

# Dividend Growth Does Not Help Predict Returns

## Compared To Likelihood-Based Tests:

### An Anatomy of the Dog\*

Erik Hjalmarsson<sup>†</sup>

Tamás Kiss<sup>‡</sup>

April 21, 2020

#### Abstract

The dividend-growth based test of return predictability, proposed by Cochrane [2008, Review of Financial Studies 21, 1533-1575], is similar to a likelihood-based test of the standard return-predictability model, treating the autoregressive parameter of the dividend-price ratio as known. In comparison to standard OLS-based inference, both tests can achieve power gains by using restrictions or prior information on the value of the autoregressive parameter. When compared to the likelihood-based test, there are no power advantages for the dividend-growth based test. In common implementations, with the autoregressive parameter set equal to the corresponding OLS estimate, Cochrane's test suffers from severe size distortions.

*JEL classification:* C22, G1.

*Keywords:* Predictive regressions; Present-value relationship; Stock-return predictability.

---

\*We would like to thank an anonymous referee and the editor, Ivo Welch, for their valuable comments. We have also benefited from comments by Daniel Buncic, Paolo Giordani, Randi Hjalmarsson, Ron Kaniel, Jesper Lindé, Anders Wilhelmsson, Marcin Zamojski, Xin Zhang, and Pär Österholm, as well as by seminar participants at Sveriges Riksbank, Örebro University, the European Summer Meeting of the Econometric Society, Cologne 2018, and the Conference on Computational and Financial Econometrics, London, 2019. Hjalmarsson gratefully acknowledges financial support from the Nasdaq Nordic Foundation.

<sup>†</sup>Department of Economics and the Centre for Finance, University of Gothenburg. Contact information: P.O. Box 640, SE 405 30 Gothenburg, Sweden. Email: erik.hjalmarsson@economics.gu.se.

<sup>‡</sup>School of Business, Örebro University. Email: tamas.kiss@oru.se.

# 1 Introduction

A standard empirical framework for testing return predictability consists of a predictive regression for returns and an autoregressive (AR) process for the dividend-price ratio. In an influential study, Cochrane (2008) adds a predictive regression for the dividend-growth rate to this standard model, and links the three equations through the Campbell and Shiller (1988) present value identity. Cochrane’s key insight is that under the identity that links returns, prices, and dividends, a given degree of predictability in returns corresponds exactly to some degree of predictability in dividend growth. Thus, if one is testing the null hypothesis of no return predictability, using the dividend-price ratio as predictor, this null hypothesis has an exact translation in terms of predictability in dividend growth.<sup>1</sup>

In this paper, we show that Cochrane’s dividend-growth based test is very similar to a test based on the full information maximum likelihood (ML) estimator for the standard return and dividend-price ratio system, with the AR parameter of the dividend-price ratio treated as known. Cochrane’s test can thus be viewed as an economically motivated (approximate) derivation of the ML estimator in the case of a known AR parameter. This finding explains why the dividend-growth based test, as implicitly formulated by Cochrane, can be more powerful than the return-based test using the simple OLS estimator. However, if one were to use the same assumptions when formulating the return-based test, one could use an ML procedure that (asymptotically) dominates the dividend-growth based test. That is, if one compares testing approaches based on the same information set, or the same set of assumptions—treating the AR parameter as known and equal to

---

<sup>1</sup>Cochrane’s proposed modelling framework and testing approach has received great interest in the profession, and there is now a host of papers that evaluates both return predictability as well as dividend-growth predictability. Papers explicitly using Cochrane’s (2008) approach to test for return predictability include Chen (2009); Engsted and Pedersen (2010); Golez and Koudijs (2018). Related works that consider both return and dividend-growth predictability include, among others, Lettau and Ludvigson (2005); Binsbergen and Koijen (2010); Lettau and Ludvigson (2010); Chen et al. (2012); Kelly and Pruitt (2013); Golez (2014); Bollerslev et al. (2015); Maio and Santa-Clara (2015). Koijen and Van Nieuwerburgh (2011) provides a review on return and dividend-growth predictability.

some given value—there are no apparent gains from using the dividend-growth regression in testing for return predictability.<sup>2</sup>

The ML procedure is extremely sensitive to the specific value used for the AR parameter of the dividend-price ratio, and by implication so is Cochrane’s procedure. Specifically, the choice to treat the OLS estimate of the AR parameter as the “true” value, which appears to have been adopted in subsequent empirical studies (e.g., Chen, 2009; Engsted and Pedersen, 2010; Golez and Koudijs, 2018), leads to severe size distortions. A test with a nominal size of five percent is shown to have actual rejection rates in excess of 20 percent under the null hypothesis. On the other hand, if the value of the AR parameter is set high enough, such that it is greater than or equal to the true parameter value in the data, the size of the resulting test can be controlled. For instance, under an assumption that the dividend-price ratio is stationary, setting the AR parameter equal to unity would ensure that it is greater than the true parameter value in the data. In this case, Cochrane’s test becomes similar to the conservative sup-bound test developed in Lewellen (2004) and analyzed further in Campbell and Yogo (2006).

To form some intuition for our results, note that Cochrane’s predictive model is made up of three regressions. The regression equations are linked together by the Campbell and Shiller (1988) present value identity, which implies an exact relationship among the slope coefficients as well as the error terms in the three regressions. Thus, any one of the three equations is redundant in the model formulation and the standard bi-variate predictive system, consisting of a predictive regression for returns and an AR process for the dividend-price ratio, must contain exactly the same information as the tri-variate system. Our results essentially confirm this basic intuition: Adding a fully redundant equation to a regression system should not lead to any statistical gains. From the perspective of

---

<sup>2</sup>Within the same framework, Cochrane also discusses tests of long-run predictability. We do not consider the properties of these tests here, as their formulation is based on a re-scaled version of the return coefficient, and does not explicitly rely on inference in the dividend-growth regression.

empirical research, our findings imply that if one is interested in testing for *return* predictability, there is no extra information available in the dividend-growth equation, and equally or more powerful tests can be formulated from the standard predictive regression setup. Cochrane’s dividend-growth based test—properly implemented such that its size is controlled—can achieve power gains over plain OLS inference on the return equation alone. But not over ML-based tests that make use of the same assumptions or information on the AR parameter of the dividend-price ratio.<sup>3</sup>

An empirical application to aggregate U.S. stock returns illustrates our main theoretical results. Specifically, we show that the dividend-growth based test provides very similar results to an ML-based test, and that both tests depend strongly upon the assumption on the maximum feasible value for the AR parameter in the dividend-price ratio. If one is not willing to impose any stronger assumption than stationarity of the dividend-price ratio, both tests fail to reject the null hypothesis of no return predictability at the five percent significance level. On the other hand, if one is willing to assume that the AR parameter in the dividend-price ratio is below about 0.97 in annual data, the evidence would point in favor of return predictability. Seemingly small changes in the assumptions on the AR parameter can thus lead to rather drastic changes in inference.

## 2 Testing return predictability

### 2.1 Model formulation

Our predictive model is identical to the one used by Cochrane (2008). Let  $r_t$  denote the log-returns from period  $t-1$  to  $t$ ,  $d_t$  the time  $t$  log-dividends, and  $d_t - p_t$  the corresponding

---

<sup>3</sup>To be clear, we make no claim to provide an exhaustive analysis of the relative merits of the many inferential methods that exist for predictive regressions (recent examples include Chen and Deo, 2009, Phillips and Chen, 2014, and Kostakis et al., 2015). Rather, we focus solely on the properties of the dividend-growth based procedure proposed by Cochrane (2008), and how it relates to likelihood-based inference.

log dividend-price ratio. The joint model of return and dividend-growth predictability is formulated as the following restricted first-order VAR system,

$$r_t = \alpha_r + \beta_r (d_{t-1} - p_{t-1}) + \epsilon_t^r, \quad (\text{E1})$$

$$\Delta d_t = \alpha_d + \beta_d (d_{t-1} - p_{t-1}) + \epsilon_t^d, \quad (\text{E2})$$

$$d_t - p_t = \alpha_{dp} + \phi (d_{t-1} - p_{t-1}) + \epsilon_t^{dp}. \quad (\text{E3})$$

By Campbell and Shiller (1988), the following (approximate) present value identity holds,

$$r_t = \rho (p_t - d_t) + \Delta d_t - (p_{t-1} - d_{t-1}). \quad (1)$$

The identity is obtained through a log-linearization of returns around the long-run mean of the dividend-price ratio, denoted by  $\rho$  and empirically defined as

$$\rho = \frac{e^{-\overline{(d-p)}}}{1 + e^{-\overline{(d-p)}}}, \quad (2)$$

where  $\overline{d-p}$  is the average dividend-price ratio. The parameter  $\rho$  is subsequently treated as a fixed and “known” quantity. In the CRSP data used by Cochrane,  $\rho = 0.9638$ , and this is the value that we use throughout this study as well.

The present value identity in equation (1) implies the following restrictions on the coefficients and error terms in the predictive equations (E1)-(E3),

$$\beta_r = \beta_d + (1 - \rho\phi), \quad (\text{R1})$$

and

$$\epsilon_t^r = \epsilon_t^d - \rho\epsilon_t^{dp}. \quad (\text{R2})$$

The restrictions in (R1) and (R2) imply that any one of the three model equations is

redundant, and an equivalent model formulation would be retained by dropping any one of the equations (E1)-(E3).<sup>4</sup>

Let  $\epsilon_t = (\epsilon_t^r, \epsilon_t^d, \epsilon_t^{dp})'$  denote the vector of mean zero innovations, and let  $\Sigma \equiv E[\epsilon_t \epsilon_t']$  be the covariance matrix for  $\epsilon_t$ , where  $\sigma_{ij}$ ,  $i, j = 1, 2, 3$ , denotes the elements of  $\Sigma$ . By restriction (R2), the covariance matrix  $\Sigma$  can be written as

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{11} + \rho\sigma_{13} & \sigma_{13} \\ \sigma_{11} + \rho\sigma_{13} & \sigma_{11} + \rho^2\sigma_{33} + 2\rho\sigma_{13} & \sigma_{13} + \rho\sigma_{33} \\ \sigma_{13} & \sigma_{13} + \rho\sigma_{33} & \sigma_{33} \end{bmatrix}. \quad (3)$$

It is easily seen that  $\Sigma$  is not full rank, reflecting the redundancy in the equation system (E1)-(E3).

## 2.2 Standard OLS-based inference

Cochrane's (2008) key idea is that the absence of return predictability must imply the presence of dividend-growth predictability. That is, the coefficient restriction (R1) implies that if  $\beta_r = 0$ ,

$$\beta_d = -(1 - \rho\phi) < 0, \quad (4)$$

provided  $\rho\phi < 1$ . Thus, under the assumption of  $\phi < 1$  (or at a minimum  $\rho\phi < 1$ ), it follows that  $\beta_r = 0 \Rightarrow \beta_d = -(1 - \rho\phi) \neq 0$  and  $\beta_d = 0 \Rightarrow \beta_r = (1 - \rho\phi) \neq 0$ . Cochrane therefore suggests that instead of testing just the usual simple null,  $\beta_r = 0$ , one should also test whether  $\beta_d = -(1 - \rho\phi)$ . Thus, the "joint" null hypothesis can be formulated

---

<sup>4</sup>The identity in equation (1) is only approximate, and the restrictions stated in equations (R1) and (R2) are therefore also approximate. However, as shown by Cochrane (2008), as well as in many subsequent papers (e.g., Binsbergen and Koijen, 2010; Koijen and Van Nieuwerburgh, 2011; Engsted et al., 2012; Kelly and Pruitt, 2013), the restrictions in (R1) and (R2) hold very closely empirically, and we will therefore treat them as exact throughout the paper. This is also in line with how Cochrane deals with them, and all his simulation results make explicit use of their exact identity.

as

$$H_0 : \beta_r = 0 \quad \text{and} \quad \beta_d + (1 - \rho\phi) = 0. \quad (5)$$

Under the maintained model specification, the coefficient restriction  $\beta_r = \beta_d + (1 - \rho\phi)$  is exact and the restriction also applies to the estimated values (see Appendix A), such that

$$\hat{\beta}_{r,LS} = \hat{\beta}_{d,LS} + (1 - \rho\hat{\phi}_{LS}), \quad (6)$$

where  $\hat{\beta}_{r,LS}$ ,  $\hat{\beta}_{d,LS}$ , and  $\hat{\phi}_{LS}$  are the OLS estimators of the corresponding parameters. Since this equality holds numerically for the OLS estimates in any sample, the distribution of  $\hat{\beta}_{r,LS}$  and  $\hat{\beta}_{d,LS} + (1 - \rho\hat{\phi}_{LS})$  must also be identical. Thus, provided all three parameters,  $\beta_r$ ,  $\beta_d$ , and  $\phi$  are estimated (by OLS), there is no distinction between testing the joint null in equation (5) viz-á-viz the standard simple null of  $\beta_r = 0$ , and no power gains can therefore be achieved through such an approach.

### 2.3 Cochrane's simulation approach

In contrast, Cochrane (2008) proposes a simulation-based approach—from which finite-sample distributions of the estimators are obtained—and reports substantially stronger evidence of return predictability when considering a test based on  $\beta_d$  rather than  $\beta_r$ . The simulations are intended to replicate the estimated model as closely as possible, while imposing the null of no return predictability.

Specifically, the simulated model is specified as follows. The AR parameter  $\phi$  is set equal to the OLS estimate from the data being used (below, we also discuss alternatives to this parameterization). Given this value of  $\phi$ ,  $\beta_d$  is set such that  $\beta_r = \beta_d + (1 - \rho\phi) = 0$ , thus imposing the null of no return predictability. That is,  $\beta_d = -(1 - \rho\phi)$ . The parameter  $\rho$  is set to 0.9638 and the covariance matrix  $\Sigma$  is also set equal to the empirical estimate from the data.

Samples from this model are simulated, and  $\beta_r$  and  $\beta_d$  are estimated in each of these samples. In particular, equations (E2) and (E3) are simulated, and the values in the return equation (E1) are inferred from the identity (1). The present value identity thus holds exactly in the simulations, and by implication the restrictions (R1) and (R2) also hold exactly.

In each draw  $i$  of the simulation, coefficient estimates  $\hat{\beta}_{r,LS}^{i,sim}$ ,  $\hat{\beta}_{d,LS}^{i,sim}$ , and  $\hat{\phi}_{LS}^{i,sim}$  are obtained through OLS estimation. For each simulated sample, these are related as

$$\hat{\beta}_{r,LS}^{i,sim} = \hat{\beta}_{d,LS}^{i,sim} + \left(1 - \rho \hat{\phi}_{LS}^{i,sim}\right). \quad (7)$$

Let  $b_r^{Data} \equiv \hat{\beta}_{r,LS}^{Data}$  and  $b_d^{Data} \equiv \hat{\beta}_{d,LS}^{Data}$  denote, respectively, the OLS coefficient estimates of  $\beta_r$  and  $\beta_d$  in the actual data. The simulations are then used to evaluate how rarely the events  $\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data}$  and  $\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}$  occur. That is, treating  $b_r^{Data}$  and  $b_d^{Data}$  as given, the probabilities

$$\Pr\left(\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data}\right) \equiv p_r, \quad (8)$$

and

$$\Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \equiv p_d, \quad (9)$$

are determined based on the empirical distributions of  $\hat{\beta}_{r,LS}^{i,sim}$  and  $\hat{\beta}_{d,LS}^{i,sim}$ . These p-values capture the likelihood of observing the empirically estimated coefficients  $b_r^{Data}$  and  $b_d^{Data}$ , if the null of no return predictability was true.<sup>5</sup> Cochrane finds that  $p_d \approx 2\%$  whereas  $p_r \approx 22\%$ , which is interpreted as the dividend-growth based test providing much stronger evidence against the null of no return predictability than the predictability test for the actual returns. That is, the  $b_d^{Data}$  outcome is highly unlikely to have been generated under

---

<sup>5</sup>Cochrane also considers the empirical distribution of the  $t$ -ratios  $\hat{\beta}_{r,LS}^{i,sim}/(\text{Var}(\hat{\beta}_{r,LS}^{i,sim}))^{1/2}$  and  $\hat{\beta}_{d,LS}^{i,sim}/(\text{Var}(\hat{\beta}_{d,LS}^{i,sim}))^{1/2}$ . The subsequent literature seems to have primarily adopted the tests based directly on the coefficients (e.g., Golez and Koudijs, 2018), and we similarly focus on these in our analysis. In non-reported Monte Carlo simulations, we find size results for the dividend-growth  $t$ -ratio test that are similar to those we document for the coefficient test in Section 3.2 below.



the null model, whereas the  $b_r^{Data}$  outcome is not that unlikely. Based on these results, it is concluded that the test based on  $\beta_d$  is more powerful.<sup>6</sup>

Why are the two p-values,  $p_r$  and  $p_d$  different? According to the discussion in the previous sub-section, one would expect the tests of  $\beta_r$  and  $\beta_d$  to be identical. Consider first the case outlined above, where the value of  $\phi$  used to simulate the dividend-price ratio process is set to  $\phi^{Data} \equiv \hat{\phi}_{LS}^{Data}$ . This specification is the one that seems to have been adopted in subsequent empirical studies (Engsted and Pedersen, 2010; Golez and Koudijs, 2018), although Cochrane also considers other scenarios that we discuss further below. To see the implications of this formulation of the simulated model, write

$$b_d^{Data} = b_r^{Data} - (1 - \rho\phi^{Data}) = b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) + \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right), \quad (10)$$

where the first step uses the restriction on the coefficients from the data and the second step simply adds and subtracts  $\rho\hat{\phi}_{LS}^{i,sim}$ .<sup>7</sup> Using this expression together with the restriction on the estimated coefficients in the simulations (equation (7)), it follows that

$$\begin{aligned} & \Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \\ &= \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) + \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right)\right) \\ &= \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data}\right) \\ &\neq \Pr\left(\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data}\right) \end{aligned} \quad (11)$$

whenever  $\hat{\phi}_{LS}^{i,sim} \neq \phi^{Data}$ .

How does this simulation-based testing approach differ from the standard OLS inference discussed in the previous sub-section, where the  $\beta_r$ - and  $\beta_d$ -based tests of return

---

<sup>6</sup>The word ‘‘power’’ is used here in a somewhat imprecise sense. Formally, power is defined as the probability of rejecting the null under a given alternative. In Cochrane’s simulations, the rejection probabilities are all obtained under the null and are therefore not, in the true sense, a measure of power.

<sup>7</sup>The restriction on the coefficients (R1) does not hold exactly for the OLS estimates in the actual data, but the discrepancy is empirically very small.

predictability were shown to be identical? The key difference stems from the fact that if one wishes to use the estimate of  $\beta_d$  to test a null of  $\beta_r = 0$ , the relevant null hypothesis for  $\beta_d$  is in fact not fully known since  $\beta_r = 0 \Leftrightarrow \beta_d = -(1 - \rho\phi)$ .

In the simulation approach, the p-values for the  $\beta_d$ -based test reflect the sampling uncertainty in the estimates of  $\beta_d$ , but ignore the uncertainty coming from the fact that the value of the AR parameter in the original data is in fact unknown. In effect, the p-values correspond to a test of the null hypothesis  $\beta_d = -(1 - \rho\phi^{Data})$ , which postulates that the true AR parameter in the data is known and equal to  $\phi^{Data}$ .

Because of the downward bias in the OLS estimator of the AR coefficient  $\phi$ ,<sup>8</sup> it follows that on average,  $\phi^{Data} - \hat{\phi}_{LS}^{i,sim} > 0$ , and it is therefore reasonable to assume that

$$\Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) = \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data}\right) \leq \Pr\left(\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data}\right), \quad (12)$$

which goes some way towards explaining the results in Cochrane (2008). In fact, one can get a stronger result. The full information ML estimator of  $\beta_r$  with a known  $\phi$  (see Campbell and Yogo, 2006) is given by<sup>9</sup>

$$\hat{\beta}_{r,ML}(\phi) = \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}}\left(\phi - \hat{\phi}_{LS}\right). \quad (13)$$

In Cochrane's data,  $\sigma_{13}/\sigma_{33} \approx -0.9$  and  $\hat{\beta}_{r,ML} \approx \hat{\beta}_{r,LS} - 0.9\left(\phi - \hat{\phi}_{LS}\right)$ . With  $\rho \approx 0.96$ , it follows that

$$\hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right) \approx \hat{\beta}_{r,LS}^{i,sim} - 0.96\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right) \approx \hat{\beta}_{r,ML}^{i,sim}\left(\phi^{Data}\right), \quad (14)$$

---

<sup>8</sup>OLS estimation of  $AR(1)$  coefficients suffers from a well-known finite-sample bias (Kendall, 1954), which becomes more severe as the AR coefficient gets closer to unity.

<sup>9</sup>The full information ML estimator of  $\beta_r$  is derived from equations (E1) and (E3) under the assumption that the joint innovations  $\left(\epsilon_t^r, \epsilon_t^{dp}\right)'$  are *iid* normally distributed and treating the AR parameter  $\phi$  as a fixed and known constant.

where  $\hat{\beta}_{r,ML}^{i,sim}(\phi^{Data})$  is calculated using the true autoregressive parameter,  $\phi^{Data}$ , in the simulated model. Further, note that the ML estimator of  $\beta_r$ , with the least squares estimate of  $\phi$  treated as the known true value, trivially reduces to the OLS estimator of  $\beta_r$ . That is,  $b_r^{Data} = \hat{\beta}_{r,LS}^{Data} = \hat{\beta}_{r,ML}^{Data}(\phi^{Data})$ , where  $\hat{\beta}_{r,ML}^{Data}(\phi^{Data})$  is the ML estimator of  $\beta_r$  in the original data, treating  $\phi^{Data} = \hat{\phi}_{LS}^{Data}$  as the true value of  $\phi$ . Defining  $b_{r,ML}^{Data}(\phi^{Data}) \equiv \hat{\beta}_{r,ML}^{Data}(\phi^{Data}) = b_r^{Data}$ , equations (12) and (14) thus give

$$\Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \approx \Pr\left(\hat{\beta}_{r,ML}^{i,sim}(\phi^{Data}) \geq b_r^{Data}\right) = \Pr\left(\hat{\beta}_{r,ML}^{i,sim}(\phi^{Data}) \geq b_{r,ML}^{Data}(\phi^{Data})\right). \quad (15)$$

The test based on  $\hat{\beta}_{d,LS}$  in this setup of Cochrane's simulation-based approach therefore turns out to be very similar to a test using the ML estimator of  $\beta_r$ , when  $\phi$  is treated as fixed and known and equal to the OLS estimate in the data.

## 2.4 Altering the value of $\phi$ in the simulations

What happens if the value for  $\phi$  used in the simulations is changed from  $\phi^{Data} = \hat{\phi}_{LS}^{Data}$  to some other value? As shown in Appendix B, a similar result to that derived above holds, except the  $\beta_d$ -based test now corresponds to a test based on the ML estimator using this alternative value for the AR parameter as the true value of  $\phi$ . In particular, if one sets the AR parameter in the simulations equal to some maximum feasible value for  $\phi$ , say  $\phi^{Max}$ , one ends up with a test that is similar to Lewellen's (2004) test, interpreted by Campbell and Yogo (2006) as a sup-bound test. Provided the assumption  $\phi \leq \phi^{Max}$  indeed holds, the resulting test will generally be conservative, in the sense that if  $\phi < \phi^{Max}$ , the rejection rate under the null hypothesis will be smaller than the nominal significance level of the test (see Section 3.1 for further details). That is, analogous to the actual ML-based tests, setting the AR parameter large enough in Cochrane's simulation-based test is a way of

constructing tests that do not over-reject the null.<sup>10</sup>

## 2.5 Is the similarity with ML coincidental?

The near-equivalence of the dividend-growth based test and the ML-based test is seemingly somewhat accidental and depends on the specific values of  $\sigma_{13}$  and  $\sigma_{33}$ . Should one in general expect the two procedures to be close? If  $\sigma_{13}/\sigma_{33} = -\rho$ , the approximate equality in equation (15) is exact. By the restrictions in the covariance matrix  $\Sigma$  (equation (3)),  $\sigma_{13}/\sigma_{33} = -\rho$  is equivalent to  $\sigma_{23} = 0$ . Therefore, the near-equivalence of the tests is a result of the near-zero correlation between the dividend-growth and dividend-price ratio innovations observed in the data used by Cochrane (2008), which cover the U.S. market from 1927 to 2004.

However,  $\sigma_{23} \approx 0$  does not seem to be a universal fact. The empirical literature has documented a relatively strong correlation (ranging from 0.3 to 0.5) between the dividend-growth and dividend-price ratio innovations in both earlier U.S. data (Chen, 2009) as well as in international data (Engsted and Pedersen, 2010). In these contexts,  $\sigma_{23}$  is significantly different from zero, and the dividend-growth based test may differ substantially from the ML-based test. Since the latter makes (asymptotically) efficient use of the information on  $\phi$ , using the dividend-growth based test in these cases will result in a loss of efficiency compared to the ML-based test.<sup>11</sup>

---

<sup>10</sup>Cochrane (2008) is clearly aware that the choice of  $\phi$  is important in his simulation-based test and he considers several different alternative values. However, the study does not seem to recognize that using the OLS estimate of  $\phi$  will almost inevitably lead to a (severely) over-sized test, as discussed further in Section 3 below.

<sup>11</sup>In a recent working paper, McLoughlin (2019) shows that any approximation error in the present value relationship in equation (1) likely further decreases the power of the dividend-growth based test, giving the ML-based test additional relative advantage.

### 3 Size and power of the test

#### 3.1 Lessons from the ML estimator

The dividend-growth based test was shown to be similar to an ML-based test, where the value of the AR parameter  $\phi$  specified in the simulation design is treated as the “true” AR parameter in the ML estimator. Cochrane’s dividend-growth based test therefore (approximately) inherits the properties of the ML procedure, for a given specification of the AR parameter.

In case the AR parameter in the simulation design is set to the OLS estimate from the data, the result is a test that is severely over-sized. We illustrate this below by reporting actual rejection rates from a Monte Carlo simulation, under the null of no return predictability. However, the general idea can quite easily be understood by considering the properties of the ML estimator. As remarked above, the ML estimator using  $\hat{\phi}_{LS}$  instead of the true value  $\phi$ , reduces to the OLS estimator.

However, if one now proceeds *as if*  $\hat{\phi}_{LS}$  was indeed the true known value for  $\phi$ , the (asymptotic) variance of the estimator would erroneously be calculated as (see Appendix A)

$$Var\left(\hat{\beta}_{r,ML}\left(\hat{\phi}_{LS}\right)\right) = (1 - \delta_{13}^2) Var\left(\hat{\beta}_{r,LS}\right) \leq Var\left(\hat{\beta}_{r,LS}\right), \quad (16)$$

where  $\delta_{13} = \sigma_{13} / \sqrt{\sigma_{11}\sigma_{33}}$  is the correlation between the return and the dividend-price ratio innovations. Since  $\hat{\beta}_{r,ML}\left(\hat{\phi}_{LS}\right) = \hat{\beta}_{r,LS}$ , the variance of the ML estimator would therefore be severely under-estimated for large absolute values of  $\delta_{13}$  (in Cochrane’s data,  $\delta_{13} \approx -0.7$ ). Resulting test statistics based on this (erroneous) result would be over-sized;<sup>12</sup> for instance, a  $t$ -statistic would use too small a denominator, leading to a test that rejects the null hypothesis too often.

In general, one never has full knowledge of  $\phi$ , but one might be willing to impose some

---

<sup>12</sup>The OLS estimator is also biased, which further invalidates inference.

upper limit on the range of possible values for  $\phi$ . Most prominently, one might assume that the dividend-price ratio is a stationary process, such that  $\phi < 1$ . More generally, suppose one imposes the assumption that  $\phi \leq \phi^{Max}$ . In that case, provided  $\sigma_{13} < 0$ , one can form a “conservative” (downward biased) estimator of  $\beta_r$ ,

$$\hat{\beta}_{r,ML}(\phi^{Max}) = \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}} (\phi^{Max} - \hat{\phi}_{LS}). \quad (17)$$

This is essentially the approach taken by Lewellen (2004), although his motivation comes from the finite sample bias result for  $\hat{\beta}_{r,LS}$ , derived in Stambaugh (1999). Campbell and Yogo (2006) also discuss this type of conservative estimator and associated test statistics. Under the assumption that  $\phi \leq \phi^{Max}$ , and provided  $\sigma_{13} < 0$ , tests based on the resulting estimator are conservative (i.e., under-sized).<sup>13</sup>

### 3.2 Monte Carlo simulations

The above results are for the actual ML procedure. Given the close similarity between the ML-based test and Cochrane’s dividend-growth based test, similar results should also apply to the latter. To verify this, we perform a Monte Carlo simulation where the actual rejection rates under the null of no return predictability are obtained. To be clear, Cochrane’s testing procedure is in itself a simulation-based procedure, and here we evaluate the finite sample properties of that procedure in a controlled Monte Carlo experiment.

To generate the Monte Carlo samples, the dividend-growth and dividend-price ratio equations ((E2) and (E3)) are simulated, and the return equation (E1) is inferred from the present-value identity in equation (1). The null hypothesis  $\beta_r = 0$  is imposed and for

---

<sup>13</sup>A conservative implementation is not the only approach to making ML-based inference feasible (i.e., dealing with the fact that the true value of  $\phi$  is unknown). For instance, Campbell and Yogo (2006) use a Bonferroni method to obtain a feasible version of ML-based inference. Analogous implementations of the dividend-growth based test might also be considered, but they add little to the principal discussion regarding the relationship between the ML and dividend-growth based tests.

a given value for  $\phi$ , which varies between 0.9 and 0.995, the dividend-growth coefficient  $\beta_d$  is implicitly determined by the parameter restriction in equation (R1). The innovations,  $\epsilon_t^d$  and  $\epsilon_t^{dp}$ , are drawn from an *iid* normal distribution with variances and covariances set equal to the empirical estimates presented in Table 2 in Cochrane (2008). Intercepts are fitted in all regressions, although these are not commented on below. The sample size in each simulation draw is set to  $T = 78$ , the size of Cochrane’s original sample. As noted earlier, in the U.S. CRSP data set used by Cochrane, the parameter  $\sigma_{23}$  is close to zero and the approximation  $\rho \approx -\sigma_{13}/\sigma_{33}$  therefore holds well (specifically,  $\sigma_{13}/\sigma_{33} \approx -0.9$  and  $\rho \approx 0.96$ ). The simulation results are based on 10,000 repetitions.

Each of the Monte Carlo samples is used to represent a fictional “data set”, satisfying equations (E1)-(E3), and we apply Cochrane’s procedure to each one of these data sets. Thus, each one of the 10,000 Monte Carlo generated samples leads to a new set of simulations, according to Cochrane’s simulation-based procedure. In the case where the least squares estimate of  $\phi$  is used as the “true” AR parameter in Cochrane’s procedure, this second-layer simulation is parametrized with  $\beta_r = 0$  and  $\beta_d = -\left(1 - \rho\hat{\phi}_{LS}^{MC,data}\right)$ , where  $\hat{\phi}_{LS}^{MC,data}$  represents the OLS estimate of  $\phi$  in the given Monte Carlo sample. 50,000 sample paths are generated from this re-parameterized system, and the OLS estimates of  $\beta_d$  are collected from each of these samples. The p-value of Cochrane’s dividend-growth based test for the given Monte Carlo sample is determined by comparing the estimate of  $\beta_d$  from this sample ( $\hat{\beta}_{d,LS}^{MC,data}$ ) to the empirical distribution of the coefficient estimates of  $\beta_d$  from the 50,000 samples in the second simulation layer. Appendix C describes the full Monte Carlo simulation in greater and more formal detail.

The average rejection rate of the dividend-growth based test is calculated across the 10,000 Monte Carlo samples, classifying p-values less than 0.05 as rejections of the null hypothesis of no return predictability. That is, we fix the nominal size of Cochrane’s test procedure to five percent in a one-sided test against a positive alternative. As a compar-

ison to Cochrane’s test, we also calculate rejection rates for the ML-based test described in Campbell and Yogo (2006), which takes the shape of an adjusted  $t$ -test (referred to as the Q-test in Campbell and Yogo’s notation). For this test, the empirical rejection frequencies are calculated using standard critical values (i.e., reject for test statistics greater than 1.65).

The results from the Monte Carlo simulations are reported in Figure 1, with the size of the tests plotted as functions of the true parameter value  $\phi$  in the data generating process. The dividend-growth based test, using the least squares estimate of  $\phi$  as the “true” AR parameter, performs very poorly in terms of size. This is true for any underlying AR parameter  $\phi$ , with rejection rates always in excess of 20 percent for a nominal-sized 5 percent test. The ML-based test implemented in an analogous manner—using the least squares estimate of  $\phi$  as the “true” AR parameter—suffer from even larger size distortions than the dividend-growth based test, as also illustrated in Figure 1. This is to be expected, since the ML-based test makes even stronger use of the value for the AR parameter.

Figure 1 also presents results for the conservative ML-based test, using equation (17), and the dividend-growth based test performed in an analogous conservative way (see Appendix C), where the upper bound for the autoregressive parameter is set to  $\phi^{Max} = 0.995$ . In this conservative implementation, both tests exhibit rejection rates that are typically well below the nominal five percent significance level. As the true parameter  $\phi$  gets closer to the specified upper bound  $\phi^{Max}$ , the rejection rates approach five percent.

[Insert Figure 1 here]

### 3.3 Power

The large size distortions of the dividend-growth based test that uses the OLS estimate of  $\phi$  as the “true” AR parameter make this specific implementation of Cochrane’s test unsuitable for inference. The conservative test maintains size well. However, as argued



above, this test is similar to Lewellen’s (2004) test and Campbell and Yogo’s (2006) sup-bound test. Based on their studies, it is therefore clear that power gains can be achieved by using the dividend-growth equation, but only viz-á-viz standard OLS tests, not against ML-based tests that also use information on the AR parameter. The ML-based test is asymptotically the most powerful (see Campbell and Yogo, 2006), and the dividend-growth based test can therefore at best (asymptotically) achieve the power of the ML-based test.<sup>14</sup>

It is also useful to briefly revisit the arguments for how the ML-based test, and in consequence the dividend-growth based test, can achieve power gains. As pointed out in previous studies, such as Lewellen (2004) and Campbell and Yogo (2006), the key idea is that knowledge of the AR parameter  $\phi$  in the predictor variable can help improve inference. If the return and predictor innovations are correlated, the OLS estimator of the slope coefficient in the predictive return regression  $(\hat{\beta}_{r,LS})$  will be correlated with the OLS estimator of the AR parameter  $(\hat{\phi}_{LS})$ . Given the correlation between the OLS estimation errors for  $\beta_r$  and  $\phi$ , one can use the knowledge of the estimation error in the latter  $(\phi - \hat{\phi}_{LS})$  to partly correct the estimation error in the former (the estimation error in  $\hat{\phi}_{LS}$  is known if  $\phi$  is known). The ML estimator stated in equation (13) captures exactly this mechanism. The dividend-growth based test derives power from the same correlation, as discussed at length in Cochrane (2008), but in contrast to the ML-based approach the dividend-growth based test is not designed to make optimal use of this information (see Section 2.5 for additional discussion).

Finally, it is crucial to recognize that the power advantages for both the ML and dividend-growth based tests follow from some knowledge or restriction on  $\phi$  (e.g.,  $\phi < 1$ ).

---

<sup>14</sup>Strictly speaking, the optimality properties of the ML-based test refers to the case when the AR parameter is fully known. In a conservative formulation of the tests, power comparisons are difficult since, by definition, the tests are not correctly sized in general. In non-reported simulations, we show that under the parameter estimates in Cochrane’s study, the power of the ML and dividend-growth based tests are very close in finite samples, both in a conservative implementation and in an implementation where one makes use of the true value of the AR parameter (which is possible in a Monte Carlo exercise).

Without such knowledge, the correlation between  $\hat{\beta}_{r,LS}$  and  $\hat{\phi}_{LS}$  cannot be used in this way to improve upon the plain OLS procedure. In particular, one cannot use the OLS estimate  $\hat{\phi}_{LS}$  in lieu of the true (unknown) value of  $\phi$  to achieve the same goal: Such an approach leads to size distortions rather than power gains.<sup>15</sup>

## 4 Empirical results

We use the annual value-weighted CRSP returns, including and excluding dividends, to calculate the dividend-price ratio and the dividend growth. The sample period is 1927 to 2016, and we use real returns calculated with inflation defined as the monthly change in the Consumer Price Index.<sup>16</sup>

Our focus is on Cochrane’s dividend-growth based test, and we compare it to the ML-based test. Specifically, we estimate equations (E1)-(E3), using the full 1927 to 2016 sample, and calculate the p-values for each test over a range of different values postulated for  $\phi$ . That is, we do not use the OLS estimator of  $\phi$  in creating either of the test statistics, since this test suffers from large size distortions. Instead we calculate the range of p-values obtained for  $\phi \in [0.95, 0.995]$ .

Figure 2 illustrates the sensitivity of the test results with respect to the assumptions made on the AR coefficient  $\phi$ . In particular, the figure shows the p-values of the ML-based and the dividend-growth based tests as a function of the (maximum) value specified for  $\phi$ .<sup>17</sup> As is seen, if one is willing to assume that  $\phi \leq 0.97$ , one starts to find significant

---

<sup>15</sup>As pointed out in Footnote 13, alternative data-driven approaches (for dealing with an unknown  $\phi$ ) are possible. Importantly, however, such schemes cannot be based solely on the OLS estimator of  $\phi$ , as illustrated very clearly from the formula for the ML estimator, which reduces to the OLS estimator if  $\hat{\phi}_{LS}$  is used as the true value for the AR parameter (the discussion in Section 2.2 provides an analogue for the dividend-growth based approach). In fact, the Campbell and Yogo (2006) Bonferroni approach relies on an assumption that the AR parameter satisfies a so-called local-to-unity specification and if this assumption is not satisfied, the procedure is no longer applicable (Phillips, 2014).

<sup>16</sup>Using excess returns (over the 3-month Treasury Bill rate) or restricting the sample to end in 2004, as in the original study by Cochrane, leads to very similar results to those reported here.

<sup>17</sup>Table 5 in Cochrane (2008) presents calculations analogous to those shown in Figure 2 here, illustrating the effect of varying the value postulated for  $\phi$ .

results.

[Insert Figure 2 here]

Figure 2 also clearly illustrates how the empirical results can be viewed as conditional on one’s beliefs regarding the autocorrelation in the dividend-price ratio. If one is willing to make stronger assumptions than merely assuming that the dividend-price ratio is stationary, more significant results are obtained. It is worth pointing out that the OLS estimate of  $\phi$  is equal to 0.94, and using this value as the “true” AR value in the test procedures would thus have led to a strong rejection of the null hypothesis.

## 5 Conclusion

We provide a detailed analysis of the properties of the dividend-growth based test of return predictability in Cochrane (2008). We show that Cochrane’s test is similar to a full-information maximum likelihood test, using an explicit assumption on the degree of persistence in the predictor variable. Cochrane’s test can thus be viewed as an economically motivated proxy for the statistically motivated efficient ML method, and as such Cochrane’s test does not add power over and above the previously existing ML-based test. Importantly, we also show that unless one specifically imposes a conservative approach to the formulation of Cochrane’s procedure, the test will tend to over reject the null hypothesis of no return predictability.

Our article is followed by a rejoinder by John Cochrane and while we cannot respond in detail to his comments here, we would like to offer a few short observations. We never intended our study as a critique of Cochrane (2008), and the only real critique that we state is against the use of the dividend-growth based test implemented with the OLS estimator of the AR coefficient  $\phi$ . This practice is *not* a recommendation of Cochrane (2008), but rather the preferred implementation approach in subsequent empirical work, and we apologize if our write-up is not sufficiently clear on this point. The primary purpose of our

paper was always to provide an understanding of how, and to what extent, information in the dividend-growth equation—a fully redundant equation under the maintained model specification—could help improve upon return predictability inference. Our conclusion is by no means that one should go back to using plain OLS tests of return predictability: The dividend-growth equation can indeed help improve upon plain OLS inference, but not upon ML procedures that use the same information or assumptions on the AR parameter in the dividend-price ratio.

# Appendix

## A Properties of the OLS and ML estimators

For ease of notation, define  $x_t = d_t - p_t$ , and let  $X_{-1}$  denote the vector of stacked lagged observations for  $x_t$ . Similarly, denote  $r$  as the vector of observations on returns, and let  $\epsilon^r$ ,  $\epsilon^d$ , and  $\epsilon^{dp}$ , denote the stacked innovations. We treat the model without intercepts, but the results generalize immediately to regressions with fitted intercepts by replacing all variables by their demeaned versions.

The OLS estimator of  $\beta = (\beta_r, \beta_d, \phi)'$  is now equal to

$$\hat{\beta}_{LS} = \begin{bmatrix} \hat{\beta}_{r,LS} \\ \hat{\beta}_{d,LS} \\ \hat{\phi}_{LS} \end{bmatrix} = \beta + (X'_{-1}X_{-1})^{-1} \begin{bmatrix} X'_{-1}\epsilon^r \\ X'_{-1}\epsilon^d \\ X'_{-1}\epsilon^{dp} \end{bmatrix}. \quad (18)$$

Using  $\beta_r = \beta_d + (1 - \rho\phi)$  and  $\epsilon^r = \epsilon^d - \rho\epsilon^{dp}$ ,

$$\begin{aligned} \hat{\beta}_{r,LS} &= \beta_r + (X'_{-1}X_{-1})^{-1} X'_{-1}\epsilon^r \\ &= \left( \beta_d + (X'_{-1}X_{-1})^{-1} X'_{-1}\epsilon^d \right) + \left( 1 - \rho \left( \phi + (X'_{-1}X_{-1})^{-1} X'_{-1}\epsilon^{dp} \right) \right) \\ &= \hat{\beta}_{d,LS} + \left( 1 - \rho\hat{\phi}_{LS} \right). \end{aligned} \quad (19)$$

If  $\phi < 1$  and  $\epsilon_t$  is an *iid* or martingale difference sequence, it follows straightforwardly from classical asymptotic results that as the sample size  $T \rightarrow \infty$ ,

$$\sqrt{T} \left( \hat{\beta}_{LS} - \beta \right) \xrightarrow{d} N(0, (Var(x_{t-1}^2))^{-1} \Sigma), \quad (20)$$

where  $\Sigma$  is given in (3), and  $Var(x_{t-1}^2) = \frac{\sigma_{33}}{1-\phi^2}$ .<sup>18</sup> The asymptotic variance of the ML

---

<sup>18</sup>For  $\phi$  close to one, the asymptotic distribution result stated in (20) does not hold up well in finite

estimator of  $\beta_r$  can now be expressed as follows,

$$\begin{aligned}
\text{Var} \left( \hat{\beta}_{r,ML} \right) &= \text{Var} \left( \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}} \left( \phi - \hat{\phi}_{LS} \right) \right) \\
&= \text{Var} \left( \hat{\beta}_{r,LS} \right) + \frac{\sigma_{13}^2}{\sigma_{33}^2} \text{Var} \left( \hat{\phi}_{LS} \right) - 2 \frac{\sigma_{13}}{\sigma_{33}} \text{Cov} \left( \hat{\beta}_{r,LS}, \hat{\phi}_{LS} \right) \\
&= \text{Var} \left( \hat{\beta}_{r,LS} \right) + \frac{\sigma_{13}^2}{\sigma_{33}^2} \frac{\sigma_{33}}{\sigma_{11}} \text{Var} \left( \hat{\beta}_{r,LS} \right) - 2 \frac{\sigma_{13}}{\sigma_{33}} \frac{\sigma_{13}}{\sigma_{11}} \text{Var} \left( \hat{\beta}_{r,LS} \right) \\
&= \text{Var} \left( \hat{\beta}_{r,LS} \right) \left( 1 - \delta_{13}^2 \right), \tag{21}
\end{aligned}$$

where the third equality uses the shape of  $\Sigma$  given in (3).

## B The simulation-based test parametrized with $\phi^{Max}$

Before analyzing the simulation-based test, recall first the “conservative” ML estimator in equation (17),

$$\hat{\beta}_{r,ML} \left( \phi^{Max} \right) = \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}} \left( \phi^{Max} - \hat{\phi}_{LS} \right). \tag{22}$$

In a conservative test, the value of  $\hat{\beta}_{r,ML} \left( \phi^{Max} \right)$  is evaluated against the critical value that would apply if the AR parameter in the data was indeed equal to  $\phi^{Max}$ . Provided the true value of the AR parameter in the data is less than or equal to  $\phi^{Max}$ , and  $\sigma_{13} < 0$ , the resulting estimator will be downward biased, and tests based on this estimator will be conservative against a positive alternative.

Consider now the simulation-based test, where the simulated model is parametrized with an AR parameter  $\phi^{Max}$ . Let  $\hat{\beta}_{r,ML}^{i,sim} \left( \phi^{Max} \right)$  be the ML estimator of  $\beta_r$  in the simulations, and let  $b_{r,ML}^{Data} \left( \phi^{Max} \right) \equiv \hat{\beta}_{r,ML}^{Data} \left( \phi^{Max} \right)$  be the conservative ML estimate of  $\beta_r$  in the actual data. In the simulated data  $\hat{\beta}_{r,ML}^{i,sim} \left( \phi^{Max} \right)$  is the “correct” ML estimator, since the simulated model has an AR parameter equal to  $\phi^{Max}$ , whereas in the original data samples, and the stationarity condition ( $\phi < 1$ ) is by no means necessary for our main analysis. However, it enables a simple comparison of the (asymptotic) variances of the OLS and ML estimators.

the estimator is conservative (provided the true AR parameter is less than or equal to  $\phi^{Max}$ ). In order to show that the simulation-based dividend-growth test is similar to the conservative ML test in this case, we need to show that

$$\Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \approx \Pr\left(\hat{\beta}_{r,ML}^{i,sim}(\phi^{Max}) \geq b_{r,ML}^{Data}(\phi^{Max})\right). \quad (23)$$

That is,  $\Pr\left(\hat{\beta}_{r,ML}^{i,sim}(\phi^{Max}) \geq b_{r,ML}^{Data}(\phi^{Max})\right)$  compares the conservative ML estimate from the data,  $b_{r,ML}^{Data}(\phi^{Max})$ , to the distribution of the ML estimator in a setting where  $\phi^{Max}$  is indeed the true value for the AR parameter.

For  $\rho \approx -\sigma_{13}/\sigma_{33}$ , the ML estimator of  $\beta_r$  in the simulations can be written as

$$\hat{\beta}_{r,ML}^{i,sim}(\phi^{Max}) = \hat{\beta}_{r,LS}^{i,sim} + \frac{\sigma_{13}}{\sigma_{33}}\left(\phi^{Max} - \hat{\phi}_{LS}^{i,sim}\right) \approx \hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Max} - \hat{\phi}_{LS}^{i,sim}\right). \quad (24)$$

As in Section 2.3, let  $b_r^{Data} \equiv \hat{\beta}_{r,LS}^{Data}$  and  $b_d^{Data} \equiv \hat{\beta}_{d,LS}^{Data}$ . By restriction (R1),

$$b_d^{Data} = b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{Data}\right) = b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) + \rho\left(\hat{\phi}_{LS}^{Data} - \hat{\phi}_{LS}^{i,sim}\right), \quad (25)$$

and  $\hat{\beta}_{d,LS}^{i,sim} = \hat{\beta}_{r,LS}^{i,sim} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right)$ . The conservative ML estimator in the actual data, using  $\rho \approx -\sigma_{13}/\sigma_{33}$ , can be written as

$$b_{r,ML}^{Data}(\phi^{Max}) \approx b_r^{Data} - \rho\left(\phi^{Max} - \hat{\phi}_{LS}^{Data}\right). \quad (26)$$

It follows that

$$\begin{aligned}
& \Pr \left( \hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data} \right) \\
&= \Pr \left( \hat{\beta}_{r,LS}^{i,sim} - \left( 1 - \rho \hat{\phi}_{LS}^{i,sim} \right) \geq b_r^{Data} - \left( 1 - \rho \hat{\phi}_{LS}^{i,sim} \right) + \rho \left( \hat{\phi}_{LS}^{Data} - \hat{\phi}_{LS}^{i,sim} \right) \right) \\
&= \Pr \left( \hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data} + \rho \left( \hat{\phi}_{LS}^{Data} - \hat{\phi}_{LS}^{i,sim} \right) \right) \\
&= \Pr \left( \hat{\beta}_{r,LS}^{i,sim} - \rho \left( \phi^{Max} - \hat{\phi}_{LS}^{i,sim} \right) \geq b_r^{Data} - \rho \left( \phi^{Max} - \hat{\phi}_{LS}^{Data} \right) \right) \\
&\approx \Pr \left( \hat{\beta}_{r,ML}^{i,sim} \left( \phi^{Max} \right) \geq b_{r,ML}^{Data} \left( \phi^{Max} \right) \right). \tag{27}
\end{aligned}$$

## C Implementation of Monte Carlo simulations

For a given set of parameter values, the procedure is summarized by the below steps. For a simulation with 10,000 repetitions, steps 1-4 are repeated 10,000 times.

1. Simulate a single sample of size  $T = 78$  of the system in equations (E1)-(E3), under restrictions (R1) and (R2).
2. From the simulated sample, obtain OLS estimates of all regression coefficients, as well as estimates of all parameters in the covariance matrix  $\Sigma$ . Denote the OLS estimates of  $\beta_r$ ,  $\beta_d$ , and  $\phi$  as  $\hat{\beta}_{r,LS}^{MC,data}$ ,  $\hat{\beta}_{d,LS}^{MC,data}$ , and  $\hat{\phi}_{LS}^{MC,data}$ , respectively.
3. Parametrize the same system as in step (1), imposing the null of no return predictability and replacing all other parameters of the model with the *empirical estimates* from step (2). That is, in this parametrization,  $\beta_r = 0$ ,  $\beta_d = - \left( 1 - \rho \hat{\phi}_{LS}^{MC,data} \right)$ , and all other parameters are set equal to the empirical estimates obtained in step (2). Simulate 50,000 sample paths of size  $T = 78$  from this empirically parametrized system that imposes the null of no return predictability. For each of the simulated samples, obtain OLS estimates of  $\beta_r$  and  $\beta_d$ .
4. Based on the resulting empirical distributions of the OLS estimators in step (3),



calculate the p-values for the coefficient estimates  $\hat{\beta}_{r,LS}^{MC,data}$  and  $\hat{\beta}_{d,LS}^{MC,data}$ , obtained from the initial simulated sample in steps (1) and (2). Classify as rejection/non-rejection depending on whether the respective p-value is less or greater than 0.05.

The above steps refer to the implementation of Cochrane's procedure where the parametrization in step (3) uses the OLS estimate of  $\phi$  obtained in step (2). We also consider a conservative test where the value of  $\phi$  used to parametrize the simulated system in step (3) is set to  $\phi^{Max} = 0.995$  and the value of  $\beta_d$  is adjusted accordingly to ensure that the null of no return predictability is imposed (i.e.,  $\beta_d = -(1 - \rho\phi^{Max})$ ).

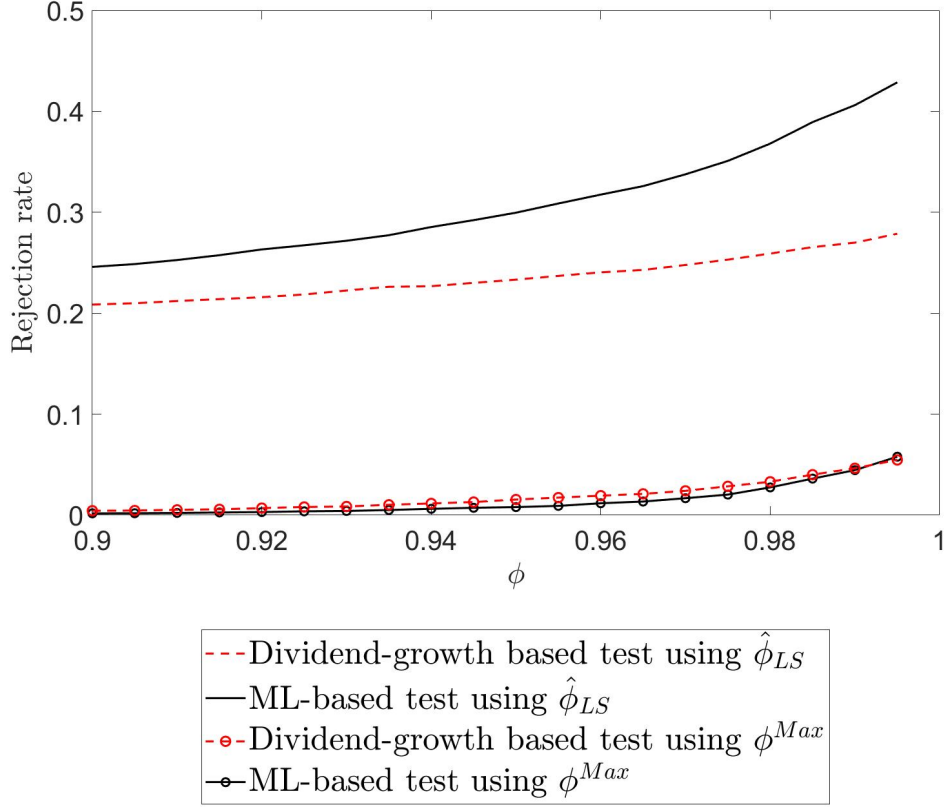
## References

- Binsbergen, V., Jules, H., and Koijen, R. S. (2010). Predictive Regressions: A Present-Value Approach. *The Journal of Finance*, 65(4):1439–1471.
- Bollerslev, T., Xu, L., and Zhou, H. (2015). Stock Return and Cash Flow Predictability: The Role of Volatility Risk. *Journal of Econometrics*, 187(2):458–471.
- Campbell, J. Y. and Shiller, R. J. (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 Yogo, M. (2006). Efficient Tests of Stock Return Predictability. *Journal of Financial Economics*, 81(1):27–60.
- 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.
- Chen, L., Da, Z., and Priestley, R. (2012). Dividend Smoothing and Predictability. *Management Science*, 58(10):1834–1853.
- Chen, W. W. and Deo, R. S. (2009). Bias Reduction and Likelihood-Based Almost Exactly Sized Hypothesis Testing in Predictive Regressions Using the Restricted Likelihood. *Econometric Theory*, 25(5):1143–1179.
- Cochrane, J. H. (2008). The Dog that did not Bark: A Defense of Return Predictability. *Review of Financial Studies*, 21(4):1533–1575.
- Engsted, T. and Pedersen, T. Q. (2010). The Dividend-Price Ratio Does Predict Dividend Growth: International Evidence. *Journal of Empirical Finance*, 17(4):585–605.
- Engsted, T., Pedersen, T. Q., and Tanggaard, C. (2012). The Log-Linear Return Approximation, Bubbles, and Predictability. *Journal of Financial and Quantitative Analysis*, 47(03):643–665.

- Golez, B. (2014). Expected Returns and Dividend Growth Rates Implied by Derivative Markets. *Review of Financial Studies*, 27(3):790–822.
- Golez, B. and Koudijs, P. (2018). Four Centuries of Return Predictability. *Journal of Financial Economics*, 127(2), 248-263.
- Kelly, B. and Pruitt, S. (2013). Market Expectations in the Cross-Section of Present Values. *Journal of Finance*, 68(5):1721–1756.
- Kendall, M. G. (1954). Note on Bias in the Estimation of Autocorrelation. *Biometrika*, 41(3-4):403–404.
- Koijen, R. S. and Van Nieuwerburgh, S. (2011). Predictability of Returns and Cash Flows. *Annual Review of Financial Economics*, 3:467–491.
- Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust Econometric Inference for Stock Return Predictability. *Review of Financial Studies*, 28(5):1506–1553.
- Lettau, M. and Ludvigson, S.C. (2005). Expected Returns and Expected Dividend Growth. *Journal of Financial Economics*, 76(3):583–626.
- Lettau, M. and Ludvigson, S.C. (2010). Measuring and Modelling Variation in the Risk-Return Trade-off. *Handbook of Financial Econometrics Volume 1*, 617-690, ed. by Y. Ait-Sahalia and L.P. Hansen. Elsevier Science B.V., North Holland, Amsterdam.
- Lewellen, J. (2004). Predicting Returns with Financial Ratios. *Journal of Financial Economics*, 74(2):209–235.
- Maio, P. and Santa-Clara, P. (2015). Dividend Yields, Dividend Growth, and Return Predictability in the Cross Section of Stocks. *Journal of Financial and Quantitative Analysis*, 50(1-2):33–60.

- McLoughlin, J. (2019). Why Neither Dividend-Growth nor Long-Horizon Regressions Provide Stronger Evidence of Return Predictability. DERA Working Paper, U.S. Securities and Exchange Commission.
- Phillips, P. C. (2014). On Confidence Intervals for Autoregressive Roots and Predictive Regression. *Econometrica*, 82(3):1177–1195.
- Phillips, P. C. and Chen, Y. (2014). Restricted Likelihood Ratio Tests in Predictive Regression. *Cowles Foundation Discussion Paper*, No. 1968.
- Stambaugh, R. F. (1999). Predictive Regressions. *Journal of Financial Economics*, 54(3):375–421.

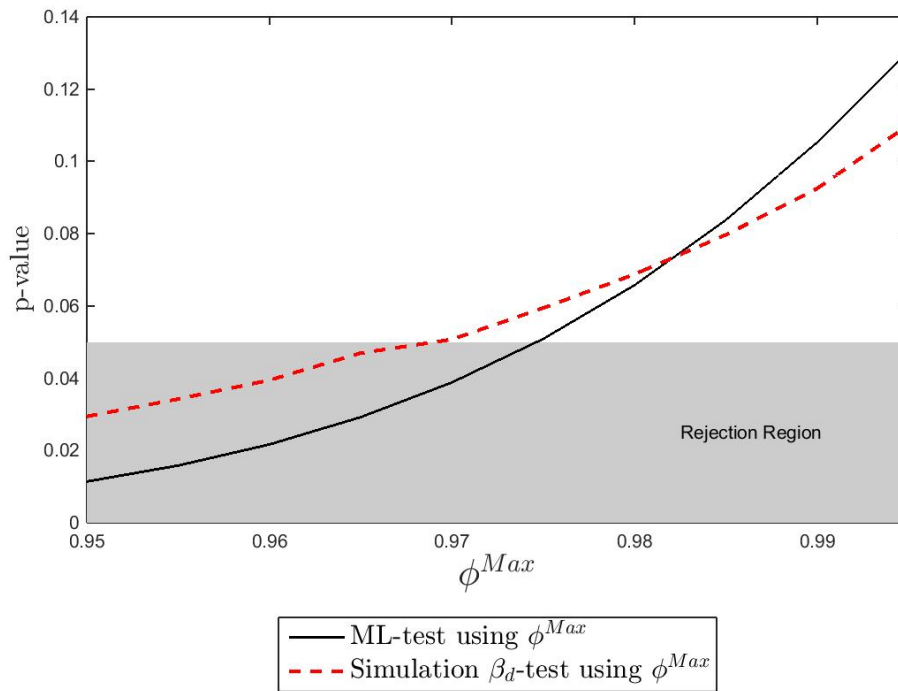
Figure 1: **Actual size of the tests**



**Description:** The figure presents the actual size of one-sided nominal-sized five percent tests of the null hypothesis of  $\beta_r = 0$  against a positive alternative. That is, the graphs indicate the average rejection rates, for the corresponding tests, under the null hypothesis of no return predictability. The dashed (solid) line is the dividend-growth (ML-) based test using the OLS estimate  $\hat{\phi}_{LS}$  as the value for the AR parameter. The dashed (solid) line with circles represents the rejection rates for the conservative dividend-growth (ML-) based test using  $\phi^{Max} = 0.995$  as the value for the AR parameter. The ML-based tests use standard normal critical values (i.e., reject for test statistics greater than 1.65). The results are based on the Monte Carlo simulation described in the main text with 10,000 repetitions.

**Interpretation:** Both the dividend-growth based test and the ML-based test suffer from size distortions when using the OLS estimate of the autoregressive parameter of the dividend-price ratio,  $\hat{\phi}_{LS}$ . The tests control size well if they are implemented in a conservative way, using an upper bound on the autoregressive parameter,  $\phi^{Max}$ .

Figure 2: Empirical p-values as a function of  $\phi^{Max}$



**Description:** The graph shows the p-values of tests of return predictability. The p-values are plotted as functions of the value specified for the upper bound,  $\phi^{Max}$ , on the AR parameter. The dashed line represents the dividend-growth based test, and the solid line represents the ML-based test (using critical values from the standard normal distribution). Both tests specify that the value for the autoregressive parameter is set to the corresponding value for  $\phi^{Max}$ .

**Interpretation:** The dividend-growth based test and the ML-based test are both sensitive to the assumptions made on (the upper bound of) the autoregressive parameter of the dividend-price ratio.