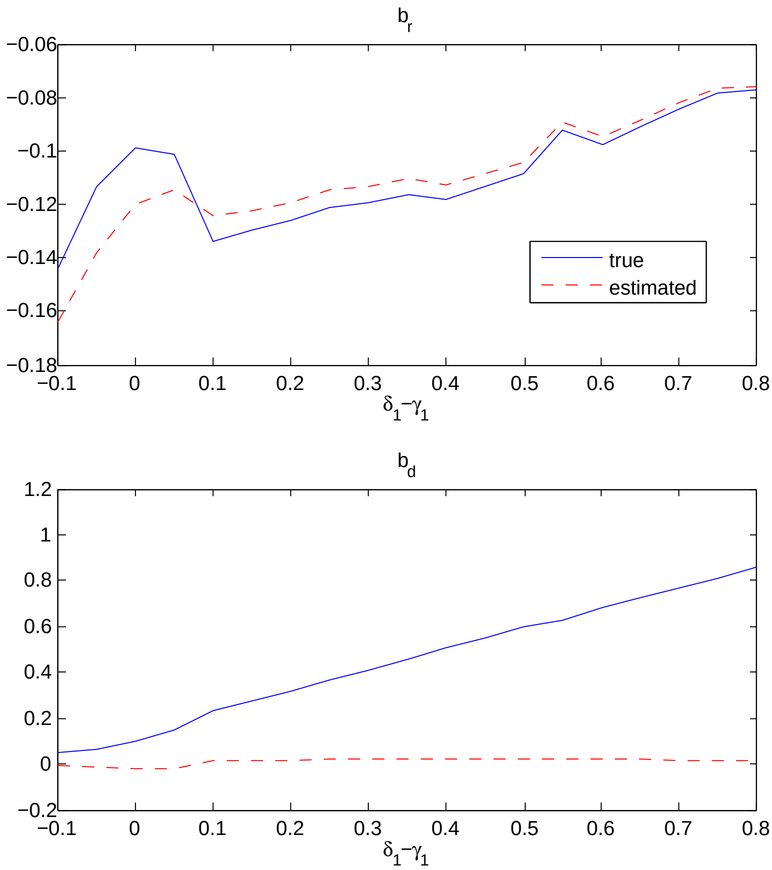


Figure IV: Percentage R-squared of returns (dashed red line) and dividend growth (solid blue line), implied by a simple present-value model as a function of the difference between the autoregressive coefficients in the dynamics of expected returns (δ_1) and expected dividend growth (γ_1). Horizontal lines denote the R-squared of standard predictive regressions of returns (dotted red line) and dividend growth (dash-dotted blue line) on lagged price-dividend ratio for the same sample period, i.e. 1946 to 2010.

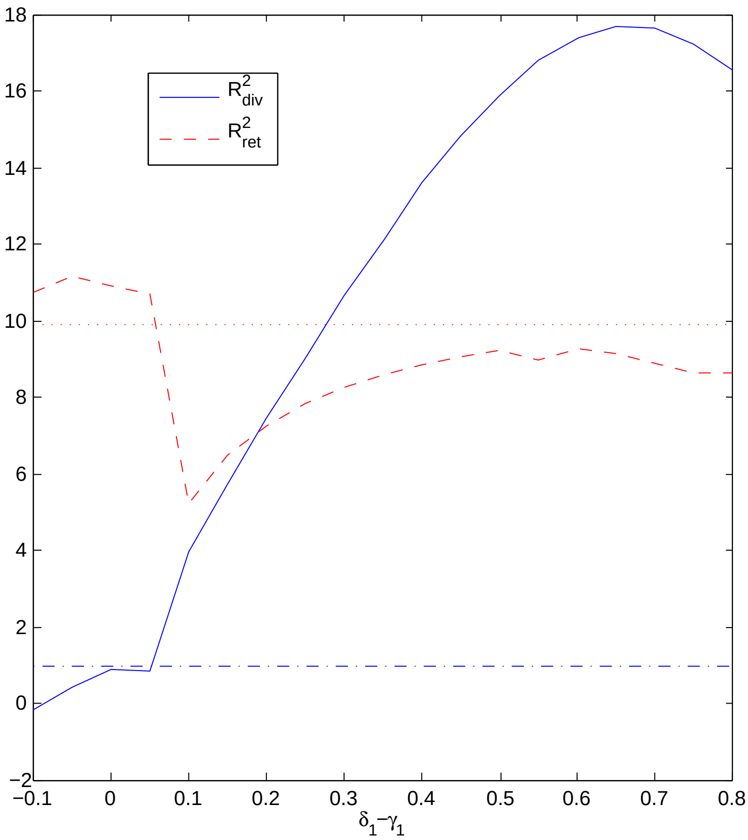


Figure V: This figure refers to Section 4.5 in the main text. The blue histograms denote the bootstrapped distribution of sample skewness of the standardized filtered innovations in the observation equations. For the panels in the first row, model parameters are taken from the unconstrained estimation, while in other rows model parameters are those estimated under the constraints of constant expected dividend growth, constant expected returns and equal persistence, respectively. Panels in the two columns refer to the standardized residuals in the dividend and price-dividend ratio observation equation, respectively. For comparison, the red histograms denote the skewness when standardized innovations are simulated from a standard normal distribution and vertical black lines denote the theoretical skewness of a normal distribution, that is zero.

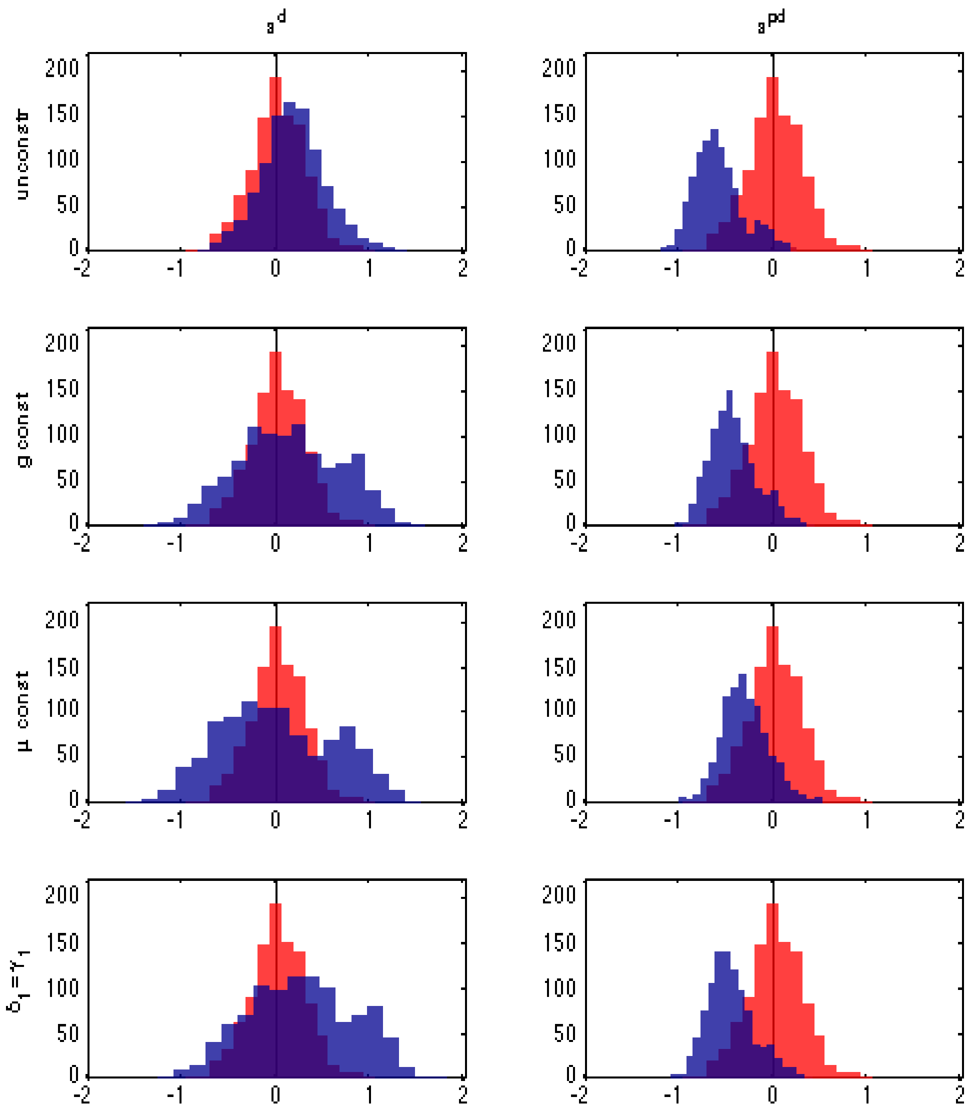


Figure VI: This figure refers to Section 4.5 in the main text. The blue histograms denote the bootstrapped distribution of sample kurtosis of the standardized filtered innovations in the observation equations. For the panels in the first row, model parameters are taken from the unconstrained estimation, while in other rows model parameters are those estimated under the constraints of constant expected dividend growth, constant expected returns and equal persistence, respectively. Panels in the two columns refer to the standardized residuals in the dividend and price-dividend ratio observation equation, respectively. For comparison, the red histograms denote the kurtosis when standardized innovations are simulated from a standard normal distribution and vertical black lines denote the theoretical kurtosis of a normal distribution, that is 3.

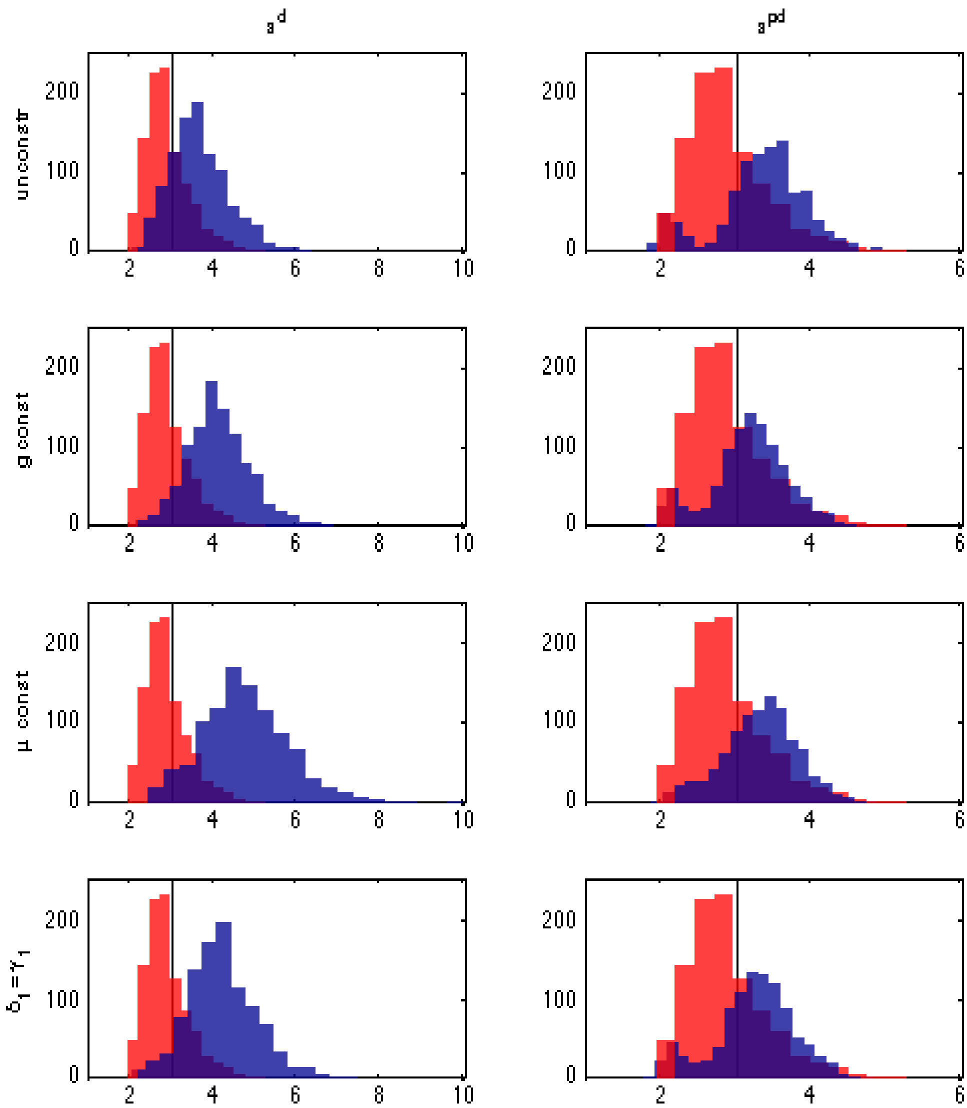
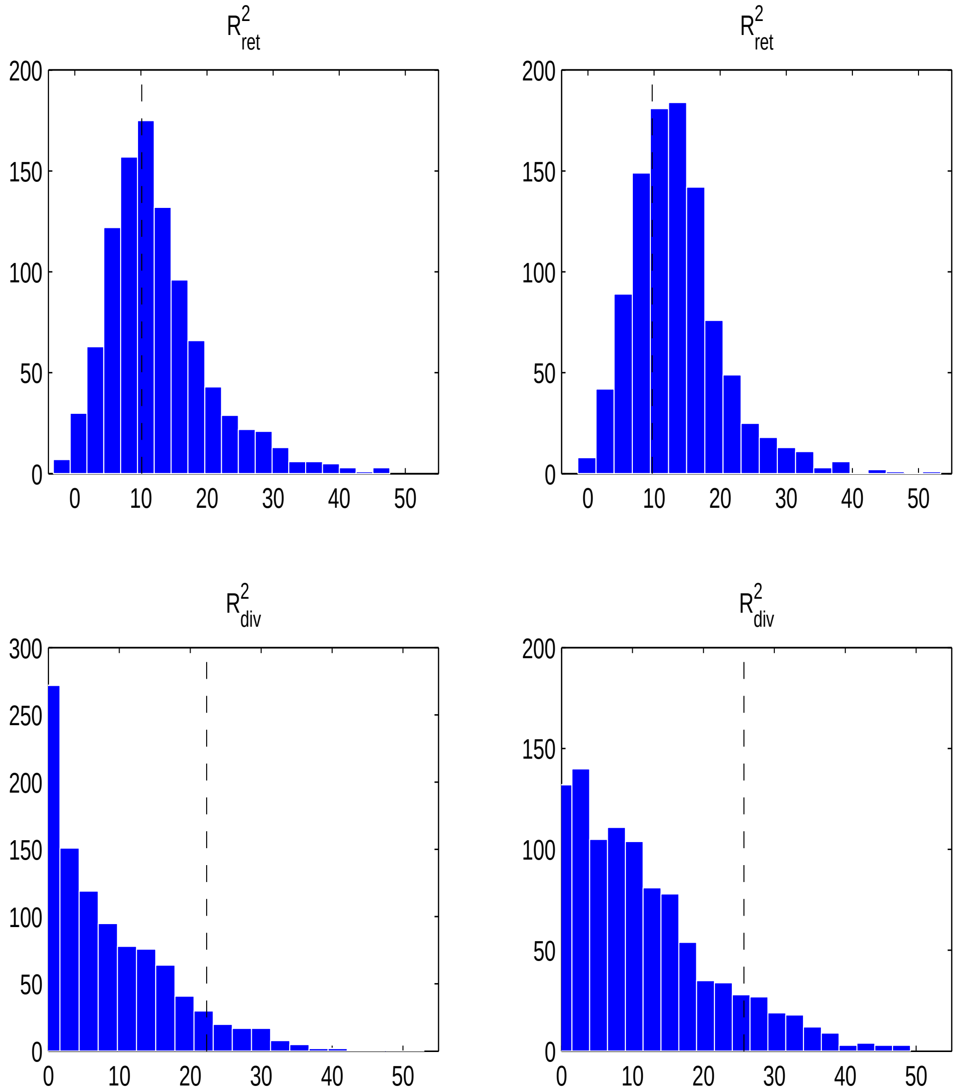


Figure VII: This figure refers to Appendix E in the main text. Bootstrapped distribution of the R-squared of returns (upper panels) and dividend growth (lower panels) implied by the present-value model with additional predictors (*bm* in the left panels and *svar* in the right panels), starting from the estimates under the constraint of constant expected dividend growth ($\gamma_1 = \gamma_2 = \sigma_g = \rho_{g\mu} = \rho_{gz} = 0$). Vertical dashed black lines denote R-squared from unconstrained estimations on real data. Distributions are based on 1000 bootstrap samples.



References

Binsbergen, J. and R. S. Koijen (2010). Predictive regressions: A present-value approach. *Journal of Finance 65*(4), 1439–1471.

Campbell, J. Y. (1991). A variance decomposition for stock returns. *Economic Journal 101*, 157–179.

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*(3), 195–228.

Cochrane, J. H. (1992). Explaining the variance of price-dividend ratios. *Review of Financial Studies 5*(2), 243–280.

Cochrane, J. H. (2008). State-space vs. var models for stock returns. Working Paper. University of Chicago.

Lettau, M. and S. Van Niewerburgh (2008). Reconciling the return predictability evidence. *Review of Financial Studies 21*(4), 1607–1652.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics 74*(2), 209–235.

Stambaugh, R. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375–421.

Yun, J. (2012). Evaluating predictors within a present-value framework. Working Paper. University of Chicago Booth School of Business.