Dividend Growth Predictability and the Price-Dividend Ratio[*]

Ilaria Piatti[†]and Fabio Trojani[‡]

First version: April 2012

This version: February 2013

Abstract

Conventional tests of present-value models tend to over-reject the null of no predictability, concluding that price-dividend ratio variations are due to both cash flow and discount rate shocks. We propose a nonparametric Monte Carlo testing method, which does not rely on distributional assumptions to aggregate the information from the time series of price-dividend ratios and dividend growth. We find evidence of return predictability, but no apparent evidence of dividend growth predictability, thus reconciling the diverging conclusions in the literature. Our findings are robust to the specification of the predictive information set and account for the intrinsic probability of detecting predictive relations by chance alone.

Keywords: Predictability, Predictive regression, Present-value model, State-space model, Bootstrap, Likelihood ratio test.

[†]University of Lugano, Via Buffi 13, CH-6900 Lugano, Switzerland; e-mail: `ilaria.piatti@usi.ch`

[‡]University of Lugano and Swiss Finance Institute, Via Buffi 13, CH-6900 Lugano, Switzerland; e-mail: `fabio.trojani@usi.ch`

1 Introduction

Are stock market returns and dividend growth predictable? Campbell and Shiller (1988) observation that the price-dividend ratio reflects information on both future expected returns and expected dividend growth has motivated a vast literature that studies predictability features based on predictive regressions of returns and cash flow growth on lagged dividend yields.

Predictive regression results typically imply some economically significant evidence of return predictability, even if the statistical significance is weaker in some subperiods, and an almost constant expected dividend growth. This evidence suggests that the price-dividend ratio varies mainly because of discount rate shocks; See Campbell (1991) and Cochrane (1992), among others.[1] In contrast, the Kalman-Bucy filter estimation of a benchmark present-value model with hidden dividend and return expectations yields both a predictable return and a predictable dividend growth, indicating that the price-dividend ratio varies because of both dividend expectation and discount rate shocks; see, e.g., Binsbergen and Koijen (2010).

In this paper, we introduce a new method with reliable finite-sample accuracy and valid asymptotic properties, for testing general predictability hypotheses in present-value models. Our approach is based on a nonparametric Monte Carlo bootstrap, which avoids distributional assumptions in aggregating the information from the time series of price-dividend ratios and dividend growth, and allows us to study more sharply the diverging predictability implications emerging from present-value models and standard predictive regression settings.

We show that while conventional testing procedures can imply significant finite-sample biases that over-reject the null of no predictability, our testing method produces more reliable finite-sample inferences. Applying our testing methodology to benchmark present-value models, which are designed to parsimoniously aggregate dividend growth and price-dividend ratio information, we find a significant evidence of return predictability, but no

---

[1]Early predictive regression studies are Rozeff (1984), Schiller (1984), Keim and Stambaugh (1986), Campbell and Shiller (1988) and Fama and French (1988). The predictive regression findings of no dividend predictability also depend on the sample period used for estimation, as the conclusions are opposite for the pre-war sample (Chen (2009)).

evidence of dividend predictability, thus reconciling the diverging predictability conclusions in the literature.

Inference on return and dividend predictability in standard predictive regressions is difficult for a number of reasons. First, the large correlation between stock returns and predictive variables, combined with the high persistence of the latter, can create finite-sample biases and a non-standard asymptotic behaviour for common tests of return or dividend predictability.[2] Second, as the dividend yield reflects expectations of both future stock returns and future cash flows, it is a noisy estimate of expected returns and expected dividend growth in univariate predictive regressions, thus creating a standard error-in-variable (EIV) problem; see, e.g., Binsbergen and Koijen (2010), among others. Third, powerful tests of predictability need to incorporate the fact that they test a joint null hypothesis on the return-dividend process. Cochrane (2008a) stresses the fact that return and dividend growth predictability have to be studied jointly, concluding that the weak evidence of return predictability in earlier univariate studies is stronger if one jointly considers the empirical evidence on dividend predictability.[3]

Standard predictive regressions are agnostic about the hidden economic link between price-dividend ratios, expected returns and expected cash flow growth, which is instead explicitly revealed within present-value models. These models offer a convenient framework to jointly estimate expected returns and dividend growth, while taking into account the joint no-arbitrage constraints on stock returns, cash flows and valuation ratios. Recent studies estimating market expectations for returns and dividends with different present-value models include Menzly, Santos, and Veronesi (2004), Lettau and Ludvigson (2005), Ang and Bekaert (2007), Lettau and Van Niewerburgh (2008), Campbell and Thompson

---

[2]Stambaugh (1999) derives the finite-sample distribution of the predictive regression parameter estimates and finds that the associated t-statistic is biased towards rejection of the null of no predictability. Torous, Valkanov, and Yan (2004) and Campbell and Yogo (2006) study asymptotic testing frameworks in predictive regression settings with nearly integrated predictors.

[3]Given the observed time variation in the price-dividend ratio, at least one between returns and dividend growth must be predictable. Cochrane (2008a) also derives upper bounds on price-dividend ratio autocorrelation, to deliver more powerful statistics in the joint testing of return and dividend growth predictability. In a present-value approach, such constraint is explicitly incorporated given the joint dynamics of the expectation processes.

(2008), Pastor, Sinha, and Swaminathan (2008), Cochrane (2008a,b), Binsbergen and Koijen (2010), and Rytchkov (2012), among others.

Consistent with the literature, the Kalman-Bucy filter maximum likelihood estimation of a benchmark present-value model on post-war US stock market data, along the lines of, e.g., Binsbergen and Koijen (2010), yields different predictability implications than standard predictive regressions. In the sample period from January 1946 to December 2010, our point estimates imply predictable returns and dividend growth rates, a quite large fraction (18%) of future dividend variability explained by dividend expectations and a much lower fraction (9%) of future returns explained by return expectations. These findings are supported by the evidence produced in a standard likelihood ratio test of the null hypotheses of constant expected returns or expected dividend growth, which are both clearly rejected at significance levels below 0.5%. Expanding the predictive information set to include additional predictive variables for return and dividend growth expectations further strengthens this evidence; see Yun (2012).

Simple univariate predictive regressions with the lagged price-dividend ratio as a predictive variable imply R^2's of 9.9% and 0.95% for market returns and aggregate dividend growth, respectively. Moreover, while the null hypothesis of no dividend predictability is not rejected, the return predictive regression test implies a significant predictability evidence with a p-value of 1.13%. What drives the diverging predictability implications between benchmark present-value models and predictive regressions?

A possible explanation is the EIV-bias inherent to predictability studies.[4] This paper focuses on a different explanation, which follows from the peculiar finite-sample properties of estimators and tests in present-value models with latent return and dividend expectations. While the finite-sample properties of tests of predictability hypotheses in standard predictive regressions have been studied in detail,[5] they have not yet been thoroughly

---

[4]While the estimation results imply a small asymptotic EIV-bias in the return predictive regression, they are linked to a large negative EIV-bias in the slope parameter of the dividend predictive regression, with the limit of the slope coefficient being equal to 0.0197, much lower than the model-implied parameter value of 0.7038.

[5]Stambaugh (1999) derives an analytic expression for the bias in univariate predictive regressions. Kothari and Shanken (1997), Amihud and Hurvich (2004), Lewellen (2004) Torous, Valkanov, and Yan (2004), Campbell and Yogo (2006) and Polk, Thompson, and Vuolteenaho (2006) develop methods for

studied in present-value models estimated with a latent-variables approach.

Similar to predictive regressions, in state-space models inference about relevant hypotheses is made tractable by the existence of an asymptotic theory, which under appropriate conditions implies consistency and asymptotic normality of parameter and latent state estimates; see, e.g., Liung and Caines (1979) and Spall and Wall (1984). However, the short or moderate length of time-series data available in many predictability studies can make the use of asymptotic inference methods potentially suspect for latent variable approaches in present-value models as well.[6]

To improve over the conventional asymptotic inference, a useful nonparametric method, which does not rely on strong assumptions about the joint distribution of dividend growth and price-dividend ratios, is the nonparametric Monte Carlo bootstrap first suggested by Efron (1979). Stoffer and Wall (1991) prove that the bootstrap applied to the innovations of a time-invariant and stable state-space model yields asymptotically correct results. They also demonstrate, by Monte Carlo simulation and in a number of real-data applications, that a bootstrap approach can improve over the finite-sample inference of conventional asymptotics.

We borrow from these insights and propose a novel class of bootstrap likelihood ratio tests of predictability hypotheses. We show their asymptotic validity and demonstrate the improved finite-sample properties over the conventional asymptotics. Using the new testing methodology, we obtain novel findings and interpretations for the predictability evidence obtained by latent variable approaches within present-value models. The more detailed contributions to the literature are the following.

First, in order to study the finite-sample properties of tests of predictability hypotheses in present-value models, without assuming a particular error distribution, such as, e.g., a normal distribution, we introduce a simple nonparametric Monte Carlo simulation

---

hypothesis testing in univariate settings. Amihud, Hurvich, and Whang (2009) propose an analytic method for hypothesis testing in regression with multiple predictors, while Lettau and Ludvigson (2001) and Ang and Bekaert (2007), among others, use bootstrap methods in this setting.

[6]The close relation between present-value models and their (VAR) reduced-form predictive regression representations (see, e.g., Cochrane (2008b)) also suggests that if samples must be fairly large before asymptotic theory is applicable, then this should similarly hold both in predictive regressions and present-value models. See also Section A of the Supplemental Appendix.

approach. Our simulations show that asymptotic likelihood ratio tests imply large finite-sample biases that often lead to an incorrect rejection of the null of no predictability. For instance, while according to the asymptotic chi-square distribution and a significance level $\alpha = 5\%$ the critical value of the asymptotic test of no dividend (return) predictability is 7.81 (9.49), the Monte-Carlo finite-sample critical value is 17.13 (15.99). Overall, the fraction of incorrect rejections of the null of no time-variation in dividend (return) expectations using the asymptotic test can be as large as 25.8% (60.5%). As a result, the evidence resulting from asymptotic tests needs to be taken with caution, because it could be generated by chance alone, using the typically moderate sample sizes available in many predictability studies.

Second, the Monte Carlo evidence shows that large estimated R^2's for dividends or returns can arise by chance alone, even under the null of constant expected dividend growth or expected return. These features are linked to the volatile point estimates for the persistence of dividend and return expectations, which is estimated unprecisely in finite samples and stays in a close relation to the estimated R^2's. This evidence stresses the importance of combining a pure estimation approach with a reliable testing method, when quantifying the actual degree of dividend or return predictability.

Third, we propose a general nonparametric likelihood ratio test of predictability, by applying the bootstrap to the innovations from the latent state dynamics, generated under the relevant null hypothesis. We prove that our bootstrap likelihood ratio test implies an asymptotically valid inference, under standard conditions, and demonstrate by Monte Carlo simulation that it improves in finite samples over the conventional asymptotic tests. Overall, these findings indicate that the bootstrap testing approach can better control the finite-sample probability of rejecting a null hypothesis because of chance alone, thus producing a more reliable predictability evidence in a number of applications.

Fourth, we apply our bootstrap test to post-war US stock market data, using several specifications of the predictive information set. Overall, we find evidence in favour of time-varying expected returns, but no significant evidence against a constant expected dividend growth. While these findings are different from those of conventional tests, they suggest that the dividend predictability of present-value models aggregating dividend growth and price-dividend ratio information is similarly weak as for standard predictive

6

regressions, thus reconciling the diverging conclusions in the recent literature.

Finally, we propose a modification of our bootstrap testing method that is useful to test the actual degree of out-of-sample predictability, while controlling the probability of detecting predictive relations by chance alone. Also in this context, our tests indicate that the larger estimated out-of-sample R-square's for dividends in the data can arise by chance alone, under the null of constant dividend growth expectations.

The paper proceeds as follows. Section 2 introduces the benchmark present-value model for aggregate dividends and market returns. It then briefly discusses the data and the estimation strategy, before reporting the standard estimation results. Section 3 studies the finite-sample biases of conventional asymptotic likelihood ratio tests, while Section 4 introduces the bootstrap likelihood ratio testing approach and studies the resulting improvements in finite-sample inference. Section 5 analyses the robustness of our main predictability findings with respect to broader specifications of the predictive information set, while Section 6 concludes.

2 Present-Value Approach

Borrowing from Binsbergen and Koijen (2010), we introduce the benchmark cash flow and discount rate dynamics. This model offers a tractable framework to estimate the expected return and expected dividend growth processes, by parsimoniusly aggregating the time-series information from dividend growth and price-dividend ratios. Even though the benchmark model restricts the information set to be spanned by the history of dividend (or returns) and price-dividend ratios, it is flexible enough to capture the essential aspects related to the estimation and testing of predictive relations.[7] Broader specifications of the predictive information set are studied in Section 5.

---

[7]The same setting can result from a general equilibrium framework with multiple securities and time-varying risk aversions; see, e.g., Menzly, Santos, and Veronesi (2004)). Recent studies have investigated predictability in the context of the model considered in this paper, including Cochrane (2008b), Binsbergen and Koijen (2010), and Rytchkov (2012), among others. Model extensions and different special cases have also been considered in Lettau and Ludvigson (2005), Ang and Bekaert (2007), Lettau and Van Niewerburgh (2008), Campbell and Thompson (2008), Pastor, Sinha, and Swaminathan (2008), and Yun (2012).

2.1 The benchmark model

Let

$$r_{t+1} \equiv \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) \tag{1}$$

be the cum-dividend log market return and denote by

$$\Delta d_{t+1} \equiv \log\left(\frac{D_{t+1}}{D_t}\right), \tag{2}$$

the aggregate log dividend growth. Expected dividend growth and return, conditional on the information at time t, are denoted by $g_t \equiv E_t[\Delta d_{t+1}]$ and $\mu_t \equiv E_t[r_{t+1}]$, respectively. They follow simple autoregressive processes:

$$g_{t+1} = \gamma_0 + \gamma_1(g_t - \gamma_0) + \varepsilon^g_{t+1}, \tag{3}$$

$$\mu_{t+1} = \delta_0 + \delta_1(\mu_t - \delta_0) + \varepsilon^\mu_{t+1}. \tag{4}$$

The dividend growth rate is the expected dividend growth plus an orthogonal shock:

$$\Delta d_{t+1} = g_t + \varepsilon^d_{t+1}. \tag{5}$$

The vector of IID shocks $(\varepsilon^g_{t+1}, \varepsilon^\mu_{t+1}, \varepsilon^d_{t+1})'$ has covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_g^2 & \sigma_{g\mu} & \sigma_{gd} \\ \sigma_{g\mu} & \sigma_\mu^2 & \sigma_{\mu d} \\ \sigma_{gd} & \sigma_{\mu d} & \sigma_d^2 \end{bmatrix}. \tag{6}$$

The affine explicit expression for the log price-dividend ratio directly follows from a Campbell and Shiller (1988) log linearisation:

$$pd_t = A - B_1(\mu_t - \delta_0) + B_2(g_t - \gamma_0), \tag{7}$$

where A, B_1 and B_2 are simple functions of the model parameters such that, consistent with intuition, pd_t is decreasing in expected returns and increasing in expected dividend growth.[8]

---

[8]See Binsbergen and Koijen (2010) and Appendix A. Campbell and Shiller (1988) approximation in our sample holds almost exactly, for yearly data, if annual dividends and prices are constructed as described in Section 2.2.

2.2 Estimation results

We obtain the with- and without-dividend monthly returns on the value-weighted portfolio of all NYSE, Amex and Nasdaq stocks, in the period from January 1946 until December 2010, from the Center for Research in Security Prices (CRSP). We construct annual time series of aggregate dividends and prices, assuming that monthly dividends are cash-reinvested at the 30-day T-bill rate. Data on 30-day T-bill rates are also obtained from CRSP.

We estimate the model with a Kalman filter based on a Gaussian quasi likelihood function, from the observable time series of dividend growth Δd_{t+1} and price-dividend ratios pd_{t+1}. Due to the present-value relations, market return r_{t+1} is redundant with respect to Δd_{t+1} and pd_{t+1}.[9]

The parameter estimates are reported in Table 1, with bootstrapped standard errors in parenthesis.[10] We find an unconditional expected log return (dividend growth) of $\delta_0 = 8.3\%$ ($\gamma_0 = 5.7\%$). Both expectation processes feature some degree of persistence, with autoregressive roots γ_1 and δ_1 equal to 0.304 and 0.927, respectively, and expected returns are substantially more persistent than expected dividend growth. Finally, expected dividend growth is estimated as very volatile ($\sigma_g = 6.5\%$), while unexpected dividend growth variability is very low ($\sigma_d = 0.2\%$).

2.3 Dividend and return predictability

Let I_t denote the econometrician's information set at time t, generated by the history of dividends and price-dividend ratios. A nice feature of the Kalman filter is to provide filtered estimates of the unknown latent states μ_{t-1} and g_{t-1}, conditional on I_{t-1}. Thus, a standard measure of the degree of predictability in model (3)-(5) can be computed by

---

[9]Using (r_{t+1}, pd_{t+1}) as observable variables, the estimation results are almost identical and one can always recover the missing variable using Campbell and Shiller (1988) approximation; see also Cochrane (2008b), among others. Details on the estimation procedure are collected in Appendix B.

[10]Parameter standard errors are obtained using the circular block-bootstrap of Politis and Romano (1992), in order to account for the potential serial correlation in the data.

the fraction of r_t and Δd_t variability explained by μ_{t-1} and g_{t-1}, respectively:

$$R_{Ret}^2 \;\; = \;\; 1 - \frac{\widehat{Var}(r_{t+1} - \mu_t)}{\widehat{Var}(r_{t+1})}, \tag{8}$$

$$R_{Div}^2 \;\; = \;\; 1 - \frac{\widehat{Var}(\Delta d_{t+1} - g_t)}{\widehat{Var}(\Delta d_{t+1})}, \tag{9}$$

where $\widehat{Var}$ denotes sample variances.

We find that $R_{Ret}^2 = 8.82\%$ and $R_{Div}^2 = 17.58\%$, indicating that the degree of dividend predictability is about twice as large as the degree of return predictability. In contrast to these findings, simple regressions of returns and dividend growth on lagged price-dividend ratios yield R^2 of about 9.9% and 0.95%, respectively.

A possible interpretation for these diverging results is the noisiness of the price-dividend ratio (7) as a signal for expected returns and expected dividend growth, respectively, which creates a potential EIV problem in predictive regressions of returns and dividend growth on lagged price-dividend ratios. Indeed, the large persistence of return expectations is linked to a large sensitivity of price-dividend ratios to expected return shocks ($B_1 = 10.332$) and a smaller sensitivity to dividend expectation shocks ($B_2 = 1.421$). This feature obfuscates the predictive power of dividend expectations in dividend predictive regressions, leading to the low (biased) R^2.

3 Testing Predictability Hypotheses

Based on the above estimation results, it is natural to ask whether the empirical evidence supports (i) the hypotheses that dividends and returns are predictable and (ii) the EIV problem interpretation for the different predictability findings with respect to standard predictive regressions.

These questions are best addressed using an appropriate hypothesis testing framework. As emphasized, e.g., in Cochrane (2008a), while a point estimate produces the most likely predictability structure according to the chosen statistical metric, an hypothesis testing framework is needed to simultaneously control for the probability that some estimated predictability features might be generated by chance alone.

3.1 Asymptotic tests

Most predictability hypotheses can be formulated by means of simple parametric constrains, which can be efficiently tested with a standard likelihood ratio (LR) test, using the statistic

$$LR_T = 2 \left(\max_{\Theta} \log \mathcal{L} \left(\theta, \{Y_t\}_{t=1}^T \right) - \max_{\Theta_0} \log \mathcal{L} \left(\theta, \{Y_t\}_{t=1}^T \right) \right), \tag{10}$$

where Θ_0 is the restricted set of parameters under the given null hypothesis H_0 and $\log \mathcal{L}$ is the log-likelihood of the model. Evidence against H_0 is collected when LR is sufficiently large:

$$\{LR_T > c_{1-\alpha}\}, \tag{11}$$

relative to a critical value $c_{1-\alpha}$ that is unlikely under H_0. As $T \to \infty$, statistic LR_T follows a χ_r^2 distribution with r degrees of freedom, where r is the number of parameter constraints defining the constrained parameter set Θ_0. Therefore, the choice $c_{1-\alpha} = \chi_{r,1-\alpha}^2$, where $\chi_{r,1-\alpha}^2$ is the $1 - \alpha$ quantile of the chi-square distribution, ensures asymptotically a small probability α of rejecting H_0 by chance alone:

$$\alpha = \lim_{T \to \infty} P_{H_0}(LR_T > \chi_{r,1-\alpha}^2). \tag{12}$$

3.2 Time-varying expectations

Testing return and cash flow predictability is equivalent to examining time-variation in expected returns and expected dividend growth, respectively. In terms of the model parameters, the null of constant return expectations is:

$$H_0 : \delta_1 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = 0. \tag{13}$$

Similarly, the null of constant dividend expectations is:[11]

$$H_0 : \gamma_1 = \sigma_g = \rho_{g\mu} = 0. \tag{14}$$

---

[11]Under the null (13) (the null (14)) all price-dividend ratio variation comes from variation in expected dividend growth (returns) and the present-value model collapses to a standard linear regression of dividend growth rates (returns) on the lagged price-dividend ratio. Note that $\rho_{gd} = 0$ is imposed also in the unconstrained model for identification purposes, as for instance in Binsbergen and Koijen (2010), see Appendix B.

Using standard asymptotic critical values from a χ_r^2 distribution, where $r = 4$ and $r = 3$, respectively, Table 2 shows that both these null hypotheses are strongly rejected according to the asymptotic likelihood ratio test based on statistic (10), for significance levels α below 0.5%.

3.3 Expectation persistence and EIV-problem

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 (7) 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 1 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.[12] 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 2 shows the R-squared of returns and dividend growth implied by the present-value model described in Section 2 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

---

[12]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 Appendix C.

in Section 2.3. Note that under the null hypothesis of equal expectation persistences:[13]

$$H_0 : \gamma_1 = \delta_1 , \tag{15}$$

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.[14]

Based on asymptotic critical values from a χ_r^2 distribution ($r = 1$), we find in Table 2 that null hypothesis (15) is also clearly rejected, at a significance level α well below 1%. In summary, the evidence from asymptotic tests supports the following predictability features:

(i) A time-variation in expected returns and expected dividends;

(ii) A large (a negligible) EIV problem in standard predictive regressions for dividends (returns);

(iii) A larger degree of predictability, in terms of estimated R^2s, for dividends than for returns.

3.4   Finite-sample reliability of asymptotic tests

How reliable is the asymptotic approximation (12) for the probability of rejecting H_0 by chance alone? Since the conventional asymptotics might provide inaccurate results (see again Stoffer and Wall (1991) and Stoffer and Wall (2004)), we investigate the quality of this approximation for tests of predictability in the benchmark present-value model.

3.4.1   Nonparametric Monte Carlo bootstrap

Under null hypothesis H_0, we can consistently estimate the quantiles of statistic LR_T, using a simulation approach that does not rely on distributional assumptions on observed

---

[13]A 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 (15) was also imposed, e.g., for the present-value models in Cochrane (2008b) and Lettau and Van Niewerburgh (2008).

[14]Under null hypothesis (15), pd_t follows a standard AR(1) process; see also Stambaugh (1999) and Lewellen (2004), among others, and Section A of the Supplemental Appendix.

dividend and price-dividend ratio shocks. We borrow from Stoffer and Wall (1991) and apply a nonparametric Monte Carlo bootstrap to the fitted innovations in the present-value model. Details and formal justification for this approach are provided in Section 4.[15]

We impose the null hypothesis H_0 using the constrained Maximum Likelihood estimator $\hat{\theta}_0$ and we simulate $B = 1000$ time series of dividend growth and price-dividend ratios. Given the actual probability $\alpha_T := P_{H_0}(LR_T \geq \chi^2_{r,1-\alpha})$ of rejecting H_0 by chance alone in the asymptotic tests, we consistently estimate α_T with the bootstrap estimator:

$$\hat{\alpha}_T := \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(LR^*_{T,b} > \chi^2_{r,1-\alpha}) \,, \tag{16}$$

where $LR^*_{T,b}$ is the value of the likelihood ratio statistic in simulated bootstrap sample $b = 1, \ldots, B$ and $\mathbb{I}(A)$ denotes the indicator function of event A.[16]

3.4.2 Constant return or dividend expectations

For null hypothesis (13) (null hypothesis (14)), the first (second) Panel of Figure 3 displays the estimated quantiles of the empirical distribution of likelihood ratio statistic (10) under H_0, against the quantiles of the asymptotic χ^2_4 (χ^2_3) distribution.

Apparently, the finite-sample distributions of the test statistics deviate substantially from their asymptotic limit. For instance, while for a significance level $\alpha = 5\%$ the asymptotic critical value for the hypothesis of constant expected dividend growth is $\chi^2_{3,0.95} = 7.81$, the finite-sample critical value is more than two times larger (17.13).

Overall, we find that both asymptotic tests tend to reject H_0 too often. For instance, the estimated probability of rejecting null hypothesis (14) (null hypothesis (13)) by chance alone in a test of asymptotic significance level $\alpha = 5\%$ is as large as 25.8% (60.5%); see the last row of Table 2.

---

[15]Rytchkov (2012) also recognizes that inference based on standard asymptotics may be incorrect and applies a parametric Monte Carlo method with Gaussian shocks to estimate the finite-sample distribution of the LR statistic for the test of no return predictability. We prefer to avoid a parametric Monte Carlo simulation with jointly normal dividend and price-dividend ratios, because for several null hypotheses relevant to our analysis we have found the fitted model residuals under H_0 to deviate quite significantly from normality.

[16]$\mathbb{I}(A)(\omega) = 1$ ($\mathbb{I}(A)(\omega) = 0$) if and only if $\omega \in A$ ($\omega \notin A$).

3.4.3 Equal expectation persistence

For null hypothesis (15), the third panel of Figure 3 plots the quantiles of the empirical distribution of likelihood ratio statistic (10) under H_0, against those of the asymptotic χ_1^2 distribution. The finite-sample quantiles under H_0 are quite different from their asymptotic limit: The asymptotic test is again excessively liberal. For instance, for a significance level $\alpha = 5\%$ a test of null hypothesis $\delta_1 = \gamma_1$ has an asymptotic critical value $\chi_{0.95,1}^2 = 3.84$, which is less than half the finite-sample critical value estimated with the bootstrap approach (9.43). This difference implies a finite-sample probability of rejecting the null by chance alone as large as 26.4%, according to the bootstrap estimate (16); see again the last row of Table 2.

4 Bootstrap Tests in the Present-Value Model

A powerful approach to obtain asymptotically valid tests that are less susceptible to finite-sample distortions or specific distributional assumptions, can rely on nonparametric Monte Carlo methods, such as the bootstrap.[17]

We first introduce our bootstrap tests of predictability hypotheses in present-value models. We then show their asymptotic validity and quantify by Monte Carlo simulation the improvements over conventional asymptotic tests. Finally, we revisit the conclusions about return and dividend predictability in the benchmark present-value model.

4.1 State-space representation

For observed variables $Y_t := (\Delta d_t, pd_t)'$ and expanded state vector $X_t := (\hat{g}_{t-1}, \varepsilon_t^g, \varepsilon_t^\mu, \varepsilon_t^d)'$, where $\hat{g}_t := g_t - \gamma_0$, the present-value model can be written in state-space form (see

---

[17]As shown in Hall and Horowitz (1996) and Andrews (2002), among others, a desirable property of the bootstrap is that it may provide more accurate finite-sample approximations of the sampling distribution of standard $t-$test statistics for testing the null of no predictability in predictive regression models. Ang and Bekaert (2007) use bootstrap methods to quantify the bias of standard estimators of regression in predictive regressions of future returns on the lagged price-dividend ratio and interest rate. Amihud, Hurvich, and Whang (2009) compare the performance of bootstrap tests to bias-corrected procedures in multi-predictor regressions and find the two to provide similar finite-sample accuracy.

Appendix B):

$$X_{t+1} = FX_t + \Gamma \varepsilon_{t+1}^X \ , \tag{17}$$

$$Y_t = M_0 + M_1 Y_{t-1} + M_2 X_t \ , \tag{18}$$

with matrices F, Γ, M_0, M_1, M_2 that are functions of parameter vector

$$\theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_g, \sigma_\mu, \sigma_d, \rho_{g\mu}, \rho_{\mu d}, \rho_{gd})' \ .$$

Let $X_{t,t-1}$ be the best linear prediction of X_t based on observable data $\{Y_s\}_{s=1}^{t-1}$, obtained via the Kalman-Bucy filter, and $\eta_t = Y_t - M_0 - M_1 Y_{t-1} - M_2 X_{t,t-1}$ the corresponding prediction error. The innovations form representation of the present-value model follows from the Kalman filter as:

$$X_{t+1,t} = FX_{t,t-1} + F\mathcal{K}_t \eta_t \ , \tag{19}$$

$$Y_t = M_0 + M_1 Y_{t-1} + M_2 X_{t,t-1} + \eta_t \ , \tag{20}$$

where the Kalman gain $\mathcal{K}_t$ and the conditional covariance matrix S_t of innovation η_t are given explicitly in Appendix B.

The advantage of representation (19)-(20) for an efficient nonparametric bootstrap procedure, is that it allows to easily simulate forward the dynamics of observable variables $\{Y_1, \ldots, Y_T\}$, given initial conditions Y_0, $X_{0,0}$ and random innovations $\{\eta_1, \ldots, \eta_T\}$.[18]

4.2 Nonparametric Monte Carlo bootstrap

Let $\hat{\theta}$ and $\hat{\theta}_0$ be the unconstrained and the constrained estimators of the model parameters, obtained by maximizing the likelihood function (49) in Appendix B over the full and the H_0-constrained parameter set, $\hat{\Theta}$ and $\hat{\Theta}_0$, respectively. The observed value of the likelihood ratio statistic LR_T then follows from definition (10).[19]

---

[18]In practice, we first apply a nonparametric bootstrap to efficiently simulate the joint distribution of innovations $\{\eta_1, \ldots, \eta_T\}$. In a second step, we simulate the joint distribution of $\{Y_1, \ldots, Y_T\}$ using the forward dynamics (19)-(20).

[19]Bootstrap inference is always conditional on the observed sample of data. With a slight abuse of notation, in the sequel we denote by LR_T the sample value of the likelihood ratio statistics.

Stoffer and Wall (1991) apply a nonparametric Monte Carlo bootstrap to the (standardized) innovations $\{\hat{e}_t := S_t^{-1/2}(\hat{\theta})\eta_t(\hat{\theta})\}_{t=1}^T$, in order to obtain the standardized bootstrap residuals $\{\hat{e}_t^*\}_{t=1}^T$. The bootstrap residuals are used to compute a bootstrap distribution of maximum likelihood estimators $\hat{\theta}^*$:

$$\hat{\theta}^* = \arg\max_{\Theta} \log \mathcal{L}\left(\theta, \{Y_t^*\}_{t=1}^T\right) \; , \tag{21}$$

where the Monte Carlo sequence $\{Y_t^*\}_{t=1}^T$ is simulated with the dynamics (19)-(20) applied to the unstandardized bootstrap residuals $\{\hat{\eta}_t^* := S_t^{1/2}(\hat{\theta})e_t^*\}_{t=1}^T$.

This approach gives rise to a valid bootstrap distribution for $\sqrt{T}(\hat{\theta}^* - \hat{\theta})$, which is equivalent in large samples to the distribution of $\sqrt{T}(\hat{\theta} - \theta^\star)$, where $\theta^\star$ is the true unknown parameter value. We start from this result, to construct a valid nonparametric bootstrap likelihood ratio test of null hypothesis H_0 in the present-value model.[20]

4.3 Nonparametric Monte Carlo bootstrap likelihood ratio test

Our bootstrap likelihood ratio test for state-space model (17)-(18) is based on the following six-steps algorithm.

1) Using the estimated parameter vector under null hypothesis H_0, construct the (constrained) time series of standardized innovations $\{\hat{e}_{0t}\}_{t=1}^T$, by setting:

$$\hat{e}_{0t} = S_t^{-1/2}(\hat{\theta}_0)\eta_t(\hat{\theta}_0) \; , \tag{22}$$

where $S_t^{-1/2}$ is the inverse of the unique square root of S_t.

2) Applying a nonparametric bootstrap procedure (such as, e.g., the circular block-bootstrap in Politis and Romano (1992)) to time series $\{\hat{e}_{0t}\}_{t=1}^T$, compute a bootstrap sample $\{\hat{e}_{0t}^*\}_{t=1}^T$ of standardized innovations.

---

[20]In a robustness check, Rytchkov (2012) applies a version of a nonparametric bootstrap method, which simulates the full state-space dynamics, in order to test the null hypothesis of constant expected returns in a present-value setting. In contrast, we develop a bootstrap method for the innovation form representation of the state-space model. In addition to producing a less computationally demanding procedure, this approach yields a more transparent bootstrap simulation scheme that allows us to prove the formal asymptotic validity of our approach.

3) Using the innovation form representation (19)-(20), construct a bootstrap sample $\{Y_t^*\}_{t=1}^T$ as follows:

$$X_{t+1,t}^* = FX_{t,t-1}^* + F\mathcal{K}_t S_t^{1/2} \hat{e}_{0t}^* , \tag{23}$$

$$Y_t^* = M_0 + M_1 Y_{t-1}^* + M_2 X_{t,t-1}^* + S_t^{1/2} \hat{e}_{0t}^* , \tag{24}$$

where matrices F, $\mathcal{K}_t$, S_t, M_0, M_1, M_2 are all evaluated in $\hat{\theta}_0$ and the initial conditions are $Y_0^* = Y_0$, $X_{0,-1}^* = X_{0,0}$.

4) Using bootstrap sample $\{Y_t^*\}_{t=1}^T$, compute constrained and unconstrained maximum likelihood point estimates $\hat{\theta}_0^*$ and $\hat{\theta}^*$, respectively, by maximizing the log likelihood function $\log \mathcal{L}\left(\theta, \{Y_t^*\}_{t=1}^T\right)$, while imposing and not imposing null hypothesis H_0, respectively.

5) Following definition (10), compute the value LR_T^* of the likelihood ratio statistic in the bootstrap sample, defined by:

$$LR_T^* = 2\left(\log \mathcal{L}\left(\hat{\theta}^*, \{Y_t^*\}_{t=1}^T\right) - \log \mathcal{L}\left(\hat{\theta}_0^*, \{Y_t^*\}_{t=1}^T\right)\right) . \tag{25}$$

6) Repeat steps 2)-5) a large number of times, B, to obtain a collection of bootstrap values of the likelihood ratio statistics, $\{LR_{T,b}^*, 1 \leq b \leq B\}$. The empirical distribution of these values provides an approximation of the distribution of the likelihood ratio statistic under the null hypothesis H_0.

Remark 1 *(i) In step 2) of the algorithm, several bootstrap procedures are applicable to the standardized innovations $\{\hat{e}_{0t}\}_{t=1}^T$. We recommend a time-series bootstrap, such as the circular block-bootstrap, in order to robustify the test against a potentially left time series dependence, not captured by the estimated conditional moment dynamics. ii) In some cases, it may help to exclude the random sampling of the innovations for the first 2-3 data points in step 2) of the algorithm, e.g., by setting $\hat{e}_{0t}^* = \hat{e}_{0t}$ for $t = 1, 2, 3$. This is useful to avoid start-up problems of the algorithm, when the Kalman filter might have an initially transient behavior, e.g., with large values of the Kalman gain $\mathcal{K}_t$.*

An important question is whether the proposed bootstrap test delivers correct results in large samples, i.e., whether the bootstrap likelihood ratio statistic LR_t^* follows the same

asymptotic distribution as LR_T under H_0. The next theorem justifies our bootstrap likelihood ratio test of null hypothesis H_0.

Theorem 1 *Under regularity conditions detailed in Appendix D, it follows as $B, T \to \infty$: $LR_T^* \longrightarrow \chi_r^2$, in distribution.*

According to Theorem 1, the bootstrap statistic LR_T^* has an asymptotically equivalent distribution to LR_T under H_0.[21] Therefore, it gives rise to bootstrap tests with the correct significance level asymptotically.

The most convenient way to define a bootstrap likelihood ratio test of H_0 is by means of the so-called bootstrap $p-$value:

$$p^*(LR_T) := P^*(LR_T^* > LR_T) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(LR_{T,b}^* > LR_T) , \qquad (26)$$

where P^* denotes the bootstrap probability measure. Using bootstrap $p-$values, the bootstrap test rejects H_0 whenever:

$$p^*(LR_T) < \alpha . \qquad (27)$$

From Theorem 1, this test implies the correct asymptotic size α. The interesting question then is whether the bootstrap test delivers more reliable results in finite samples.

A useful property in this respect is that the inference based on bootstrap procedures applied to asymptotically pivotal statistics, such as the likelihood ratio statistic, is generally more accurate than the inference of conventional asymptotics, in the sense that the errors made are of lower order in the sample size T;[22] see Beran (1988), Davidson and MacKinnon (1999b), Hall and Horowitz (1996) and Andrews (2002), among others. As a consequence, we can hope that bootstrap likelihood ratio tests will improve over the conventional asymptotic inference in realistic applications.

4.4 Finite-sample reliability of bootstrap likelihood ratio tests

In this section, we investigate by Monte Carlo simulation the finite-sample properties of our bootstrap tests, in the context of the benchmark present-value model.

---

[21]A similar result can be proven with respect to sequences of shrinking local alternative hypotheses $H_{A,T}$, by applying the above algorithm to innovations defined by $\hat{e}_{At} = S_t^{-1/2}(\hat{\theta}_{A,T})\eta_t(\hat{\theta}_{A,T})$ in step 1), where $\hat{\theta}_{A,T}$ is the constrained maximum likelihood estimator computed under the local alternative $H_{A,T}$.

[22]A pivotal statistic is a statistic with sampling distribution independent of nuisance parameters.

We first impose the null hypothesis H_0, using the constrained ML estimator $\hat{\theta}_0$, and simulate S time series of dividend growth and price dividend ratios, using our nonparametric bootstrap procedure applied to the fitted innovations in the present-value model. In this way, we can simulate the empirical distribution of observed data, under the null hypothesis H_0, without making strong parametric assumptions on the joint distribution of dividend and price-dividend ratio shocks. For each simulated time series $s = 1, \ldots, S$, we compute the corresponding value of the likelihood ratio statistic, denoted by $LR_{T,s}$.

We then apply our bootstrap testing method to the simulated data. For each simulated time series $s = 1, \ldots, S$, we compute a bootstrap distribution of likelihood ratio statistics $LR^*_{T,b,s}$, $b = 1, \ldots, B$, and we compute the resulting bootstrap p-value:

$$p^*(LR_{T,s}) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(LR^*_{T,b,s} > LR_{T,s}) \ ,$$

following the algorithm in Section 4.3. For a significance level $\alpha = 5\%$, we finally compute the frequency of rejections of H_0 in the bootstrap test, i.e., the fraction of time series $s = 1, \ldots, S$ in which $p^*(LR_{T,s}) < \alpha$, and compare it to the frequency of rejections obtained by following the asymptotic testing approach. We denote these rejection frequencies by α^*_T and α_T, respectively. Overall, our Monte Carlo simulation is based on a double-bootstrap simulation scheme with $2S(B + 1)$ estimations of the parameters in the present-value model, which is a computationally demanding procedure. We present our Monte Carlo results for the parameter choices $S = 200$, $B = 99$ and an optimal bootstrap block size of 2.[23] Other parameter choices produce similar results.

For null hypothesis (13), we obtain $\alpha^*_T = 5\%$ for the bootstrap test, which is exactly equal to the nominal level ($\alpha = 5\%$), while for null hypothesis (14), the empirical size of the bootstrap test is $\alpha^*_T = 8\%$, which is clearly closer to the given nominal level ($\alpha = 5\%$) than the rejection frequency $\alpha_T = 25.8\%$ implied by the asymptotic test. Even though the bootstrap test is slightly too liberal in the Monte Carlo simulation, it does not reject the null of constant dividend expectations in our data. For null hypothesis (15), we obtain

---

[23]We apply a data driven calibration method for the selection of the block size, similar to the one introduced in Romano and Wolf (2001) and discussed by Camponovo, Scaillet, and Trojani (2009). We choose the block size that minimizes the difference between empirical and nominal size of the bootstrap test for equal expectation persistences.

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 in the Monte Carlo simulation.

4.5   The empirical evidence revisited

We make use of the bootstrap likelihood ratio test in Section 4.3 and compare the results with those of conventional asymptotic tests in Table 2. Based on a bootstrap size $B = 1000$ and an optimal block size of 2, Table 3 shows that the null hypothesis of a constant expected return is rejected at a significance level $\alpha = 1\%$ by the bootstrap test, but the null hypothesis of constant expected dividend growth is not rejected, with a bootstrap p-value of 9.5%.

The p-value of the null hypothesis of equal autoregressive coefficients in Table 2 is 2.4% (see Table 3) for the bootstrap test, compared to the p-value of about 0.05% implied by the standard asymptotic results.[24] Thus, null hypothesis (15) 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 the slightly too liberal behaviour of bootstrap tests in our Monte Carlo simulations, the evidence against a similar persistence of expected returns and expected dividend growth is more ambiguous than under the conventional asymptotic tests.

In summary, the non-rejection of null hypotheses (14) and (15) based on bootstrap tests suggests the following different predictability features, when compared to the findings in Section 3 for the asymptotic tests:

(i) A time-variation in expected returns, but no apparent evidence of a time-varying expected dividend growth;

(ii) A moderate EIV-problem in dividends and return predictive regressions;

---

[24]Consistent with the standard asymptotic results, a parametric bootstrap likelihood ratio test with Gaussian errors yields p-values of 0%, 0.3% and 0.2% for the null hypotheses of constant expected return, constant expected dividend growth and equal autoregressive coefficients, respectively. This evidence underlines the importance of avoiding too stringent distributional assumptions when testing predictability hypotheses in present-value models.

(iii) Returns and dividend growth predictability features roughly consistent with those of standard predictive regressions.

4.6   How much predictability?

The weak evidence of dividend growth predictability produced by bootstrap likelihood-ratio tests raises the question of the interpretation of the large R-squared ($R^2_{div} = 17.58\%$) estimated in Section 2.3 for future dividends.

Differently from standard predictive regressions, the asymptotic distribution of estimated R-squares in the present-value model is not known in closed-form. Therefore, the conventional asymptotic approach cannot be used, e.g., to quantify the probability of estimating large R-squares because of chance alone. In contrast, our bootstrap methodology can be applied with no major modification to consistently estimate such probability, under the assumption that an asymptotic distribution for the estimated R-squares exists.

Using steps 1)-3) of the algorithm in Section 4.3, we can compute bootstrap estimates of parameter θ in the present-value model and obtain the bootstrap distribution of estimated R-square statistics under a given null hypothesis H_0. Figure 4 displays the histogram of the bootstrap distribution of estimated R-squares for future returns and future dividend growth, under the null hypotheses of a constant expected dividend growth and an equal persistence of dividend and return expectations, respectively.

Apparently, the bootstrap distribution of estimated R-squares under the two null hypotheses is similar. Moreover, even though the model-implied R-squared for dividend growth under H_0 is 0% and 0.9%, respectively, we find that the variability of estimated R-squares is quite large. For instance, the median estimated R^2_{Div}-value is 6.02% under the null of constant expected dividend growth (5.34% under the null of equal persistence parameters) and the most frequently estimated R-squared value is 0% in both cases, but the probability of estimating a dividend R-squared of at least 17.58%, as in the data, is 11.3% (10.5%).

Overall, these findings highlight that finite-sample variability is important for appropriately interpreting the finite-sample distribution of estimated R-squares, as large R-squares as in the data can arise by chance alone, in a present-value model where divi-

dend predictability is absent or weak.

4.7   Out-of-sample predictability

All R^2 values reported in the previous sections are estimated using in-sample data. From the perspective of real-time predictability, out-of-sample prediction is an additional important aspect. For instance, Goyal and Welch (2008) study the out-of-sample predictive power of a large set of variables for market returns and find that most of them perform worse than the historical mean.

Following Campbell and Thompson (2008) and Goyal and Welch (2008), the incremental out-of-sample predictive power for returns and dividend growth in the present-value model of Section 2 can be estimated using the metrics:

$$R^2_{Ret,OS} = 1 - \frac{\sum_{t=0}^{T} (r_{t+1} - \tilde{\mu}_t)^2}{\sum_{t=0}^{T} (r_{t+1} - \bar{r}_t)^2} \ , \tag{28}$$

$$R^2_{Div,OS} = 1 - \frac{\sum_{t=0}^{T} (\Delta d_{t+1} - \tilde{g}_t)^2}{\sum_{t=0}^{T} (\Delta d_{t+1} - \overline{\Delta d}_t)^2} \ , \tag{29}$$

where $\tilde{\mu}_t$ and $\tilde{g}_t$ are the estimated expected return and expected dividend growth in the present value model, using observations up to time t, while $\bar{r}_t$ and $\overline{\Delta d}_t$ are the sample means of returns and dividend growth using data up to time t.

We estimate the degree of out-of-sample predictability according to measures (28) and (29), using an out-of-sample period starting in 1985. Standard predictive regressions of returns and dividend growth on the lagged price-dividend ratio yield $R^2_{Ret,OS} = -12.32\%$ and $R^2_{Div,OS} = -4.38\%$, while we obtain $R^2_{Ret,OS} = -7.31\%$ and $R^2_{Div,OS} = 5.88\%$ for the present-value model.[25] Thus, the point estimates for the benchmark present-value model might indicate an incremental degree of out-of-sample predictability for dividend growth with respect to the sample mean forecast.

---

[25]Precisely, we use data between 1946 and 1985 to estimate the parameters of the model and compute expected return and expected dividend growth for 1986, which are compared to the realized return and dividend growth in the same year. We then use data between 1946 and 1986 to compute predictions for 1987 and proceed in this way until the end of the sample. Using data from 1946 to 2007 and starting the out-of-sample computations in 1972, Binsbergen and Koijen (2010) find $R^2_{Ret,OS} = 1.06\%$ and $R^2_{Div,OS} = 5.76\%$.

Using a slight modification of our bootstrap method, we can estimate the distribution of out-of-sample R-squares (28) and (29) under the null of no return or dividend predictability; details of the procedure are given in Appendix E. This approach is useful, e.g., to better quantify the probability of estimating large out-of-sample R^2 values as in the data by chance alone.

Given the variability of estimated in-sample R^2_{Div} values highlighted in Section 4.6, it is plausible that the out-of-sample R-squared distribution might inherit similar features. Figure 5 illustrates the properties of the bootstrap distributions of out-of-sample R-squares (28) and (29), generated under the null hypothesis of constant expected cash flow growth in the present-value model. Both distributions imply a large variability of estimated out-of-sample measures of predictability for returns (upper panel) and dividend growth (lower panel). Even though under H_0 expected returns are time-varying, the estimated $R^2_{Ret,OS}$ distribution puts a large mass in regions where no evidence of incremental predictability is estimated. Moreover, despite the absence of dividend predictability under the null, the distribution of estimated $R^2_{Div,OS}$'s puts a significant mass of about 15% in regions of positive $R^2_{Div,OS}$ values, with a probability of estimating an out-of-sample R-squared for dividends at least as large as in the data that is almost 10%.

Overall, these findings show that the conclusions produced by estimated common measures of out-of-sample predictability in present-value models have to be taken with caution and put in relation to the finite-sample variability of these quantities under the null of no predictability. On the one side, the limited amount of data information available can lead to a difficulty in detecting predictive relations for returns when they are there. On the other side, high out-of-sample R-squares for dividends can arise by chance alone, in a setting with constant expected dividend growth. In this respect, our nonparametric bootstrap approach provides a useful tool to better interpret also the information provided by estimated out-of-sample measures of predictability.

5 Broader Specifications of the Predictive Information Set

While the benchmark present-value model in Section 2 is useful for highlighting the main issues of tests of predictability hypotheses, it might not provide the most accurate description for the dynamics of dividend-return expectations and their link to price-dividend ratios. Richer specifications might improve the evidence of predictability and it is useful to study the robustness of our previous results, with respect to an enlarged specification of the predictive information set.

Several potential predictors have been considered in the literature, to improve the statistical evidence of univariate predictive regressions with the lagged price-dividend ratio.[26] Such predictive variables can naturally extend the benchmark present-value model, in order to parsimoniously aggregate the joint information generated by the time series of dividend growth, price-dividend ratios and additional predictors, following the present-value approach proposed in Yun (2012).

Using the conventional asymptotic approach, variables such as the book-to-market ratio (bm), the stock market variance ($svar$), the consumption-wealth-income ratio (cay) and the BAA-rated corporate bond yield (BAA) significantly improve the forecasts of future returns and future dividend growth in the present-value model.[27] Using our general bootstrap tests of Section 4, we study the robustness of our findings on dividend and return predictability, with respect to the choice of the predictive information set.

---

[26]Goyal and Welch (2008) and Koijen and Van Nieuwerburgh (2011) give an excellent review of this literature. Even though less studies have focused on dividend growth predictability, Lettau and Ludvigson (2005) and Favero, Gozluklu, and Tamoni (2011), among others, provide evidence that predictive variables like cay and proxies of demographics help forecasting cash flow growth.

[27]The benchmark present-value model assumes a constant return volatility. Piatti and Trojani (2012) develop a present-value approach with time-varying return and dividend growth risks to predictive regression.

5.1 The present-value model with additional predictive variables: estimation results

Expected dividend growth, expected return and an additional predictive variable, z_t, follow the following first-order vector autoregression:

$$g_{t+1} = \gamma_0 + \gamma_1(g_t - \gamma_0) + \gamma_2(z_t - \xi_0) + \varepsilon_{t+1}^g, \tag{30}$$

$$\mu_{t+1} = \delta_0 + \delta_1(\mu_t - \delta_0) + \delta_2(z_t - \xi_0) + \varepsilon_{t+1}^\mu, \tag{31}$$

$$z_{t+1} = \xi_0 + \xi_1(z_t - \xi_0) + \varepsilon_{t+1}^z. \tag{32}$$

In contrast to the benchmark dynamics (3)-(4), the additional predictive variable z_t can help to better explain expected returns or expected dividend growth. As such, it appears in the price-dividend ratio implied by a standard Campbell and Shiller (1988) log linearization:

$$pd_t = A - (B_1\hat\mu_t + B_3\hat z_t) + (B_2\hat g_t + B_4\hat z_t), \tag{33}$$

where $B_3 = \frac{\delta_2}{\delta_1 - \xi_1}\left(\frac{1}{1-\rho\delta_1} - \frac{1}{1-\rho\xi_1}\right)$, $B_4 = \frac{\gamma_2}{\gamma_1 - \xi_1}\left(\frac{1}{1-\rho\gamma_1} - \frac{1}{1-\rho\xi_1}\right)$ and $\hat z_t = z_t - \xi_0$ is the demeaned additional predictive variable at time t; see, e.g., Yun (2012).

The model is estimated in state-space form with a Kalman filter.[28] For brevity, we report results only for additional predictive variables that significantly predict returns and dividend growth using standard asymptotic tests. These include the book-to-market ratio (bm), the stock market variance ($svar$), the consumption-wealth-income ratio (cay) and the corporate bond yield on BAA-rated bonds (BAA). The description of the variables is provided by Goyal and Welch (2008) and their updated time series through 2010 are available at Goyal's website.[29]

Estimated present-value model parameters and R-squared for returns and dividend growth are collected in Table 4, together with the R-squared estimated from standard predictive regressions with the additional predictive variable z_t. Consistently with the findings in Yun (2012), in each present-value model the predictive information set enlarged by the additional predictor z_t increases the estimated R-squares for dividends and returns,

---

[28]For completeness, Appendix B also describes the state-space representation and the Kalman filter estimation procedure for the present-value model with the additional predictor z_t.

[29]See the web page http://www.hec.unil.ch/agoyal/.

relative to the findings for the benchmark model in Section 2.3. While estimated R-squares for returns are similar to those obtained from the standard predictive regressions in Panel C of Table 4, the estimated R-squared values for dividend growth are much higher, consistently with the findings of Section 2.3 for the benchmark present-value model.

5.2 Tests of constant dividend and return expectations

Cash flow predictability is again tested by testing the null hypothesis of constant expected dividend growth. In the extended present-value model, this hypothesis is equivalent to the following constraints, which are tested using a standard LR statistic that is asymptotically χ^2_5 distributed:

$$H_0 : \gamma_1 = \gamma_2 = \sigma_g = \rho_{g\mu} = \rho_{gz} = 0 \ . \tag{34}$$

We test this null hypothesis for $z_t = bm$ and $z_t = svar$, which are the variables that seem to increase more model-implied dividend growth predictability, measured in terms of R-squared, compared to the benchmark model (see again Panel B of Table 4). Panel B of Table 5 shows that the asymptotic likelihood ratio test rejects null hypothesis (34) for both choices of predictive variable z_t, with a p-value below 0.5%.

To apply our bootstrap testing approach, we introduce the extended vectors of observed variables $Y_t := (\Delta d_t, pd_t, z_t)'$ and state variables $X_t := (\hat{g}_{t-1}, \epsilon^g_t, \epsilon^\mu_t, \epsilon^d_t, \epsilon^z_t)'$, in order to write the present-value model (30)-(32) in state-space form (see Appendix B):

$$X_{t+1} \ = \ FX_t + Bu_{t+1} + \Gamma\epsilon^X_{t+1} \ , \tag{35}$$

$$Y_t \ = \ M_0 + M_1 Y_{t-1} + M_2 X_t \ , \tag{36}$$

with parameter-dependent matrices F, B, Γ, M_0, M_1, M_2 and variable $u_t := z_{t-1} - \xi_0$.

Given $X_{t,t-1}$ the best linear prediction of X_t based on data $\{Y_t\}^{t-1}_{s=1}$ and $\eta_t = Y_t - M_0 - M_1 Y_{t-1} - M_2 X_{t,t-1}$, the innovations form representation of model (30)-(32) follows from the Kalman filter:

$$X_{t+1,t} \ = \ FX_{t,t-1} + Bu_{t+1} + F\mathcal{K}_t\eta_t \ , \tag{37}$$

$$Y_t \ = \ M_0 + M_1 Y_{t-1} + M_2 X_{t,t-1} + \eta_t \ , \tag{38}$$

where the Kalman gain $\mathcal{K}_t$ is given explicitly in Appendix B. From this dynamics, the bootstrap likelihood ratio test in the extended present-value model is performed with the algorithm presented in Section 4.3.[30]

Panel C of Table 5 shows that the bootstrap likelihood ratio test produces different conclusions from the asymptotic test. The bootstrap test p-values are always bigger than the asymptotic p-values and we can never reject null hypothesis (34) at the 5% significance level, indicating that the evidence of dividend growth predictability is still weak in the extended present-value models, as it was in Section 4.5 for the benchmark model.

The null hypothesis of no return predictability in the extended present-value model is equivalent to the following parametric constraints:

$$H_0 : \delta_1 = \delta_2 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = \rho_{\mu z} = 0 \ . \tag{39}$$

For brevity, we test again this null hypothesis using the two predictive variables that mostly increase the return predictability evidence, as measured by the model-implied R^2_{Ret}, namely $z_t = bm$ and $z_t = cay$; see again Panel B of Table 4. Panel B of Table 6 shows that the asymptotic likelihood ratio test rejects null hypothesis (39) for all choices of the predictive variable z_t, with a p-value below 0.05%. The p-values for the bootstrap test are reported in Panel C of Table 6. Consistently with the asymptotic test results, null hypothesis (39) is again clearly rejected, with p-values of about 0.5%, as it was in Section 4.5 for the benchmark model.

5.3   Variability of estimated R-squared values

How can the weak evidence of dividend growth predictability be consistent with the large estimated dividend R-squares in the extended present-value models? To understand this, Figure 6 plots the bootstrap distribution of estimated R-squares for returns and future

---

[30]To run the bootstrap algorithm in the extended present-value model, we replace equation (19) in step 3) of the algorithm in Section 4 by the following bootstrap simulation scheme:

$$X^*_{t+1,t} = F X^*_{t,t-1} + B u_{t+1} + F \mathcal{K}_t S_t^{1/2} \hat{e}^*_{0t} \ ,$$

using parameter matrices detailed in Appendix B.

dividend growth, simulated under null hypothesis (34), for two different choices $z_t = bm$ (left panels) and $z_t = svar$ (right panels) of the additional predictive variable.

The bootstrap distribution of estimated R-squares under the null of constant expected dividend growth is similar to the distribution estimated in the benchmark model, with a large variability of estimated R-squares. The increased predictive information generated by z_t tends to rise the probability of correctly estimating an R-squared of 0% for dividend growth under the given null hypothesis for $z_t = bm$, while the distribution of R^2_{Div} displays more variability for $z_t = svar$. For instance, while the median estimated R^2_{Div}-value is 6.23% for $z_t = bm$ (9.41% for $z_t = svar$), the most frequently estimated R-squared value is 0%, but the probability of estimating a dividend R-squared of at least 22.32% (25.71%), as in the data, is still as large as 8.20% (10.60%).

In summary, finite-sample variability can again produce large estimated R-squares by chance alone, within a present-value model where dividend predictability is absent.

6 Conclusion

The Campbell and Shiller (1988) present-value logic, implying that price-dividend ratios vary because of shocks to expected returns or expected dividend growth, has motivated a vast literature studying the predictability of market returns and aggregate dividend growth. Predictive regressions of future returns and dividend growth on predictive variables including the lagged price-dividend ratio have produced no apparent evidence of dividend predictability in the post-war sample, suggesting that price-dividend ratios have mostly varied because of discount rate shocks in that period. In contrast, latent variable approaches within present-value models, which parsimoniously incorporate information from the joint time-series of dividends and returns, have found a stronger evidence of a time-varying expected dividend growth.

A natural explanation for these contrasting conclusions is the error-in-variable (EIV) problem inherent to predictability studies, which can be explicitly modelled using the present-value relations that connect return and dividend growth dynamics to the price-dividend ratio. This paper focuses on a different explanation, motivated by the peculiar finite-sample properties of conventional tests of predictability in models with latent return

and dividend expectations. Using a nonparametric Monte Carlo simulation approach, which avoids restrictive distributional assumption, such as, e.g., a normal distribution for dividend and return shocks, we show that the conventional tests have similar finite-sample drawbacks as many tests of predictability in predictive regressions with lagged persistent predictors and correlated innovations.

First, we show that conventional tests frequently reject the null of no dividend predictability because of chance alone. Moreover, we find that large estimated R-squares for dividends can arise by chance alone, even under the null of a constant expected dividend growth. These findings stress the importance of combining a pure estimation approach with a reliable testing method, when testing and quantifying the actual degree of predictability within present-value models.

Second, in order to introduce a general and more reliable testing approach, we propose a class of nonparametric bootstrap tests of predictability hypotheses in present-value models, by applying the bootstrap to the innovations from the latent state dynamics, generated under the relevant null hypothesis. We prove that the bootstrap tests imply a valid asymptotic inference and demonstrate their improved properties in finite samples. Precisely, we find that the bootstrap test can better control the finite-sample probability of rejecting a null hypothesis by chance alone, thus producing a more reliable predictability evidence in a number of applications.

Third, we apply our bootstrap tests to post-war US stock market data, based on a variety of specifications of the predictive information set, and we detect a significant evidence in favour of time-varying expected returns, but no sharp evidence of a time-variation in dividend expectations. The weak evidence of dividend predictability emerging from the different present-value models considered in the paper is consistent with the results of tests of predictability in standard predictive regression models, indicating that the most significant source of price-dividend ratio variation in post-war US data are discount rate shocks.

We finally propose a slight modification of our bootstrap testing method, which can be used also to test the presence of out-of-sample predictability, while controlling the probability of detecting predictive relations by chance alone. We find that the conclusions produced by estimated common measures of out-of-sample predictability in present-value

models have to be taken with caution and need to be set in relation to the finite-sample variability of these quantities under the null of no predictability. In this respect, our nonparametric bootstrap testing method provides a useful tool for more comprehensively interpreting also the information provided by estimated measures of out-of-sample predictability.

From a broader methodological perspective, our bootstrap testing approach and our results have implication for a number of potentially more general aspects. First, while our bootstrap tests can help to control more systematically the probability of rejecting a null hypothesis by chance alone, our results also indicate that the information generated by the joint time series of stock market returns and dividends might be insufficient to reliably identify time-variations in dividend expectations, i.e., tests of dividend predictability in such settings may have a low power.

A low power might arise because of the short time series available for many predictability studies or because market price-dividend ratios aggregate into a single observable signal the expectations of future dividends for different horizons, which are potentially difficult to identify separately. As shown in Binsbergen, Brandt, and Koijen (2012) and Binsbergen, Hueskes, Koijen, and Vrugt (2012), a more direct identification of dividend expectations at distinct horizons can rely on the equity yield of dividend strips, which are dividend claims for single maturities. Annual dividend growth is strongly predictable in the period from October 2002 to April 2011, with univariate predictive regression R^2s between 48% for the 5 year yield and 76% for the 1 year yield. This evidence suggests that dividend strip information can potentially improve the power of tests of dividend predictability more generally. Unfortunately, quotes of liquid dividend claims are available only since recently. Another possibility is to replicate synthetically the prices of dividend strips from quoted index option or futures prices, in which case data are available starting approximately in 1986; see, e.g., Binsbergen, Brandt, and Koijen (2012).

The study of reliable inference methods in present-value models estimating the joint dynamics of dividend growth, stock returns and dividend strip returns is an interesting direction for future research. Also in this domain, our bootstrap testing methods can prove useful in order to better control the probability of rejecting a null of no predictability by chance alone.

Finally, our bootstrap testing method is also applicable more generally, in order to more reliably test the relevant null hypotheses in models estimated by a latent variable approach using their state-space form. Concrete but not exhaustive examples of possible applications include the testing of the expectation and similar hypotheses in the context of affine factor models for the yield curve (see, e.g., Piazzesi (2010) for a review) or tests of predictability hypotheses in present-value models with time-varying return and dividend risks (see, e.g., Piatti and Trojani (2012)).

References

Amihud, Y. and C. Hurvich (2004). Predictive regression: a reduced bias estimation method. *Journal of Financial and Quantitative Analysis 39*, 813–841.

Amihud, Y., C. Hurvich, and Y. Whang (2009). Multiple-predictor regressions: Hypothesis testing. *Review of Financial Studies 22*, 413–434.

Andrews, D. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica 70*(6), 119–162.

Ang, A. and G. Bekaert (2007). Stock return predictability: Is it there? *Review of Financial Studies 20*(3), 651–708.

Beran, R. (1988). Prepivoting test statistics. *Journal of the American Statistical Association 83*, 687–697.

Binsbergen, J., M. Brandt, and R. S. Koijen (2012). On the timing and pricing of dividends. *American Economic Review 102*(4), 1596–1618.

Binsbergen, J., W. H. Hueskes, R. S. Koijen, and E. B. Vrugt (2012). Equity yields. Working Paper. Northwestern Kellogg, APG Asset Management, Chicago Booth and Hedge Fund - Singapore.

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.

Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies 21*(4), 1509–1531.

Campbell, J. Y. and M. Yogo (2006). Efficient tests of stock return predictability. *Journal of Financial Economics 81*(1), 27–60.

Camponovo, L., O. Scaillet, and F. Trojani (2009). Robust resampling methods for time series. Swiss Finance Institute Research Paper No. 09-38.

Chen, L. (2009). On the reversal of return and dividend growth predictability: A tale of two periods. *Journal of Financial Economics 92*(1), 128–151.

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

Cochrane, J. H. (2008a). The dog that did not bark: A defense of return predictability. *Review of Financial Studies 21*(4), 1533–1576.

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

Davidson, R. and J. MacKinnon (1999a). *Econometric Theory and Methods.* Oxford University Press.

Davidson, R. and J. MacKinnon (1999b). The size distortion of bootstrap tests. *Econometric Theory 15*, 361–376.

Efron, B. (1979). Bootstrap methods: Another look at the jacknife. *Annals of Statistics 7*, 1–26.

Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics 22*(3), 3–25.

Favero, C. A., A. E. Gozluklu, and A. Tamoni (2011). Demographic trends, the dividend-price ratio, and the predictability of long-run stock market returns. *Journal of Financial and Quantitative Analysis 46*(5), 1493–1520.

Goyal, A. and I. Welch (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

Hall, P. and J. Horowitz (1996). Bootstrap critical values for tests based on generalized-method-of-moment estimators. *Econometrica 64*(4), 891–916.

Keim, D. B. and R. F. Stambaugh (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics 17*(2), 357–390.

Koijen, R. S. and S. Van Nieuwerburgh (2011). Predictability of returns and cash flows. *Annual Review of Financial Economics 3*, 467–491.

Kothari, S. P. and J. Shanken (1997). Book-to-market, dividend yield, and expected market returns: A time series analysis. *Journal of Financial Economics 44*(2), 169–203.

Lettau, M. and S. C. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance 3*, 815–849.

Lettau, M. and S. C. Ludvigson (2005). Expected returns and expected dividend growth. *Journal of Financial Economics 76*(3), 583–626.

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.

Liung, L. and P. Caines (1979). Asymptotic normality of prediction error estimatorsfor approximatesystem models. *Stochastics 3*, 29–46.

Menzly, L., T. Santos, and P. Veronesi (2004). Understanding predictability. *Journal of Political Economy 112*(1), 1–47.

Pastor, L., M. Sinha, and B. Swaminathan (2008). Estimating the intertemporal risk-return tradeoff using the implied cost of capital. *Journal of Finance 63*(6), 2859–2897.

Piatti, I. and F. Trojani (2012). Predictable risks and predictive regression in present-value models. *SSRN eLibrary*.

Piazzesi, M. (2010). *Handbook of Financial Econometrics*, Volume 1, Chapter Affine Term Structure Models, pp. 691–766. North-Holland, Amsterdam.

Politis, D. N. and J. P. Romano (1992). *Exploring the Limits of Bootstrap*, Chapter A circular block-resampling procedure for stationary data, pp. 263–270. Wiley, New York.

Polk, C., S. Thompson, and T. Vuolteenaho (2006). Cross-sectional forecasts of the equity premium. *Journal of Financial Economics 81*, 101–141.

Rodriguez, A. and E. Ruiz (2009). Bootstrap prediction intervals in state-space models. *Journal of Time Series Analysis 30*(2), 167–178.

Rozeff, M. S. (1984). Dividend yields are equity risk premiums. *Journal of Portfolio Management 11*(1), 68–75.

Rytchkov, O. (2012). Filtering out expected dividends and expected returns. *Quarterly Journal of Finance 2*(3), 1250012.

Schiller, R. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity 2*, 457–498.

Spall, J. and K. Wall (1984). Asymptotic distribution theory for the kalman filter state estimator. *Communications in Statistics-Theory and Methods 13*, 1981–2003.

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

Stoffer, D. and K. Wall (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the kalman filter. *Journal of the American Statistical Association 86*, 1024–1033.

Stoffer, D. and K. Wall (2004). *State Space and Unobserved Components Models: Theory and Applications*, Chapter 9: Resampling in State Space Models, pp. 171–202. Cambridge University Press.

Torous, W., R. Valkanov, and S. Yan (2004). On predicting stock returns with nearly integrated explanatory variables. *Journal of Business 77*, 937–966.

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

A   Price-dividend ratio

In this section we present the detailed derivation of equation (7) in the text. From Campbell and Shiller (1988) we have

$$pd_t \simeq \kappa + \rho pd_{t+1} + \Delta d_{t+1} - r_{t+1}. \tag{40}$$

By iterating this equation we find:

$$
\begin{aligned}
pd_t & \simeq \kappa + \rho(\kappa + \rho pd_{t+2} + \Delta d_{t+2} - r_{t+2}) + \Delta d_{t+1} - r_{t+1} \\
& = \sum_{j=0}^{\infty} \rho^j \kappa + \rho^{\infty} pd_{\infty} + \sum_{j=1}^{\infty} \rho^{j-1}(\Delta d_{t+j} - r_{t+j}) \\
& = \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1}(\Delta d_{t+j} - r_{t+j}),
\end{aligned}
\tag{41}
$$

assuming that $\rho^{\infty} pd_{\infty} = \lim_{j \to \infty} \rho^j pd_{t+j} = 0$, at least in expectation. Then, we take expectation conditional to time t:

$$
\begin{aligned}
pd_t & \simeq \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} E_t[\Delta d_{t+j} - r_{t+j}] \\
& = \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} E_t[g_{t+j-1} - \mu_{t+j-1}] \\
& = \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^{j} E_t[g_{t+j} - \mu_{t+j}].
\end{aligned}
\tag{42}
$$

Iterating the dynamics of $\hat{\mu}_{t+1}$ and $\hat{g}_{t+1}$ and taking conditional expectation we find

$$E_t[\hat{\mu}_{t+j}] = \delta_1^j \hat{\mu}_t$$

and

$$E_t[\hat{g}_{t+j}] = \gamma_1^j \hat{g}_t.$$

Therefore,

$$
\begin{aligned}
pd_t & \simeq \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j [\gamma_0 + \gamma_1^j \hat{g}_t - \delta_0 - \delta_1^j \hat{\mu}_t] \\
& = \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho} + \frac{\hat{g}_t}{1-\rho\gamma_1} - \frac{\hat{\mu}_t}{1-\rho\delta_1} \\
& = A + B_2 \hat{g}_t - B_1 \hat{\mu}_t.
\end{aligned}
\tag{43}
$$

The explicit expressions for the present-value coefficients A, B_1 and B_2 are the following:

$$
\begin{aligned}
A &= \frac{\kappa + \gamma_0 - \delta_0}{1 - \rho}, \\
B_1 &= \frac{1}{1 - \rho\delta_1}, \\
B_2 &= \frac{1}{1 - \rho\gamma_1}.
\end{aligned}
$$

B   Estimation Methodology

This appendix describes in detail the estimation procedure, first for the benchmark model in Section 2 and then for the extended model in Section 5.

B.1   Benchmark model

For estimation purposes, we cast the model in state-space form, using demeaned state variables $\hat{\mu}_t \equiv \mu_t - \delta_0$ and $\hat{g}_t \equiv g_t - \gamma_0$. We obtain the following linear transition dynamics:

$$
\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \varepsilon^g_{t+1}, \tag{44}
$$

$$
\hat{\mu}_{t+1} = \delta_1 \hat{\mu}_t + \varepsilon^\mu_{t+1}. \tag{45}
$$

The observable variables are dividend growth Δd_{t+1} and the price-dividend ratio pd_{t+1}. Measurement equations for Δd_{t+1} and pd_{t+1} are derived from the model-implied expressions for dividend growth and price-dividend ratio. The measurement equation for dividend growth is given by (5) while log price-dividend ratio is given by (7). Note however that Equation (7) contains no error term, and as shown by Binsbergen and Koijen (2010), this feature can be exploited to reduce the number of transition equations in the model. By substituting the equation for pd_t in the measurement equation for dividend growth, we arrive at a final system with one transition equation, (44), and two measurement equations:

$$
\Delta d_{t+1} = g_t + \varepsilon^d_{t+1}. \tag{46}
$$

$$
pd_{t+1} = (1 - \delta_1)A + B_2(\gamma_1 - \delta_1)\hat{g}_t + \delta_1 pd_t - B_1\varepsilon^\mu_{t+1} + B_2\varepsilon^g_{t+1}. \tag{47}
$$

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:

$$\theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_g, \sigma_\mu, \sigma_d, \rho_{g\mu}, \rho_{\mu d}, \rho_{gd}).$$

We assume that expectation processes are stationary, therefore parameters δ_1 and γ_1 are bounded to be less than one in absolute value. The covariance matrix of the shocks, (6), has to be positive definite, thus σ_g, σ_μ and σ_d are constrained to be positive, while the correlation parameters are between -1 and 1.[31] Rytchkov (2012) shows that it is impossible to identify the whole covariance structure of shocks even when an infinitely long history of returns and dividends is given, but only one element of Σ must be fixed to identify the whole matrix. Thus, for identification purposes, we impose the constraint $\rho_{gd} = 0$, as in Binsbergen and Koijen (2010). Overall the model implies 9 free parameters to estimate. The estimation procedure is the following: We first define an expanded 4-dimensional state vector by the concatenation of the original state variable $\hat{g}$ and the process and observation noise random variables:

$$X_t = \begin{pmatrix} \hat{g}_{t-1} \\ \varepsilon_t^g \\ \varepsilon_t^\mu \\ \varepsilon_t^d \end{pmatrix},$$

which satisfies:

$$X_{t+1} = F X_t + \Gamma \varepsilon_{t+1}^X,$$

where

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

with conditional variance Σ, given in (6). Moreover,

$$F = \begin{bmatrix} \gamma_1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad \Gamma = \begin{bmatrix} 0_{1\times3} \\ I_3 \end{bmatrix},$$

---

[31]Moreover, the condition $\rho_{g\mu}^2 + \rho_{\mu d}^2 + \rho_{gd}^2 < 1$ has to hold for Σ to be positive definite

The measurement equation,

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

is of the form

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

where

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

and

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

The steps of the filter algorithm are the following:

- Initialize with the unconditional mean and covariance of the expanded state:

$$
\begin{aligned}
X_{0,0} &= 0_{4\times 1}, \\
P_{0,0} &= E(X_t X_t').
\end{aligned}
$$

- The time-update equations are

$$
\begin{aligned}
X_{t,t-1} &= F X_{t-1,t-1}, \\
P_{t,t-1} &= F P_{t-1,t-1} F' + \Gamma \Sigma \Gamma',
\end{aligned}
$$

- The prediction error η_t and the variance-covariance matrix of the measurement equations are then:

$$
\begin{aligned}
\eta_t &= Y_t - M_0 - M_1 Y_{t-1} - M_2 X_{t,t-1}, \\
S_t &= M_2 P_{t,t-1} M_2',
\end{aligned}
\tag{48}
$$

where Y_t is the observed value of the measurement equation at time t.

- Update filtering:

$$
\begin{aligned}
\mathcal{K}_t &= P_{t,t-1} M_2' S_t^{-1}, \\
X_{t,t} &= X_{t,t-1} + \mathcal{K}_t \eta_t, \\
P_{t,t} &= (I - K_t M_2) P_{t,t-1},
\end{aligned}
$$

where $\mathcal{K}_t$ is the *Kalman gain*.

To estimate model parameters, θ, we define the log-likelihood for each time t, assuming normally distributed observation errors, as

$$l_t(\theta) = -\frac{1}{2}\log|S_t| - \frac{1}{2}\eta_t' S_t^{-1}\eta_t,$$

where η_t and S_t denote prediction error of the measurement series and the covariance of the measurement series, respectively, obtained from the KF. Model parameters are chosen to maximize the log-likelihood of the data series:

$$\hat{\theta} \equiv \arg\max_{\Theta} \mathcal{L}\left(\theta, \{Y_t\}_{t=1}^T\right), \tag{49}$$

with

$$\mathcal{L}\left(\theta, \{Y_t\}_{t=1}^T\right) = \sum_{t=1}^T l_t(\theta),$$

where T denotes the number of time periods in the sample of estimation.[32]

B.2 Extended Model

In the case of the extended model in Section 5.1, the transition dynamics are the following:

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \gamma_2 \hat{z}_t + \varepsilon_{t+1}^g, \tag{50}$$

$$\hat{\mu}_{t+1} = \delta_1 \hat{\mu}_t + \delta_2 \hat{z}_t + \varepsilon_{t+1}^\mu. \tag{51}$$

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 Equation (33) 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, (50), and three measurement equations:

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \varepsilon_{t+1}^d. \tag{52}$$

$$pd_{t+1} = (1-\delta_1)A + B_2(\gamma_1 - \delta_1)\hat{g}_t + [\gamma_2 B_2 + (\xi_1 - \delta_1)(B_4 - B_3) - \delta_2 B_1]\hat{z}_t +$$
$$+ \delta_1 pd_t - B_1 \varepsilon_{t+1}^\mu + B_2 \varepsilon_{t+1}^g + (B_4 - B_3)\varepsilon_{t+1}^z. \tag{53}$$

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:

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

---

[32]For yearly data, as in our application, T is the number of years in the sample.

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:

$$X_t = \begin{pmatrix} \hat{g}_{t-1} \\ \varepsilon_t^g \\ \varepsilon_t^\mu \\ \varepsilon_t^d \\ \varepsilon_t^z \end{pmatrix},$$

which satisfies:

$$X_{t+1} = FX_t + Bu_{t+1} + \Gamma \varepsilon_{t+1}^X,$$

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

$$\Sigma = \begin{bmatrix} \sigma_g^2 & \sigma_{g\mu} & \sigma_{gd} & \sigma_{gz} \\ \sigma_{g\mu} & \sigma_\mu^2 & \sigma_{\mu d} & \sigma_{\mu z} \\ \sigma_{gd} & \sigma_{\mu d} & \sigma_d^2 & \sigma_{dz} \\ \sigma_{gz} & \sigma_{\mu z} & \sigma_{dz} & \sigma_z^2 \end{bmatrix}. \tag{54}$$

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 previous subsection) apart from a slight change in the time-update equation for the state, which becomes:

$$X_{t,t-1} = F X_{t-1,t-1} + B u_t.$$

C 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:

$$r_{t+1} = \delta_0 + \hat{\mu}_t + \varepsilon_{t+1}^r,$$

but we wrongly assume the following model to hold:

$$r_{t+1} = a_r + b_r pd_t + \tilde{\varepsilon}_{t+1}^r, \tag{55}$$

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 (55). The p-limit of the OLS slope coefficient is the following:[33]

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

where

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

so that

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

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

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

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:

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

while the wrong model is:

$$\Delta d_{t+1} = a_d + b_d pd_t + \tilde{\varepsilon}^d_{t+1}, \tag{56}$$

and we try to estimate the true parameter $b_d = 1/B_2$ from (56). The p-limit of the OLS slope is the following:

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

---

[33]Note that here we denote with $\hat{b}_r$ the OLS estimate of the slope coefficient b_r in (55).

where

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

so that

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

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.

D Asymptotic Validity of the Bootstrap Likelihood Ratio Test

In this appendix we prove the validity of our nonparametric bootstrap likelihood ratio testing procedure, i.e., the equivalence in distribution of LR_T and LR_T^* in equations (10) and (25), respectively, when $B, T \to \infty$, under the null hypothesis H_0. It is well known that if H_0 holds, as $T \to \infty$, LR_T follows a χ_r^2 distribution with r degrees of freedom, where r is the number of parameter constraints defining the null hypothesis H_0. Therefore, we only need to show that also LR_T^* is asymptotically χ_r^2 distributed.

Without loss of generality, let us consider for brevity the case in which the null hypothesis to be tested is formed by zero restrictions, i.e., some of the model parameters are equal to zero. In such cases, the r restrictions can be written as $\theta_2 = 0_{r \times 1}$, where the parameter vector θ is partitioned as $\theta = [\theta_1' \quad \theta_2']'$, possibly after some reordering of the elements, where θ_1 is $(k-r) \times 1$ and θ_2 is $r \times 1$-dimensional.

Let $\hat{\theta}$ be the unconstrained ML estimator of θ, while the pseudo-true value of θ in the population under H_0 is denoted by $\theta^\star = [\theta_1^{\star\prime} \quad 0_{1 \times r}]'$, where $\theta_1^\star$ is the *pseudo-true value* of θ_1, i.e., the maximum of the population expected log likelihood function with respect to θ_1 under the (potentially) incorrect assumption $H_0 : \theta_2 = 0_{r \times 1}$.

Stoffer and Wall (1991) show that nonparametric Monte Carlo bootstrap applied to the (standardized) innovations $\{\hat{e}_t := S_t^{-1/2}(\hat{\theta}) \eta_t(\hat{\theta})\}_{t=1}^T$ yields a distribution of bootstrap

residuals $\{\hat{e}_t^*\}_{t=1}^T$, which can be used to compute a bootstrap distribution of ML estimators $\hat{\theta}^*$:

$$\hat{\theta}^* = \arg\max_{\Theta} \log \mathcal{L}\left(\theta, \{Y_t^*\}_{t=1}^T\right) , \tag{57}$$

where the Monte Carlo sequence $\{Y_t^*\}_{t=1}^T$ is obtained by simulating dynamics (19)-(20) based on bootstrap residuals $\{\hat{e}_t^*\}_{t=1}^T$ (see steps 1)-3) in Section 4.3). Stoffer and Wall (1991) also provide an asymptotic justification of this procedure, showing, under general conditions, the equivalence in distribution of $\sqrt{T}(\hat{\theta}^* - \hat{\theta})$ and $\sqrt{T}(\hat{\theta} - \theta^\star)$ as $B, T \to \infty$, and assuming for simplicity $B = T$. For simplicity of notation we assume that the ML setting holds, but all results hold true with obvious modifications in a PML setting, using sandwich variance-covariance matrix estimators, see Stoffer and Wall (1991):

$$\sqrt{T}(\hat{\theta} - \theta^\star) \xrightarrow{d} N\left(0, \mathcal{I}(\theta^\star)^{-1}\right), \tag{58}$$

where $\mathcal{I}(\theta) = \text{plim}_{T\to\infty} \frac{1}{T} E[-\partial^2 \log \mathcal{L}(\theta)/\partial\theta\partial\theta']$ is the asymptotic information matrix, and

$$\sqrt{T}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{d} N\left(0, \mathcal{I}(\theta^\star)^{-1}\right). \tag{59}$$

The constrained ML estimator $\hat{\theta}_0$ can then be expressed as $\hat{\theta}_0 = \begin{bmatrix} \hat{\theta}_1' & 0_{1\times r} \end{bmatrix}'$, and the asymptotic distribution of $\hat{\theta}_1$ is given by:[34]

$$\sqrt{T}(\hat{\theta}_1 - \theta_1^\star) \xrightarrow{d} N\left(0, \mathcal{I}_{11}(\theta_1^\star)^{-1}\right), \tag{60}$$

where $\mathcal{I}_{11}(.)$ is the $(k-r)\times(k-r)$ top left block of the asymptotic information matrix $\mathcal{I}(.)$ of the unrestricted model. Analogously, the constrained bootstrap Maximum Likelihood estimator $\hat{\theta}_0^*$ can be partitioned as $\hat{\theta}_0^* = \begin{bmatrix} \hat{\theta}_1^{*\prime} & 0_{1\times r} \end{bmatrix}'$ and its asymptotic distribution is given by:

$$\sqrt{T}(\hat{\theta}_1^* - \hat{\theta}_1) \xrightarrow{d} N\left(0, \mathcal{I}_{11}(\theta_1^\star)^{-1}\right). \tag{61}$$

For ease of notation, let us denote by $l(\theta, y)$ the log-likelihood of the model, i.e. $l(\theta, y) \equiv \log \mathcal{L}\left(\theta, \{Y_t\}_{t=1}^T\right)$. Using a second order Taylor expansion around $\hat{\theta}^*$, the bootstrap log-likelihood $l(\hat{\theta}_0^*, y^*)$ can be written as

$$l(\hat{\theta}_0^*, y^*) = l(\hat{\theta}^*, y^*) - \frac{1}{2}(\hat{\theta}_0^* - \hat{\theta}^*)'H(\bar{\theta})(\hat{\theta}_0^* - \hat{\theta}^*). \tag{62}$$

---

[34]See e.g. Davidson and MacKinnon (1999a), chapter 10.

where $H(.)$ is the Hessian matrix,[35] and $\bar{\theta} \in (\hat{\theta}_0^*, \hat{\theta}^*)$. Using (62), the bootstrap likelihood ratio statistics LR_T^* in (25) becomes

$$
\begin{aligned}
LR_T^* &= 2\left(l(\hat{\theta}^*, y^*) - l(\hat{\theta}_0^*, y^*)\right) \\
&= -(\hat{\theta}_0^* - \hat{\theta}^*)'H(\bar{\theta})(\hat{\theta}_0^* - \hat{\theta}^*).
\end{aligned}
$$

Consistency of $\hat{\theta}^*$ implies consistency of $\bar{\theta}$, and using information matrix inequality[36] we get:

$$
LR_T^* \overset{a}{=} T(\hat{\theta}_0^* - \hat{\theta}^*)'\mathcal{I}(\theta^\star)(\hat{\theta}_0^* - \hat{\theta}^*). \tag{63}
$$

Let we now define the score vector $g(\theta, y)$ of first derivatives of $l(\theta, y)$ with respect to the elements of θ,[37] and the asymptotic score vector $s \equiv plim T^{-1/2}g(\theta^*, y)$. From a Taylor expansion of the likelihood equation $g(\hat{\theta}^*, y^*) = 0$ we obtain the following asymptotic equalities:

$$
\begin{aligned}
T^{1/2}(\hat{\theta}^* - \hat{\theta}) &\overset{a}{=} \mathcal{I}^{-1}T^{-1/2}g(\theta^\star) \\
T^{1/2}(\hat{\theta}_1^* - \hat{\theta}_1) &\overset{a}{=} \mathcal{I}_{11}^{-1}T^{-1/2}g_1(\theta^\star),
\end{aligned}
$$

which can be used to eliminate the estimators in (63) when we take the limit, obtaining an expression that involves only asymptotic information matrix and asymptotic score vector, as follows:

$$
\begin{aligned}
\operatorname{plim} T^{1/2}(\hat{\theta}^* - \hat{\theta}_0^*) &= \operatorname{plim} T^{1/2}(\hat{\theta}^* - \hat{\theta}) - \operatorname{plim} T^{1/2}(\hat{\theta}_0^* - \hat{\theta}) \\
&= \mathcal{I}^{-1}s - \mathcal{I}_{11}^{-1}s_1 \\
&= \mathbf{J}s, \tag{64}
\end{aligned}
$$

where s_1 is the subvector of s that corresponds to θ_1, and

$$
\mathbf{J} \equiv \mathcal{I}^{-1} - \begin{bmatrix} \mathcal{I}_{11}^{-1} & 0_{(k-r)\times r} \\ 0_{r\times(k-r)} & 0_{r\times r} \end{bmatrix}. \tag{65}
$$

---

[35]The $k \times k$ matrix of second derivatives of the log-likelihood with respect to θ.

[36]Let the asymptotic Hessian matrix be defined as $\mathcal{H}(\theta) \equiv plim\frac{1}{T}H(\theta)$. The information matrix equality, which assumes correct specification of the model, implies that $\mathcal{I}(\theta) = -\mathcal{H}(\theta)$.

[37]In the same way, $g_1(\theta, y)$ is the subvector of first derivatives of $l(\theta, y)$ with respect to the elements of θ_1

Using (64), the probability limit of LR_T^* for $T \to \infty$ becomes:

$$\text{plim } LR_T^* = s' \mathbf{J} \mathcal{I} \mathbf{J} s. \tag{66}$$

Moreover, from (65), we have that

$$\mathcal{I}\mathbf{J} = I_k - \begin{bmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{bmatrix} \begin{bmatrix} \mathcal{I}_{11}^{-1} & 0_{(k-r)\times r} \\ 0_{r\times(k-r)} & 0_{r\times r} \end{bmatrix} = \begin{bmatrix} 0_{(k-r)\times(k-r)} & 0_{(k-r)\times r} \\ -\mathcal{I}_{21}\mathcal{I}_{11}^{-1} & I_r \end{bmatrix} \equiv \mathbf{Q}, \tag{67}$$

which implies $\mathcal{I}^{-1}\mathbf{Q} = \mathbf{J}$, $\mathbf{J}\mathbf{Q} = \mathbf{J}$ and $\mathbf{J}\mathcal{I}\mathbf{J} = \mathbf{J}$, from which we conclude that (66) can be written as

$$\text{plim } LR_T^* = s'\mathbf{J}s. \tag{68}$$

Now, notice that s is asymptotically $N(0,\mathcal{I})$, thus $s = \mathcal{I}^{1/2}\tilde{s}$, where $\tilde{s}$ is asymptotically standard normal. Therefore, (68) can be written as

$$\text{plim } LR_T^* = \tilde{s}'\mathcal{I}^{1/2}\mathbf{J}\mathcal{I}^{1/2}\tilde{s}, \tag{69}$$

which is χ^2 distributed with degrees of freedom equal to the rank of matrix $\mathcal{I}^{1/2}\mathbf{J}\mathcal{I}^{1/2}$:

$$r(\mathcal{I}^{1/2}\mathbf{J}\mathcal{I}^{1/2}) = r(\mathcal{I}^{-1/2}\mathbf{Q}\mathcal{I}^{1/2}) = r,$$

using the fact that $\mathcal{I}$ has full rank and that the rank of $\mathbf{Q}$ is r since its first $k - r$ rows are zero. Therefore, we can conclude that $LR_T^* \xrightarrow{d} \chi_r^2$, as we wanted to show.

E   Bootstrap Distribution of out-of-sample R-squares

The distribution of the out-of-sample R-squares of returns and dividend growth under the null hypothesis H_0 is computed based on the following algorithm:

1) Using the estimated model parameters obtained using the first T years of data, under the null hypothesis H_0, denoted $\hat{\theta}_{T,0}$, construct the (constrained) time series of standardized innovations $\{\hat{e}_{0t}\}_{t=1}^T$, and a bootstrap sample of observations, $\{Y_t^*\}_{t=1}^T$ as in steps 1)-3) in Section 4.3.

2) Using bootstrap sample $\{Y_t^*\}_{t=1}^T$, compute unconstrained maximum likelihood point estimates $\hat{\theta}_T^*$, by maximizing the log likelihood function $\log \mathcal{L}\left(\theta, \{Y_t^*\}_{t=1}^T\right)$, without imposing null hypothesis H_0.

3) Based on estimated parameters $\hat{\theta}_T^*$ and filtered state using data until time T, compute the expected return and dividend growth for year $T+1$, $\tilde{\mu}_T$ and $\tilde{g}_T$, respectively.

4) Repeat steps 1)-3) for $T = T_{in}, \ldots, T_{max} - 1$, where T_{in} is the minimum length of the in-sample period and T_{max} is the length of the full sample of data.[38]

5) The out-of-sample R^2 statistics for returns and dividend growth are computed as

$$
\begin{aligned}
R^2_{Ret,OS} &= 1 - \frac{\sum_{T=T_{in}}^{T_{max}-1} \left(r_{T+1} - \tilde{\mu}_T \right)^2}{\sum_{T=T_{in}}^{T_{max}-1} \left(r_{T+1} - \bar{r}_T \right)^2}, \\
R^2_{Div,OS} &= 1 - \frac{\sum_{T=T_{in}}^{T_{max}-1} \left(\Delta d_{T+1} - \tilde{g}_T \right)^2}{\sum_{t=T_{in}}^{T_{max}-1} \left(\Delta d_{T+1} - \overline{\Delta d}_T \right)^2},
\end{aligned}
$$

where $\bar{r}_T$ and $\overline{\Delta d}_T$ are historical means or returns and dividend growth up until time T.

6) Repeat steps 1)-5) a large number of times, B, to obtain a collection of bootstrap values of the out-of-sample R^2 statistics. The empirical distribution of these values provides an approximation of the distribution of the $R^2_{Ret,O}$ and $R^2_{Div,OS}$ statistics under the null hypothesis H_0.

This procedure borrows from Rodriguez and Ruiz (2009), who show how to to compute nonparametric bootstrap prediction intervals in state space models, while taking into account the uncertainty linked to parameter estimation and not resorting to parametric assumptions for the shock distribution in the model.

---

[38]We start our out-of-sample computations in 1985, which means that the first estimation is done using $T_{in} = 40$ years of data, and the length of the full sample in our case is $T_{max} = 65$ years.

F Tables and Figures

Table 1: Results of the estimation of the present-value model in Section 2. The model is estimated by maximum likelihood, using yearly data from 1946 to 2010 on log dividend growth rates and log price-dividend ratio. Panel A presents estimates of the coefficients of the underlying processes. Panel B reports resulting coefficients of the present-value decomposition $pd_t = A - B_1\hat{\mu}_t + B_2\hat{g}_t$. Bootstrapped standard errors are in parentheses.

Panel A: Maximum likelihood estimates				
γ_0	δ_0	γ_1	δ_1	
0.057	0.083	0.304	0.927	
(0.009)	(0.010)	(0.337)	(0.089)	
σ_g	σ_μ	σ_D	$\rho_{g,\mu}$	$\rho_{\mu,D}$
0.065	0.015	0.002	0.231	-0.972
(0.023)	(0.024)	(0.028)	(0.419)	(0.606)

Panel B: Implied present-value parameters			
ρ	A	B_1	B_2
0.974	3.637	10.332	1.421
(0.004)	(0.140)	(2.418)	(6.396)

Table 2: Constrained ML estimates of the present-value model and LR statistics for the tests of constant expected returns ($H_0 : \delta_1 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = 0$), constant expected dividend growth ($H_0 : \gamma_1 = \sigma_g = \rho_{g\mu} = 0$) and equal autoregressive parameters ($H_0 : \delta_1 = \gamma_1$). The first column reports the results of the unconstrained estimation, from Table 1. $LogL$ denotes the pseudo log-likelihood obtained, LR is the value of the Likelihood Ratio statistic computed using (10), $p-value$ denotes percentage p-values of the tests, and the last row reports finite-sample sizes of the tests, in percentage.

	Unconstrained	$H_0 : \delta_1 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = 0$	$H_0 : \gamma_1 = \sigma_g = \rho_{g\mu} = 0$	$H_0 : \delta_1 = \gamma_1$
γ_0	0.057	0.072	0.055	0.054
δ_0	0.083	0.079	0.082	0.081
γ_1	0.304	0.996	0	0.926
δ_1	0.927	0	0.903	0.926
σ_g	0.065	0.002	0	0.004
σ_μ	0.015	0	0.021	0.021
σ_d	0.002	0.069	0.068	0.068
$\rho_{g\mu}$	0.231	0	0	0.950
$\rho_{\mu d}$	-0.972	0	0.357	0.312
$LogL$	230.84	215.91	224.33	224.79
LR		29.87	13.02	12.11
$p-value$ (%)		0.00	0.46	0.05
$size$ (%)		60.5	25.8	26.40

Table 3: Results of the bootstrap LR tests of constant expected returns ($H_0 : \delta_1 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = 0$), constant expected dividend growth ($H_0 : \gamma_1 = \sigma_g = \rho_{g\mu} = 0$) and equal autoregressive parameters ($H_0 : \delta_1 = \gamma_1$). $p-value$ denotes percentage p-values of the tests.

	$H_0 : \delta_1 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = 0$	$H_0 : \gamma_1 = \sigma_g = \rho_{g\mu} = 0$	$H_0 : \delta_1 = \gamma_1$
$p-value$ (%)	0.5	9.5	2.4

Table 4: Panel A reports estimation results of the present-value model in Section 5.1, using as predictor variables the book-to-market ratio (BM), stock variance ($SVAR$), CAY and the BAA corporate bond yield (BAA), respectively. The models are estimated using annual data from 1946 to 2010. Panel B reports the model-implied R-squared values for return and dividend growth, in percentage, computed as in (8)-(9), while Panel C reports R-squared from standard OLS predictive regressions of returns and dividend growth on lagged price-dividend ratio and each predictive variable z_t:

$$r_{t+1} = a_r + b_r pd_t + \gamma_r z_t + \epsilon^r_{t+1}$$
$$\Delta d_{t+1} = a_d + b_d pd_t + \gamma_d z_t + \epsilon^d_{t+1}.$$

	BM	$SVAR$	CAY	BAA
Panel A: Maximum-likelihood estimates				
γ_0	0.051	0.056	0.050	0.057
δ_0	0.071	0.129	0.077	0.090
γ_1	0.234	0.475	0.338	0.296
δ_1	0.878	0.993	0.926	0.920
σ_g	0.064	0.078	0.066	0.065
σ_μ	0.016	0.018	0.031	0.018
σ_d	0.013	0.018	0.015	0.008
$\rho_{g\mu}$	0.220	-0.454	-0.308	0.177
$\rho_{\mu d}$	-0.144	-0.758	-0.596	-0.167
ξ_0	0.487	0.020	0	0.082
ξ_1	0.913	0.418	0.733	0.939
ρ_{gz}	-0.298	-0.764	-0.528	-0.250
$\rho_{\mu z}$	0.567	0.804	0.882	0.485
ρ_{dz}	0.597	-0.594	-0.775	0.702
σ_z	0.102	0.021	0.014	0.010
δ_2	0.019	-0.285	-0.396	-0.025
γ_2	0.067	1.652	0.837	-0.014
Panel B: Model-implied R-squared				
R^2_{ret}	10.13	9.70	17.58	9.65
R^2_{div}	22.32	25.71	19.81	18.29
Panel C: Predictive regression R-squared				
R^2_{ret}	10.34	12.24	15.40	13.81
R^2_{div}	4.26	8.02	0.95	3.03

Table 5: Test of no dividend growth predictability in the context of the present-value model in Section 5.1:

$$H_0 : \gamma_1 = \gamma_2 = \sigma_g = \rho_{g\mu} = \rho_{gz} = 0.$$

Panel A reports constrained estimation results, using as predictor variables the book-to-market ratio (BM) and stock variance $(SVAR)$, respectively. The models are estimated using annual data from 1946 to 2010. Panel B reports the p-values of the test, using the asymptotic distribution of the LR statistic, and the effective size of the asymptotic test, for a nominal size $\alpha = 5\%$, while Panel C reports the p-values of the bootstrap test. Finite sample size computations and bootstrap tests are based on 1000 bootstrap samples.

|	BM	$SVAR$
Panel A: Constrained Maximum-likelihood estimates		
γ_0	0.056	0.055
δ_0	0.078	0.094
γ_1	0	0
δ_1	0.887	0.960
σ_g	0	0
σ_μ	0.019	0.031
σ_d	0.068	0.068
$\rho_{g\mu}$	0	0
$\rho_{\mu d}$	0.382	-0.029
ξ_0	0.504	0.021
ξ_1	0.904	0.363
ρ_{gz}	0	0
$\rho_{\mu z}$	0.681	0.927
ρ_{dz}	0.136	-0.216
σ_z	0.101	0.022
δ_2	0.006	-0.701
γ_2	0	0
Panel B: Asymptotic test		
$p-value$ $(\%)$	0.22	0.19
$size$ $(\%)$	22.70	28.60
Panel C: Bootstrap test		
$p-value$ $(\%)$	7.40	10.20

Table 6: Test of no return predictability in the context of the present-value model in Section 5.1:

$$H_0 : \delta_1 = \delta_2 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = \rho_{\mu z} = 0.$$

Panel A reports constrained estimation results, using as predictor variables the book-to-market ratio (BM) and CAY, respectively. The models are estimated using annual data from 1946 to 2010. Panel B reports the p-values of the test, using the asymptotic distribution of the LR statistic, and the effective size of the asymptotic test, for a nominal size $\alpha = 5\%$, while Panel C reports the p-values of the bootstrap test. Finite sample size computations and bootstrap tests are based on 1000 bootstrap samples.

|	BM	CAY
Panel A: Constrained Maximum-likelihood estimates		
γ_0	0.070	0.072
δ_0	0.075	0.080
γ_1	0.960	0.996
δ_1	0	0
σ_g	0.008	0.003
σ_μ	0	0
σ_d	0.067	0.069
ξ_0	0.234	-0.006
ξ_1	0.997	0.857
ρ_{gz}	0.811	-0.796
$\rho_{\mu z}$	0	0
ρ_{dz}	0.215	-0.253
σ_z	0.104	0.014
δ_2	0	0
γ_2	-0.001	0.015
Panel B: Asymptotic test		
$p-value$ (%)	0.01	0.00
$size$ (%)	14.10	28.10
Panel C: Bootstrap test		
$p-value$ (%)	0.50	0.30

Figure 1: 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.
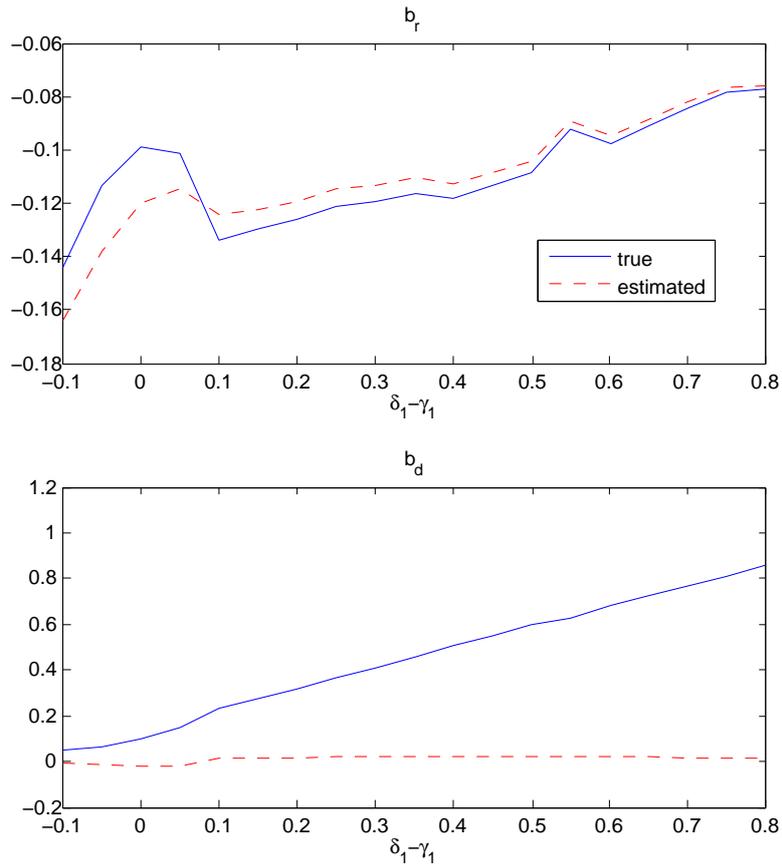
Figure 2: Percentage R-squared of returns (dashed red line) and dividend growth (solid blue line), implied by a simple present-value model as a function of the difference between the autoregressive coefficients in the dynamics of expected returns (δ_1) and expected dividend growth (γ_1). Horizontal lines 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.
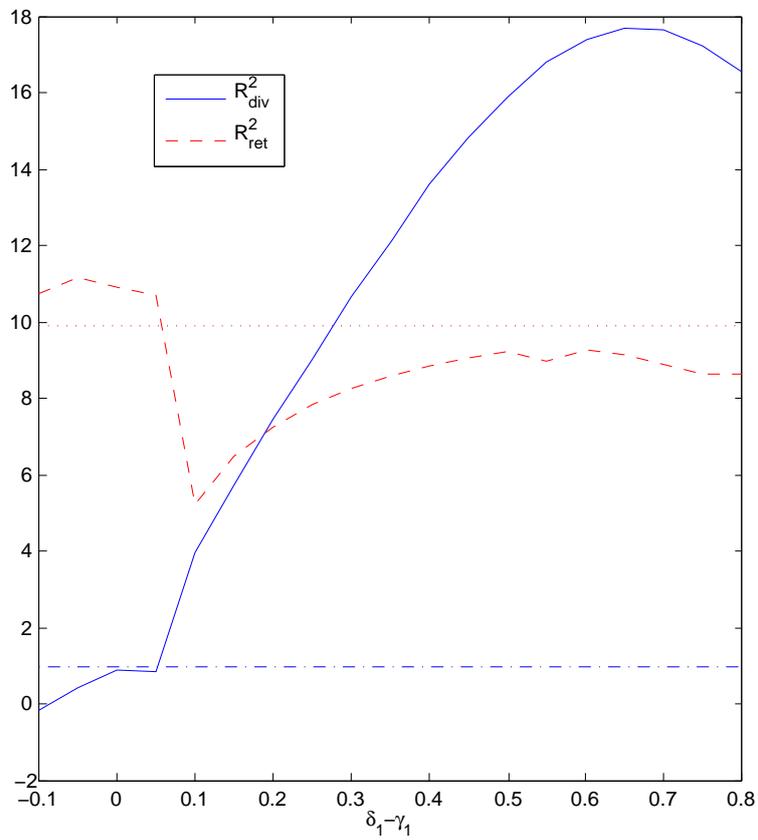
Figure 3: First (second) panel displays the quantiles of the empirical distribution of the LR statistics for the tests of constant expected returns (dividend growth), while third panel shows the quantiles of the empirical distribution of the LR statistics for the test of equal persistence parameters, all obtained through a nonparametric bootstrap simulation procedure, against the quantiles of the asymptotic chi-squared distribution of the statistics (dotted red line). The vertical dotted line denotes the 95% quantile of this distribution.
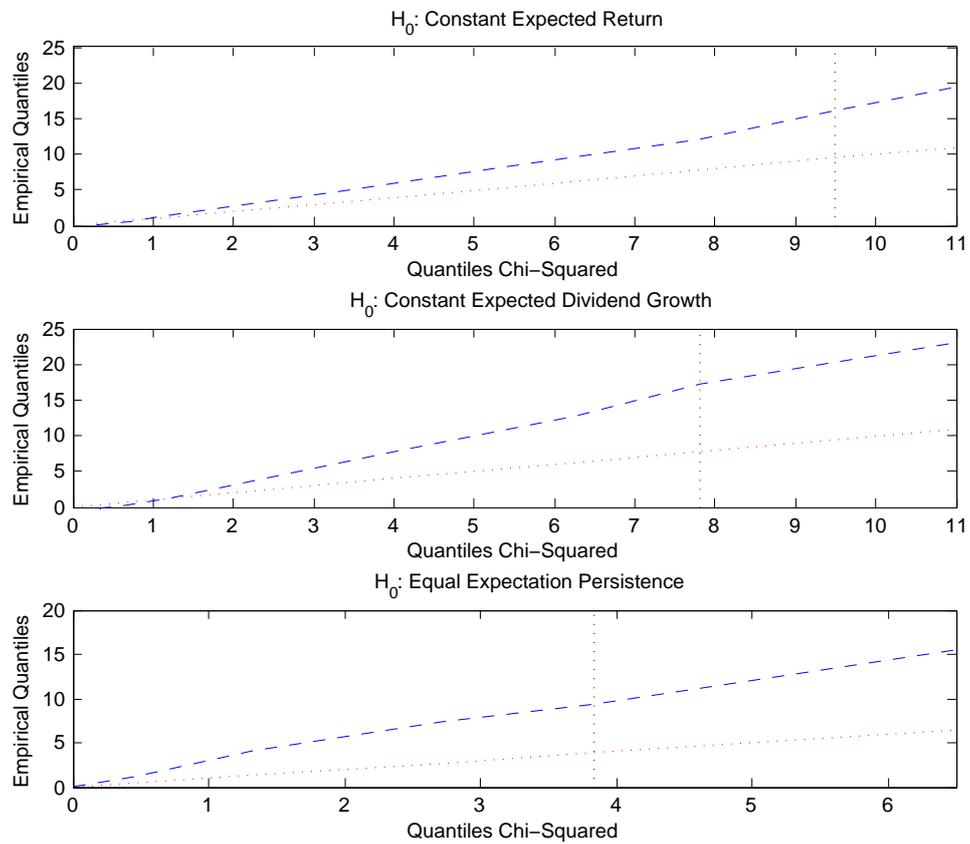
Figure 4: Bootstrapped distribution of the R-squared of returns (upper panels) and dividend growth (lower panels), starting from the estimates under the constraint of constant expected dividend growth ($\gamma_1 = \sigma_g = \rho_{g\mu} = 0$, left panels) and of equal persistence parameters ($\delta_1 = \gamma_1$, right panels). Vertical red lines and dashed black lines denote R-squared from constrained and unconstrained estimations on real data, respectively. Distributions are based on 1000 bootstrap samples.
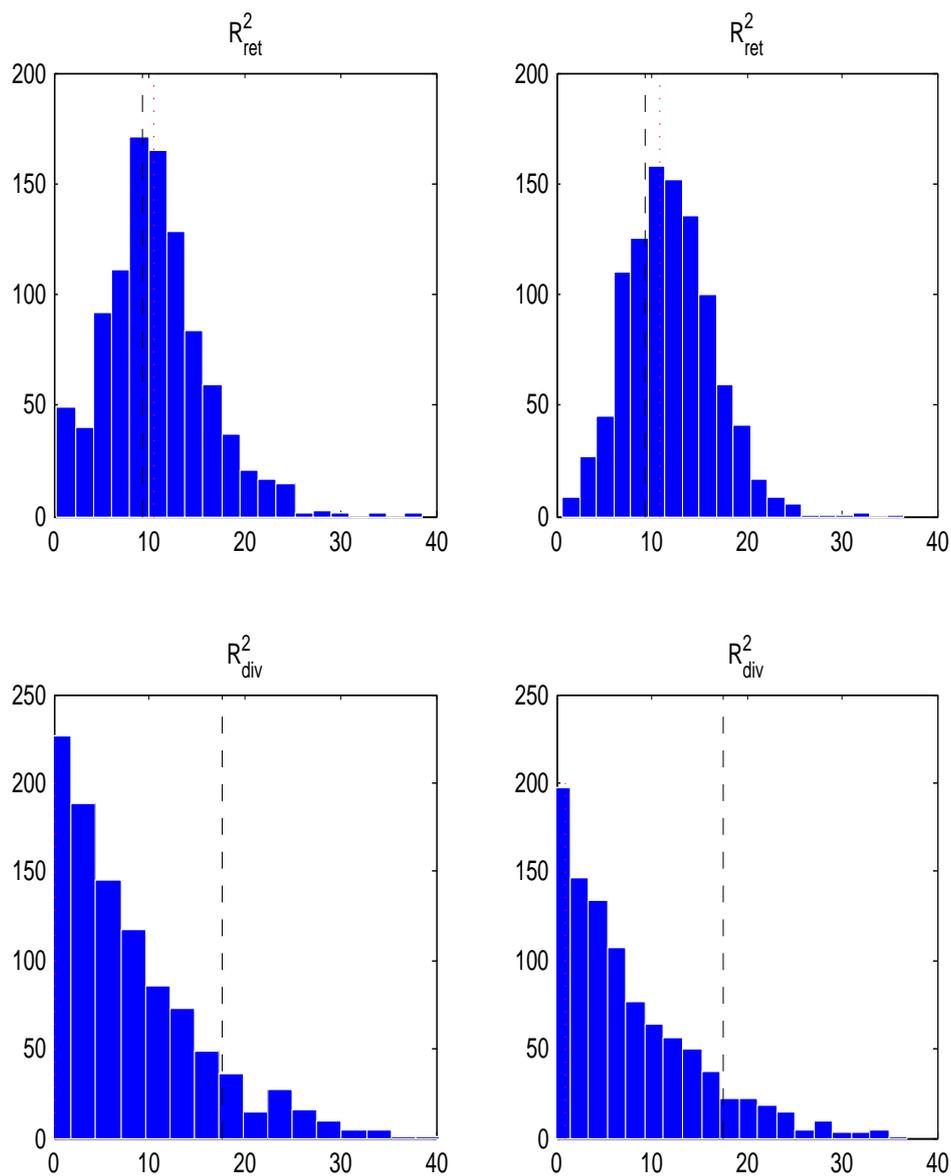
Figure 5: Bootstrapped distribution of the out-of-sample R-squared of returns (upper panel) and dividend growth (lower panel), starting from the estimates under the constraint of constant expected dividend growth ($\gamma_1 = \sigma_g = \rho_{g\mu} = 0$). Vertical red lines and dashed black lines denote out-of-sample R-squared from constrained and unconstrained estimations on real data, respectively. Distributions are based on 1000 bootstrap samples.
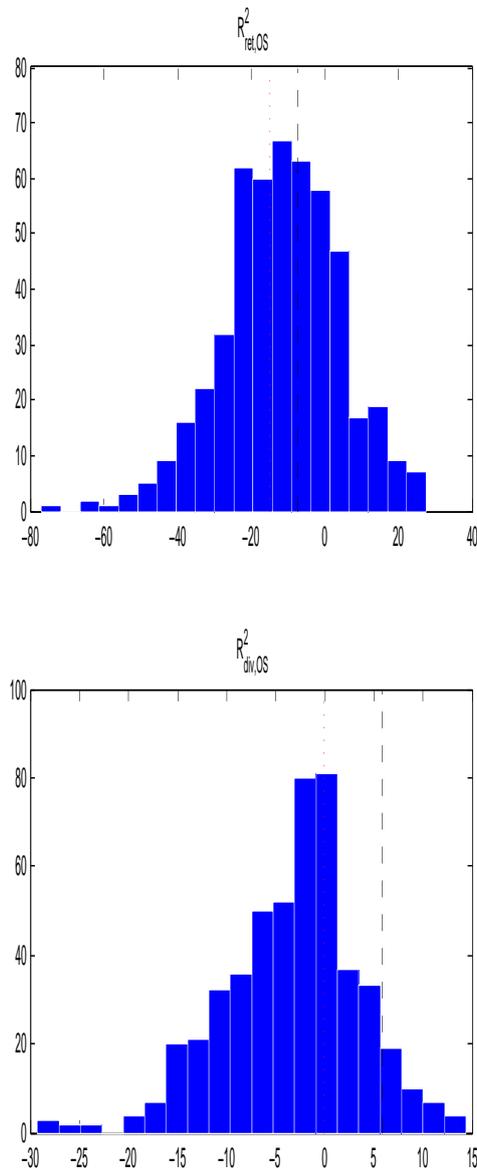
Figure 6: Bootstrapped distribution of the R-squared of returns (upper panels) and dividend growth (lower panels) implied by the present-value model with additional predictors (*bm* in the left panels and *svar* in the right panels), starting from the estimates under the constraint of constant expected 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.