Testing Return Predictability with the Dividend-Growth Equation: An Anatomy of the Dog

Erik Hjalmarsson Tamás Kiss*

October 23, 2018

Abstract

The power gains of the dividend-growth based test of return predictability, proposed by Cochrane [2008, Review of Financial Studies 21, 1533-1575], are due to a strong use of the exact value postulated for the autoregressive parameter of the dividend-price ratio. In particular, we show that Cochrane's test is similar to a likelihood-based test of the standard return-predictability model, treating the autoregressive parameter as known. When compared to the likelihood-based test, there are no power gains. In its most common implementation, with the autoregressive parameter set equal to the corresponding OLS estimate, Cochrane's test also suffers from severe size distortions.

JEL classification: C22, G1.

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

*Both authors are with the 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 and tamas.kiss@cff.gu.se. We have benefitted from comments by Daniel Buncic, Paolo Giordani, Randi Hjalmarsson, Ron Kaniel, Jesper Lindé, Anders Wilhelmsson, Xin Zhang, and Pär Österholm, as well as by seminar participants at Sveriges Riksbank, Örebro University, and the European Summer Meeting of the Econometric Society, Cologne 2018. Hjalmarsson gratefully acknowledges financial support from the Nasdaq Nordic Foundation.

1 Introduction

In an influential study, Cochrane (2008) argues that rather than directly testing return predictability, one should instead consider dividend-growth predictability. 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. Most importantly, Cochrane finds that such an indirect test, based on predictability in dividend growth, is considerably more powerful at detecting predictability *in returns*, than a test based directly on returns. 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.[1]

In this paper, we provide a formal analysis of the properties of Cochrane's (2008) procedure, with the aim of forming a better understanding of the power gains in the dividend-growth based test of return predictability.[2] 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. Cochrane 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. Our main finding is that

[1]Papers explicitly using Cochrane's (2008) approach to test for return predictability include Chen (2009); Engsted and Pedersen (2010); Golez and Koudijs (2017). Other related works that consider both return and dividend-growth predictability include Lettau and Ludvigson (2005); Boudoukh et al. (2007); Lettau and Van Nieuwerburgh (2008); Binsbergen and Koijen (2010); Lacerda and Santa-Clara (2010); Lettau and Ludvigson (2010); Chen et al. (2012); Kelly and Pruitt (2013); Golez (2014); Bollerslev et al. (2015); Maio and Santa-Clara (2015); Detzel and Strauss (2016). Koijen and Van Nieuwerburgh (2011) provides a review on return and dividend-growth predictability.

[2]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.

Cochrane's dividend-growth based test is very similar to a test based on the full information maximum likelihood (ML) estimator for the standard bi-variate system, where the AR parameter of the dividend-price ratio is treated as known.[3] This finding explains why the dividend-growth based test, as implicitly formulated by Cochrane, is more powerful than the return-based test using the simple OLS estimator. However, this power gain is achieved because the dividend-growth based test makes strong use of the postulated value of the AR coefficient. We show that if one were to use the same "knowledge" in the return-based test, one could use an ML procedure that 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 (i.e., treating the AR parameter as known and equal to some given value), there are no power gains from using the dividend-growth regression in testing for return predictability.

In fact, 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, Cochrane's choice to treat the OLS estimate of the AR parameter as the "true" value, which also appears to have been adopted in subsequent empirical studies (e.g., Engsted and Pedersen, 2010; Golez and Koudijs, 2017), 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.

Cochrane also considers alternative specifications, where the AR parameter takes on different values than the OLS estimate. 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

[3]The very close similarity with the ML approach is somewhat coincidental, and depends on the specific parameter values in a given data set. However, as seen in the formal analysis, the test implied by Cochrane's method and the ML-based test both use an adjusted form of the OLS estimator as their main input, and the only difference is the weight put on this adjustment factor.

that it is greater than the true parameter value in the data. In this case, we show that Cochrane's test becomes similar to the conservative sup-bound test developed in Lewellen (2004) and analyzed further in Campbell and Yogo (2006).[4]

To form some intuition for our results, note that Cochrane's predictive model is made up of three regressions, where the dependent variables are returns, dividend growth, and the dividend-price ratio. In all three equations, the right-hand side variable is given by the lagged dividend-price ratio. These equations are linked together by the Campbell and Shiller (1988) present value identity, which implies an exact relationship between the slope coefficients as well as the error terms in the three regressions, such that any one of the three equations is redundant in the model formulation. The tri-variate system can thus be reduced to a bi-variate system, using any two of the three equations. Therefore, the standard 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 empirical research, our findings have important implications. 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. For a given assumption regarding the AR parameter of the dividend-price ratio, the dividend-growth based test is similar to an ML procedure. Since the ML procedure is much easier to implement—it relies on standard critical values rather than a simulated distribution as in Cochrane's test—and ML is (asymptotically) more efficient, there are no apparent gains from using the dividend-growth based approach.

[4]In this implementation, Cochrane's test is sometimes slightly more powerful than the ML-based approach of Lewellen (2004) and Campbell and Yogo (2006), although the observed differences are very small. The conservative nature of the tests also makes power comparisons difficult (see Footnote 20).

If one has strong prior views on the (maximum) value of the AR parameter, the ML procedure therefore seems to be the preferred choice. However, some caution is required: Unless the postulated value for the AR parameter is in fact greater than or equal to the true (unknown) AR parameter in the dividend-price ratio process, the resulting test will be over-sized (e.g., Lewellen, 2004, and Campbell and Yogo, 2006). Importantly, the OLS estimator of the AR parameter is not precise enough to be of much use in informing about the correct value of the AR parameter. Treating the OLS estimate of the AR coefficient as the "true" and known parameter value in the likelihood function results in severely over-sized tests (see Section 2.2, as well as Amihud and Hurvich, 2004, for a related discussion). Similarly, relying upon the OLS-estimated value of the AR parameter in Cochrane's dividend-growth based test of return predictability results in large size distortions.[5]

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. Having said that, our study

[5] We also illustrate that the test of no return predictability formulated in Binsbergen and Koijen (2010) is severely size distorted. Like Cochrane, Binsbergen and Koijen work with a predictive model, imposing a present value relationship. However, they perform a full-system maximum likelihood estimation of their state space system and impose no prior value on the AR parameter of the dividend-price ratio. The over-rejection in the Binsbergen and Koijen test is not particularly surprising: Standard tests in predictive models tend not to be correctly sized—essentially as a result of the so-called Stambaugh (1999) bias—and there is no attempt at correcting for this bias in the Binsbergen and Koijen methodology.

makes no claim to provide an exhaustive empirical analysis of return predictability, or 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). Instead, we focus solely on the properties of the dividend-growth based procedure proposed by Cochrane (2008), and how it relates to likelihood-based inference.

The rest of the paper is organized as follows. Section 2 describes the predictive model, with the main theoretical findings presented in Section 3. Section 4 provides further results from Monte Carlo simulations and the empirical analysis is found in Section 5. Section 6 concludes and the Appendix contains some supplementary material as well as some technical derivations.

2 The return and dividend-growth system

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

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

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

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

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

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

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

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

where $\overline{d-p}$ is the average dividend-price ratio. The parameter ρ 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),

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

and

$$\epsilon_t^r = \epsilon_t^d - \rho\epsilon_t^{dp}. \tag{R2}$$

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. 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).

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

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

It is easily seen that Σ is not full rank, reflecting the redundancy in the equation system (E1)-(E3). Define also $\delta_{13} = \sigma_{13} \big/ \sqrt{\sigma_{11}\sigma_{33}}$ as the correlation between the innovations of the dividend-price ratio and the return equation.

The above model formulation also implies that the coefficient restriction (R1) will apply to the OLS estimates as well (see Appendix A), such that

$$\hat{\beta}_{r,LS} = \hat{\beta}_{d,LS} + \left(1 - \rho\hat{\phi}_{LS}\right). \tag{4}$$

Thus, if one estimates β_r directly from the return regression, or indirectly from the dividend-growth and dividend-price ratio regressions, the OLS estimates are numerically identical.

2.2 Estimation when ϕ is known or restricted

The main parameter of interest is the slope coefficient in the return regression, β_r, where the null hypothesis of no return predictability clearly corresponds to $\beta_r = 0$.[6] As seen from the restriction in (R1), this null can equivalently be formulated as $\beta_d = -(1 - \rho\phi)$.

The AR parameter ϕ is generally not of direct interest as such, but plays a major role in the properties of the estimators and test statistics for the predictive model (e.g., Cavanagh et al., 1995; Stambaugh, 1999; Lewellen, 2004; Campbell and Yogo, 2006; Kostakis et al., 2015). If ϕ is close to one, and the innovations to returns (ϵ_t^r) and the dividend-price

[6]There is now a large literature separately testing both return and dividend growth predictability (see references in Footnote 1). Here we follow Cochrane (2008) and focus on tests of return predictability.

ratio $\left(\epsilon_t^{dp}\right)$ are (strongly) correlated, the OLS estimator of β_r suffers from a (substantial) finite-sample bias, and its finite sample distribution will be far from normal. Furthermore, ϕ plays a direct role in the formulation of the null hypothesis of no return predictability, when this null is formulated in terms of the dividend-growth coefficient, β_d. Similarly to previous studies, the treatment of ϕ therefore plays a determining role in Cochrane's analysis, as discussed in detail in the next section. For that reason, we briefly review some basic but relevant results on the role of ϕ in the estimation of the return predictability coefficient, β_r.

If one has some prior knowledge of ϕ, one can use this information to improve upon plain OLS inference, as discussed at length in Campbell and Yogo (2006). Suppose first that ϕ is completely known. Following the derivations in Campbell and Yogo (2006), one can easily show that the full information ML estimator of β_r, treating ϕ as known,[7] is given by

$$\hat{\beta}_{r,ML} = \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}}\left(\phi - \hat{\phi}_{LS}\right). \tag{5}$$

The OLS estimator, $\hat{\beta}_{r,LS}$, is in general biased, and this bias is a function of the persistence in the regressor, ϕ. Using knowledge of ϕ, the ML estimator controls for this bias by adjusting the OLS estimator.[8] By standard arguments, the ML estimator is (asymptotically) efficient, and the corresponding tests are the most powerful (Campbell and Yogo, 2006).

Suppose now that one does not have full knowledge of ϕ, but one is willing to impose some upper limit on the range of possible values for ϕ. Most prominently, one might be willing to 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,

[7]For simplicity, Σ is also treated as known here. This does not affect the derivations of the ML estimator, and the only change would be the replacement of the true quantities σ_{13} and σ_{33} with estimates of these.

[8]In general, the distribution of the OLS estimator is not only biased, but also non-normal for ϕ close to unity. The ML estimator not only controls for the bias, but also makes the distribution "more" normal.

provided $\sigma_{13} < 0$, one can form a "conservative" (downward biased) estimator of β_r,

$$\hat{\beta}_{r,ML}\left(\phi^{Max}\right) = \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}}\left(\phi^{Max} - \hat{\phi}_{LS}\right). \quad (6)$$

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.

The ML estimator relies on either full prior knowledge of ϕ, as in (5), or some upper bound on ϕ, as in (6). Importantly, in case ϕ is unknown, one cannot simply replace the true value of ϕ with the least squares estimate in the ML formula, and proceed as if $\hat{\phi}_{LS}$ was indeed the true known value. In that case, one would end up with,

$$\hat{\beta}_{r,ML}\left(\hat{\phi}_{LS}\right) = \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}}\left(\hat{\phi}_{LS} - \hat{\phi}_{LS}\right) = \hat{\beta}_{r,LS}. \quad (7)$$

That is, the ML estimator, using $\hat{\phi}_{LS}$ instead of the true value ϕ, reduces to the OLS estimator. However, if one now proceeds *as if* $\hat{\phi}_{LS}$ was indeed the true known value for ϕ, the perceived (asymptotic) variance of the estimator would erroneously be calculated as (see Appendix C)

$$Var\left(\hat{\beta}_{r,ML}\left(\hat{\phi}_{LS}\right)\right) = \left(1 - \delta_{13}^2\right) Var\left(\hat{\beta}_{r,LS}\right) \leq Var\left(\hat{\beta}_{r,LS}\right). \quad (8)$$

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 δ_{13} close to unity, and resulting test statistics based on this (erroneous) result would be over-sized.[9,10]

3 Testing return predictability

3.1 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,

$$\beta_r = \beta_d + (1 - \rho\phi)\,, \tag{9}$$

implies that if $\beta_r = 0$,

$$\beta_d = -(1 - \rho\phi) < 0, \tag{10}$$

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 as

$$H_0 : \beta_r = 0 \quad \text{and} \quad \beta_d + (1 - \rho\phi) = 0. \tag{11}$$

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, such that $\hat{\beta}_{r,LS} \equiv \hat{\beta}_{d,LS} + \left(1 - \rho\hat{\phi}_{LS}\right)$. Since this equality holds numerically for the OLS estimates in any sample, the distribution of $\hat{\beta}_{r,LS}$ and $\hat{\beta}_{d,LS} + \left(1 - \rho\hat{\phi}_{LS}\right)$ must also be identical. Therefore, if one uses the OLS estimates of β_r, β_d, and ϕ to test either of the simple nulls, $\beta_r = 0$

[9] The OLS estimator is also biased, which further invalidates inference.

[10] Amihud and Hurvich (2004) and Amihud et al. (2008, 2010) analyze how bias-corrected versions of the OLS estimator of ϕ can be used to achieve bias-corrections in the estimator of β_r. Campbell and Yogo (2006) use a bonferroni approach to obtain a feasible version of ML-based inference, although Phillips (2014) has subsequently leveled a critique against the reliability of this type of procedure.

or $\beta_d + (1 - \rho\phi) = 0$, the two tests using $\hat{\beta}_{r,LS}$ or $\hat{\beta}_{d,LS} + \left(1 - \rho\hat{\phi}_{LS}\right)$ must have the same rejection regions for a given significance level. Appendix A illustrates this point, by showing that the individual standard t-statistics for testing $\beta_r = 0$ or $\beta_d + (1 - \rho\phi) = 0$ are numerically identical.

Thus, provided all three parameters, β_r, β_d, and ϕ are estimated (with OLS), there is no distinction between testing the joint null vis a vis the standard simple null of $\beta_r = 0$, and no power gains can therefore be achieved through such an approach.

3.2 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 what one might interpret as considerable power gains when considering a test based on β_d rather than β_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 ϕ is set equal to the OLS estimate from the data being used. Given this value of ϕ, β_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 ρ is set to 0.9638 and the covariance matrix Σ is also set equal to the empirical estimate from the data.

Samples from this model are simulated, and β_r and β_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 $\hat{\beta}^{i,sim}_{r,LS}$, $\hat{\beta}^{i,sim}_{d,LS}$, and $\hat{\phi}^{i,sim}_{LS}$ are obtained through OLS estimation. For each simulated sample, these are related as

$$\hat{\beta}_{r,LS}^{i,sim} = \hat{\beta}_{d,LS}^{i,sim} + \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right). \tag{12}$$

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 β_r and β_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

$$\Pr\left(\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data}\right) \equiv p_r, \tag{13}$$

and

$$\Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \equiv p_d, \tag{14}$$

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.[11] Cochrane finds that $p_d \approx 2\%$ whereas $p_r \approx 22\%$, which is interpreted as the dividend-growth 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 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 β_d is more powerful.[12]

Where does the additional power of the β_d-test come from? According to the discussion

[11]Cochrane also considers the empirical distribution of the t-ratios $\hat{\beta}_{r,LS}^{i,sim} \Big/ \sqrt{Var\left(\hat{\beta}_{r,LS}^{i,sim}\right)}$ and $\hat{\beta}_{d}^{i,sim} \Big/ \sqrt{Var\left(\hat{\beta}_{d,LS}^{i,sim}\right)}$. The subsequent literature seems to have primarily adopted the tests based directly on the coefficients (e.g., Golez and Koudijs, 2017), and we similarly focus on these in our analysis. In non-reported Monte Carlo simulations, we find results for the t-ratio tests that are very similar to those we document for the coefficient tests in Section 4 below.

[12]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.

in the previous sub-section, one would expect the tests of β_r and β_d to be identical. In Cochrane's main simulation setup, the value of ϕ used to simulate the dividend-price ratio process is set to $\phi^{Data} \equiv \hat{\phi}_{LS}^{Data}$ (other scenarios are also considered, and we discuss those further below). To see the implications of this formulation of the simulated model, write

$$b_d^{Data} = b_r^{Data} - \left(1 - \rho\phi^{Data}\right) = b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) + \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right), \tag{15}$$

where the first step uses the restriction on the coefficients from the data and the second step simply adds and subtracts $\rho\hat{\phi}_{LS}^{i,sim}$.[13] Using the restriction on the estimated coefficients in the simulations stated in equation (12), it follows that

$$\begin{aligned}
& \Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \\
= \; & \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) + \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right)\right) \\
= \; & \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data}\right) \\
\neq \; & \Pr\left(\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data}\right)
\end{aligned} \tag{16}$$

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 β_r- and β_d-based tests of return predictability were shown to be identical? The key difference stems from the fact that if one wishes to use the estimate of β_d to test a null of $\beta_r = 0$, the relevant null hypothesis for β_d is in fact not fully known since $\beta_r = 0 \Leftrightarrow \beta_d = -(1 - \rho\phi)$. Without any further information on ϕ, one would therefore test the null $\beta_r = 0$ by evaluating whether $\hat{\beta}_{d,LS}^{Data} + \left(1 - \rho\hat{\phi}_{LS}^{Data}\right)$ is "close to" zero, which is then identical to simply considering $\hat{\beta}_{r,LS}^{Data}$, since $\hat{\beta}_{r,LS}^{Data} = \hat{\beta}_{d,LS}^{Data} + \left(1 - \rho\hat{\phi}_{LS}^{Data}\right)$, as discussed previously. In the simulation approach,

[13] The restriction on the coefficients (equation (4)) does not hold exactly for the OLS estimates in the actual data, but the discrepancy is very small as seen in Table 2 in the empirical section.

the p-values for the β_d-based test reflect the sampling uncertainty in the estimates of β_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 the test of a known hypothesis, $\beta_d = -\left(1 - \rho\phi^{Data}\right)$, which postulates that the AR parameter is known and equal to ϕ^{Data}. That is, the β_d-based test makes implicit use of the true value of the AR parameter in the simulated data (set to $\phi^{Data} = \hat{\phi}_{LS}^{Data}$), and effectively treats this value as if it was also the true and known parameter value in the original data.

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

$$\Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) = \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data}\right) \leq \Pr\left(\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data}\right), \tag{17}$$

which goes some way towards explaining the results in Cochrane (2008). In fact, one can get a stronger result. Recall the ML estimator of β_r with a known ϕ, given in equation (5) and re-stated here,

$$\hat{\beta}_{r,ML} = \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}}\left(\phi - \hat{\phi}_{LS}\right). \tag{18}$$

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

$$\hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right) \approx \hat{\beta}_{r,LS}^{i,sim} - 0.96\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right) \approx \hat{\beta}_{r,ML}^{i,sim}\left(\phi^{Data}\right), \tag{19}$$

where $\hat{\beta}_{r,ML}^{i,sim}\left(\phi^{Data}\right)$ is calculated using the true autoregressive parameter, ϕ^{Data}, in the simulated model. Recall further that the ML estimator of β_r, with the least squares estimate of ϕ treated as the known true value, reduces to the OLS estimator of β_r (equation (7)). That is, $b_r^{Data} = \hat{\beta}_{r,ML}^{Data}\left(\phi^{Data}\right)$, where $\hat{\beta}_{r,ML}^{Data}\left(\phi^{Data}\right)$ is the ML estimator of β_r in the original data, treating $\phi^{Data} = \hat{\phi}_{LS}^{Data}$ as the true value of ϕ. Defining $b_{r,ML}^{Data}\left(\phi^{Data}\right) \equiv$

$\hat{\beta}_{r,ML}^{Data}\left(\phi^{Data}\right) = b_r^{Data}$, equations (17) and (19) thus give,

$$\Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \approx \Pr\left(\hat{\beta}_{r,ML}^{i,sim}\left(\phi^{Data}\right) \geq b_r^{Data}\right) = \Pr\left(\hat{\beta}_{r,ML}^{i,sim}\left(\phi^{Data}\right) \geq b_{r,ML}^{Data}\left(\phi^{Data}\right)\right). \tag{20}$$

The test based on $\hat{\beta}_{d,LS}$ in Cochrane's simulation-based approach therefore turns out to be very similar to a test using the ML estimator of β_r, when ϕ is treated as fixed and known and equal to the OLS estimate in the data. The perceived gains from using the test based on β_d is therefore not surprising, given that the test based on the ML estimator is (asymptotically) efficient (see discussion in Campbell and Yogo, 2006). This near numerical equality with the ML estimator is, of course, somewhat accidental, and depends on the specific values of σ_{13} and σ_{33}. More generally, the functional form of the ML estimator and the implied estimator used in Cochrane's test are identical—an adjusted version of the OLS estimator—but the weight put on the adjustment term will typically differ; the weight applied in the ML procedure is asymptotically efficient and, in Section 3.4 below, we analyze how much efficiency is lost in the dividend-growth based test for different values of σ_{13} and σ_{33}.

In summary, the simulation-based test is similar to an ML-based test, using the OLS estimate of ϕ as the "true" value of the AR parameter. As discussed in Section 2.2, such an approach can lead to severely biased inference, and the Monte Carlo simulations in Section 4 illustrate that this is indeed the case.

3.3 Altering the value of ϕ in the simulations

What happens if the value for ϕ 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 β_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 ϕ. In particular, if one sets the

AR parameter in the simulations equal to some maximum feasible value for ϕ, say ϕ^{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. 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.

3.4 Relative efficiency of ML and the β_d-based approach

As discussed previously, if one treats ϕ as completely unknown, and estimates this parameter with OLS, it follows that $\hat{\beta}_{r,LS} = \hat{\beta}_{d,LS} + \left(1 - \rho\hat{\phi}_{LS}\right)$. In this case, the dividend-growth based test of no return predictability uses the coefficient combination $\hat{\beta}_{d,LS} + \left(1 - \rho\hat{\phi}_{LS}\right)$ and clearly $Var\left(\hat{\beta}_{r,LS}\right) = Var\left(\hat{\beta}_{d,LS} + \left(1 - \rho\hat{\phi}_{LS}\right)\right)$. If one instead treats ϕ as "known", which is implicitly done in the simulation-based test described above, the relevant variance of the dividend-growth based parameter combination is given by $Var\left(\hat{\beta}_{d,LS} + (1 - \rho\phi)\right) = Var\left(\hat{\beta}_{d,LS}\right) \neq Var\left(\hat{\beta}_{r,LS}\right)$. As shown by the above discussion, the test based on $\hat{\beta}_{d,LS} + (1 - \rho\phi)$, with ϕ known, is close to the test based on the ML estimator of β_r when ϕ is treated as known. In order to get a sense of how similar the two approaches are, it is illustrative to compare the variance of $\hat{\beta}_{d,LS}$ with the variance of the ML estimator $\hat{\beta}_{r,ML}$.

As shown in Appendix C, the ratio of the (asymptotic) variances of $\hat{\beta}_{d,LS}$ and $\hat{\beta}_{r,ML}$ is given by,

$$\frac{Var(\hat{\beta}_{d,LS})}{Var(\hat{\beta}_{r,ML})} = \frac{1 + \frac{\sigma_{33}}{\sigma_{11}}\rho^2 + 2\rho\delta_{13}\sqrt{\frac{\sigma_{33}}{\sigma_{11}}}}{1 - \delta_{13}^2}, \tag{21}$$

where $\delta_{13} = \sigma_{13} \big/ \sqrt{\sigma_{11}\sigma_{33}}$ is the correlation between ϵ_t^r and ϵ_t^{dp}. It is also shown that this ratio is always greater than or equal to unity, which is not surprising given that ML estimation is efficient. Equivalence of the variances is achieved for $\rho = -\sigma_{13}/\sigma_{33} = -\delta_{13}\sqrt{\sigma_{11}/\sigma_{33}}$. In Cochrane's data,

$$\delta_{13} = -0.7, \quad \sqrt{\sigma_{11}} = 0.196, \quad \sqrt{\sigma_{33}} = 0.153. \tag{22}$$

Fixing, $\sqrt{\sigma_{11}} = 0.196$, $\sqrt{\sigma_{33}} = 0.153$, and varying the correlation δ_{13} one obtains different values for the ratio, presented in Table 1.

[Insert Table 1 here]

By coincidence, the actual parameter values in Cochrane's data imply a near equivalence between the ML estimator and the dividend-growth based estimator. As is seen in Table 1, this holds less true as the correlation δ_{13} increases in absolute value towards unity.

3.5 Alternative present-value formulations

Binsbergen and Koijen (2010) also specify a predictive model under a present-value relationship. Their model specification is arguably more elaborate than the one used in Cochrane (2008), relying on a latent-variable formulation for expected returns and expected dividend growth. Importantly, unlike Cochrane (2008), Binsbergen and Koijen (2010) perform a full joint estimation of all parameters in their model, including those specifying the time-series persistence of the variables. There are therefore no prior values imposed on any parameters. Within their modelling framework, Binsbergen and Koijen (2010) also specify a test of the null of no return predictability.

We briefly analyze the properties of this test of no return predictability, as the Binsbergen and Koijen (2010) framework might partly be viewed as a more general version of the ideas proposed by Cochrane (2008). Specifically, in Appendix D, we perform a simulation exercise of the Binsbergen and Koijen (2010) model and show that the corresponding test of return predictability also suffers from severe size distortions, similar to those we document for Cochrane's test in the next section. A test with a nominal size of five percent exhibits an actual rejection rate of 34 percent under the null hypothesis, using a simulated model with parameter values corresponding to those estimated in Binsbergen and Koijen (2010) with the null hypothesis of no return predictability imposed.

These results for the Binsbergen and Koijen (2010) model add to the overall message of our analysis. Namely, while present-value models are useful for furthering our understanding of return predictability, they do not seem to contribute much in terms of formulating better-behaved tests of the actual null of return predictability. Differently put, if one is already (mostly) certain that returns are predictable, the present value formulation of, for instance, Binsbergen and Koijen (2010), is a good device for understanding and interpreting this return predictability. It is not a particularly great device for establishing that predictability exists in the first place.

4 Monte Carlo simulations

This section provides evidence from Monte Carlo simulations, illustrating the theoretical results derived above. That is, these simulations provide finite-sample rejection rates, both under the null and the alternative of return predictability, for the procedures described previously in the paper. 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 simulation.

4.1 Simulation setup

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 parameter restriction in equation (R1) thus applies, and given a hypothesis for β_r and a fixed value for ϕ, the dividend-growth coefficient β_d is implicitly determined. The innovations, ϵ_t, are drawn from an *iid* normal distribution and their covariance matrix Σ satisfies the restrictions stated in equation (3). Intercepts are fitted in all regressions, although these are not commented on below. All simulation results are based on 50,000 repetitions.

For comparability, we use the empirical estimates presented in Table 2 in Cochrane (2008) to parametrize the model used in the simulations. Similarly, we use the same sample size of $T = 78$ in each simulation draw. These parameter values are also reproduced in the current paper in Panel A in Table 2.[14] As discussed above, the unknown AR coefficient ϕ plays the key role in determining the sampling properties of the OLS estimators of the main parameters β_r and β_d in the model. The extent to which the value of ϕ impacts these estimators is mostly determined by the absolute value of the correlation $\delta_{13} = \sigma_{13}/\sqrt{\sigma_{11}\sigma_{33}}$ (see discussion in Campbell and Yogo, 2006).[15] We therefore show results for alternative values of ϕ and δ_{13}. The empirical estimate of δ_{13} in Cochrane's original data is -0.70, but we also show results for $\delta_{13} = -0.95$. When changing δ_{13} from -0.70 (the estimate in the data) to the alternative value -0.95, σ_{11} and σ_{33} are kept fixed, and σ_{13} is adjusted to reach the new value of δ_{13}. The other elements of the covariance matrix Σ are adjusted accordingly to satisfy the restrictions in equation (3). For ϕ, we consider values from 0.9 to 0.995 in the size simulations, and values of 0.94 (Cochrane's original estimate) and 0.99 in the power simulations.

[14] The small deviations between the estimates presented here, in Panel A in Table 2, and those presented in Table 2 in Cochrane (2008) are, as far as we can tell, due to data revisions in the CRSP indexes.

[15] $\delta_{23} = \sigma_{23}/\sqrt{\sigma_{11}\sigma_{33}}$ plays the key role for the estimator of β_d, but for a given δ_{13}, δ_{23} is pinned down by the restrictions in equation (3).

4.2 Implementation of Cochrane's procedure

Cochrane's procedure is based on obtaining the finite sample distribution of the OLS estimators for β_r and β_d under the null of no return predictability, and calculating p-values for the estimated coefficients based on this distribution. To compare the size and power of Cochrane's procedure with other tests, we fix the nominal size of Cochrane's test procedure to five percent in a one-sided test against a positive alternative. That is, in each round of the simulations, we reject the null hypothesis of no return predictability for p-values less than or equal to 0.05.

In order to implement Cochrane's procedure in a repeated Monte Carlo simulation, a "two-layered" simulation is implemented. For a given set of true parameter values, the procedure can most easily be summarized by the below steps. For a simulation with 50,000 repetitions, steps 1-4 are repeated 50,000 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 Σ. Denote the OLS estimates of β_r, β_d, and ϕ as $\hat{\beta}_{r,LS}^{sim}$, $\hat{\beta}_{d,LS}^{sim}$, and $\hat{\phi}_{LS}^{sim}$, 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, $\beta_r = 0$, $\beta_d = -\left(1 - \rho\hat{\phi}_{LS}^{sim}\right)$, 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 β_r and β_d.

4. Based on the resulting empirical distributions of the OLS estimators in step (3),

calculate the p-values for the coefficient estimates $\hat{\beta}^{sim}_{r,LS}$ and $\hat{\beta}^{sim}_{d,LS}$, 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 than or greater than 0.05.

Each initial simulated sample path in step (1) leads to a simulation of another 50,000 sample paths in step (3), and a straightforward implementation of this simulation exercise, with 50,000 repetitions, results in $50000^2 = 2.5$ billion samples being simulated and estimated. Since this needs to be re-done for all different combinations of parameter values that we use in the size and power simulations below, such a naive approach is not computationally feasible. We describe in Appendix E how the simulation is actually implemented, in a manner such that the same results are achieved as if one followed steps (1)-(4).

The above steps refer to the usual implementation of Cochrane's procedure, where the parametrization in step (3) uses the OLS estimate of ϕ obtained in step (2). Alternative implementations differ only in their treatment of ϕ in step (3). Specifically, in the power simulations, we consider either the infeasible test where ϕ is assumed to be known,[16] or the conservative test where ϕ is set equal to some $\phi^{Max} \approx 1$. Thus, in these alternative implementations, the value of ϕ used to parametrize the simulated system in step (3) is set to either the same ("true") value used in the simulation in step (1), or some upper bound ϕ^{Max} (set equal to 0.995 in the below implementation). The value of β_d is adjusted accordingly to ensure that the null of no return predictability is imposed. I.e., $\beta_d = -\left(1 - \rho\phi^{True}\right)$ or $\beta_d = -\left(1 - \rho\phi^{Max}\right)$, respectively.

When simulating the size and power properties of other estimators that use known asymptotic critical values (e.g., the standard OLS t-test), we simply calculate the corresponding test statistics from each of the 50,000 samples generated in the repetitions of step (1), and obtain the empirical rejection frequencies.

[16] In the Monte Carlo simulations, ϕ *is* known, and we use the tests relying on this information (which is not available in actual empirical situations) as benchmarks.

4.3 Size

In the typical implementation of Cochrane's approach, it is assumed that the AR parameter ϕ is equal to its OLS estimate in the data. However, the least squares estimator for the AR parameter suffers from a well-known downward bias (Kendall, 1954). In consequence, Cochrane's test specifies an AR parameter that is likely smaller than the true value in the underlying data and the test based on the dividend-growth coefficient is likely over-sized, per the discussion in Section 3.2. This is illustrated in Figure 1, which shows the actual size of Cochrane's tests based on the simulation procedure described above. Panel A in the figure shows results for $\delta_{13} = -0.7$ (the estimate in Cochrane's original data) and Panel B shows results for $\delta_{13} = -0.95$. The size of the tests are plotted as a function of the true parameter value ϕ in the data generating process. As a comparison, the size of the standard OLS-based t-test of return predictability is also presented. That is, this is the test based on the standard t-statistic, defined as t_{β_r} in equation (28) in Appendix A, using standard critical values (i.e., reject for $t_{\beta_r} > 1.65$). It is well known that the standard OLS t-test suffers from size distortions (e.g., Stambaugh, 1999), but as is evident, the dividend-growth based test actually performs considerably worse in terms of size. This is true for any underlying AR parameter ϕ, with rejections rates in excess of 20 percent in all cases for the dividend-growth based test when $\delta_{13} = -0.7$. For $\delta_{13} = -0.95$, the rejection rates exceed 40 percent for most values of ϕ.[17] As is also seen in Figure 1, Cochrane's test based on the return coefficient is not at all as sensitive to the exact value of ϕ in the simulations, and is somewhat under-sized when one parametrizes the simulation with the OLS estimate of the autoregressive parameter.

Given these large size distortions, there is little use in analyzing the power properties of this implementation of Cochrane's dividend-growth based test. Instead, in the next

[17] As discussed throughout the paper, and analyzed further below, Cochrane's test is similar to an ML-based test treating the OLS estimate of ϕ as the true known value of the AR coefficient. Although not shown in Figure 1, the ML-based test thus implemented would suffer from similar size distortions to the β_d-based test.

sub-section we consider the power of the infeasible test with ϕ known, and a conservative test where ϕ is assumed to be less than or equal to some upper bound ϕ^{Max}.

[Insert Figure 1 here]

4.4 Power

The size distortions shown in Figure 1 result from a lack of knowledge of the true value of ϕ. If one had *a-priori* knowledge of the AR parameter ϕ, one would re-simulate a correctly parameterized system in step (3) of the simulations described above, and there would be (almost) no size distortions in the simulation-based test.[18] However, such knowledge of ϕ would also enable the use of the full information ML estimator in equation (5) and the corresponding likelihood-ratio test proposed by Campbell and Yogo (2006). The power of the dividend-growth based test in this situation is therefore most relevantly compared with the power of the ML-based test. The ML-based test is the likelihood-ratio 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). This test is asymptotically normally distributed, and the results below are based on rejections of the null for test statistics greater than 1.65 (a one-sided five percent test).

Figure 2 shows power curves for the ML-based test and the dividend-growth based test, with both tests using knowledge of the true value of ϕ. Both tests have negligible size distortions. In terms of power, the ML-based test dominates the dividend-growth based test for all alternatives close to the null, although under Cochrane's original parametrization ($\phi = 0.94, \delta_{13} = -0.7$) the two power curves are exceptionally close to each other (Panel A1). For some alternatives more distant from the null, Cochrane's test is marginally more powerful than the ML-based test, although the actual differences are very small.

[18] The other parameters (specifically Σ) of the model are still assumed unknown, but the impact of this parameter uncertainty is very small compared to a lack of knowledge of ϕ.

The power of Cochrane's test for β_r is also shown, and it is seen to be rather inefficient in comparison to the other two tests, as it does not make any real use of the knowledge of ϕ.

[Insert Figure 2 here]

Figure 2 also highlights that the similarity between the ML-based test and the dividend-growth based test depends crucially on the specific value of the correlation between the innovations of the dividend-price ratio and the returns $\left(\delta_{13} = \sigma_{13}/\sqrt{\sigma_{33}\sigma_{11}}\right)$. If $\delta_{13} = -0.7$, the ML and dividend-growth based tests are very close to each other (Panels A1 and A2 in Figure 2). However, if one increases the absolute value of δ_{13}, the dividend-growth based test loses power relative to the ML-based test (Panels B1 and B2 in Figure 2). These results echo the relative efficiency of the ML estimator and the dividend-growth estimator discussed in Section 3.4. Regardless of parameter values, virtually no power gains can be achieved if one uses the dividend-growth equation to test for return predictability, compared to the full information ML test.[19]

Specific knowledge of ϕ is not available in practice, and the power results shown in Figure 2 are therefore mostly of a benchmark nature. However, as discussed previously, a feasible test of return predictability can be achieved when the assumption $\phi \leq \phi^{Max}$, for some given ϕ^{Max}, is imposed. A conservative test is then obtained by forming the likelihood ratio test using ϕ^{Max}. Similarly, one can perform Cochrane's tests in a conservative way, setting the AR parameter equal to ϕ^{Max} in step (3) of the simulations. Below, we consider conservative tests of this kind, setting $\phi^{Max} = 0.995$, thus ruling out non-stationary dividend-price ratio processes.

Panel A1 in Figure 3 shows that the conservative ML-based test and the conservative dividend-growth based test are close to each other if the true parameters of the model

[19] The ML estimator is asymptotically efficient and, as argued by Campbell and Yogo (2006), the corresponding likelihood ratio test is conditionally uniformly most powerful. The instances where the dividend-growth based test is seen to achieve somewhat greater power than the ML-based test should therefore be finite sample effects.

correspond to the values in Cochrane (2008). In particular, if the true value of the AR parameter is close to its OLS estimate in Cochrane's data ($\phi = 0.94$), the simulation-based test performs slightly better than the ML-based procedure (Panels A1 and B1). However, for $\phi = 0.99$, the ML-based test is generally more powerful (Panels A2 and B2 in Figure 3). For alternatives close to the null and with $\phi = 0.94$, the power of the return-based simulation test is larger than the power of both the ML-based and the dividend-growth based test. This result reflects the fact that Cochrane's return-based test (which is a direct test of the return coefficient, using critical-values from a simulated distribution) does not turn nearly as conservative as the other two tests when the actual value of ϕ is far away from ϕ^{Max}. However, as the alternative drifts further away from the null hypothesis, the return-based test quickly loses in relative power.[20]

[Insert Figure 3 here]

The purpose of these simulations is to highlight the size distortions present in the typical implementation of the dividend-growth based test, and to show that in a comparison with the ML-based test, there are virtually no power gains. However, we do not in any way attempt to provide a full treatment of the best practice for conducting inference in predictive regressions.

5 Empirical results

We end with an empirical illustration of the analytical results. We use the annual value-weighted CRSP returns, including and excluding dividends, to calculate the dividend-

[20]Comparing the power of conservative tests is difficult since the tests are, by definition, not correctly sized in general. Hence we urge some caution in interpreting the power simulations for the conservative tests. One way of interpreting the results in Figure 3 is that the ML approach makes optimal use of the value postulated for the AR parameter. Therefore, the dividend-growth based test performs better for $\phi << \phi^{Max}$ because it makes less good use of the value of ϕ in the first place. The return-based simulation test is an extreme outcome of this reasoning: It makes no direct use of the knowledge of ϕ and therefore achieve the greatest power in some cases. These results highlight some of the limitations of the straightforward conservative implementations of the infeasible tests.

price ratio and the dividend growth. Inflation is the monthly change in the Consumer Price Index, and the risk-free rate is the interest rate on the 3-month Treasury Bill. The estimates of equations (E1)-(E3), not explicitly imposing the Campbell-Shiller present-value restrictions, are shown in Table 2. Panel A1 uses real returns between 1927 and 2004, reproducing Cochrane's parameter values.[21] Updating the data set until 2016, and/or changing to excess returns, does not materially change the estimates (Panels A2, B1, and B2 in Table 2).

[Insert Table 2 here]

Next, we compare how Cochrane's simulation-based tests relate to other tests of return predictability. Table 3 shows the p-values for several one-sided tests against a positive alternative. As in Table 2, the results in Table 3 are very similar for real and excess returns, as well as across the two different sample periods. In all cases, the simulation-based β_d-test, using the OLS estimate of ϕ in the simulated model, shows the lowest, or close to the lowest, p-values. These p-values are all around 0.02, and thus signal strong evidence against the null hypothesis of no return predictability. Likewise, the standard, asymptotically motivated, OLS t-test leads to p-values that strongly reject the null hypothesis. However, as documented in the analysis of this paper, the simulation-based β_d-test is likely very biased and over-sized, and so is the standard OLS t-test due to the well-known Stambaugh (1999) bias. If one instead performs the dividend-growth based test in a conservative way, using $\phi^{Max} = 0.995$, the p-values are around 0.07 and 0.11 in the samples ending in 2004 (Panels A1 and B1) and 2016 (Panels A2 and B2), respectively. Thus, by adopting a more robust test, the evidence in favor of predictability weakens substantially and the null of no predictability can no longer be rejected at the five percent significance level. The ML-based test, also formulated in a conservative manner using $\phi^{Max} = 0.995$, leads to very similar p-values, in line with the theoretical findings showing

[21] As mentioned in Footnote 14, the negligible deviations from the values reported in Table 2 in Cochrane (2008) are, as far as we can tell, due to data revisions in the CRSP indexes.

a near-equivalence between the simulation-based dividend-growth test and a likelihood approach. Finally, the simulation-based β_r-test (i.e., the test based directly on the return coefficient, but using p-values from the simulated model), yields p-values around 0.20 and thus provides no evidence against the null of no return predictability. This test was shown to control size reasonably well in the Monte Carlo simulations, but also to typically lack power relative to the other tests (see Figure 3).

Figure 4 further illustrates the sensitivity of the test results with respect to the assumptions made on the AR coefficient ϕ. In particular, the figure shows the p-values of the ML test and the dividend-growth test as a function of the maximum value specified for ϕ. The results are based on the sample with real returns ending in 2016 (thus corresponding to Panel A2 in Table 3). As is seen, if one is willing to assume that $\phi \leq 0.97$, one starts to find significant results.

[Insert Table 3 here]

[Insert Figure 4 here]

Our empirical analysis is mostly meant to be illustrative, but the empirical results presented in Table 3 and Figure 4 are very much in line with many previous studies. Without imposing additional assumptions, we find that tests that are robust to the bias inherent in these types of predictive regressions are at best borderline significant. Figure 4 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.

6 Conclusion

Cochrane (2008) shows how the present value identity of Campbell and Shiller (1988) can be used to relate return and dividend-growth predictability, and he argues that a test based on dividend growth is more powerful than the ordinary return-based test. In this study, we provide a detailed analysis of the properties of the dividend-growth based test of return predictability, and our main results can be summarized as follows. (i) If one uses standard OLS inference, the return and dividend-growth based tests are identical. (ii) Cochrane's specific dividend-growth based test is more powerful than the ordinary return-based test because the former makes strong use of the assumption on the exact degree of persistence in the predictor variable. If one uses a similarly strong assumption in the return-based test, one can formulate an (asymptotically) even more powerful likelihood-based test. Thus, if one uses the same information set when formulating the dividend-growth based and return-based tests, there are no power gains to the dividend-growth based approach. (iii) In the standard formulation, the dividend-growth based test suffers from severe size distortions. These can be controlled by altering the assumption on the persistence in the dividend-price ratio, but again there is virtually no power gain compared to a maximum-likelihood based approach.

In the wake of Cochrane's article, there is now a large literature that analyzes both dividend-growth and return predictability. Our study does not in any way invalidate these works. However, it highlights that the specific test proposed by Cochrane does not add power over and above previously existing tests, once the same assumptions are used for both types of tests. Importantly, our analysis also shows that unless one specifically imposes a conservative approach to the test formulation, the test procedure will tend to over reject the null hypothesis of no return predictability.

We also briefly consider the properties of the return predictability test in Binsbergen and Koijen (2010). The model in Binsbergen and Koijen might be seen as a refinement and extension of the ideas initially proposed in Cochrane (2008) and their estimation and testing methods do not impose any prior assumptions on the time-series persistence in the data. However, our simulation results show that the test of the null of no return predictability in the Binsbergen and Koijen framework also suffers from severe size distortions. In light of the large literature on inference in predictive regressions this is not particularly surprising, as any method not explicitly dealing with the so-called Stambaugh (1999) bias is likely to suffer problems. Together with our main analysis of Cochrane's procedure, these findings highlight that while extending the simple predictive regression to more elaborate present-value frameworks help provide a deeper economic understanding of return predictability, it does not help escape the basic inferential issues associated with testing for predictability.

A Properties of the OLS 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 ϵ^r, ϵ^d, and ϵ^{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

$$\hat{\beta}_{LS} = \begin{bmatrix} \hat{\beta}_{r,LS} \\ \hat{\beta}_{d,LS} \\ \hat{\phi}_{LS} \end{bmatrix} = \beta + \left(X_{-1}'X_{-1}\right)^{-1} \begin{bmatrix} X_{-1}'\epsilon^r \\ X_{-1}'\epsilon^d \\ X_{-1}'\epsilon^{dp} \end{bmatrix}. \tag{23}$$

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

$$\begin{aligned} \hat{\beta}_{r,LS} &= \beta_r + \left(X_{-1}'X_{-1}\right)^{-1} X_{-1}'\epsilon^r \\ &= \left(\beta_d + \left(X_{-1}'X_{-1}\right)^{-1} X_{-1}'\epsilon^d\right) + \left(1 - \rho\left(\phi + \left(X_{-1}'X_{-1}\right)^{-1} X_{-1}'\epsilon^{dp}\right)\right) \\ &= \hat{\beta}_{d,LS} + \left(1 - \rho\hat{\phi}_{LS}\right). \end{aligned} \tag{24}$$

Further, letting $M_{X_{-1}} = I - X_{-1}\left(X_{-1}'X_{-1}\right)^{-1} X_{-1}'$, and using standard results, the fitted residuals satisfy,

$$\begin{aligned} \hat{\epsilon}^r &\equiv r - X_{-1}\hat{\beta}_{r,LS} \\ &= M_{X_{-1}}\epsilon^r \\ &= M_{X_{-1}}\left(\epsilon^d - \rho\epsilon^{dp}\right) \\ &= \hat{\epsilon}^d - \rho\hat{\epsilon}^{dp}. \end{aligned} \tag{25}$$

In order to motivate the formulation of the standard t-statistics below, suppose that

$\phi < 1$ and that ϵ_t is an *iid* or martingale difference sequence.[22] Under classical asymptotic results, it follows straightforwardly that as the sample size $T \to \infty$,

$$\sqrt{T}\left(\hat{\beta}_{LS} - \beta\right) \xrightarrow{d} N(0, \left(Var\left(x_{t-1}^2\right)\right)^{-1} \Sigma) \equiv N\left(0, \Omega\right), \tag{26}$$

where Σ is given in (3), and $Var\left(x_{t-1}^2\right) = \frac{\sigma_{33}}{1-\phi^2}$. For ϕ close to one, the asymptotic distribution result stated in (26) does not hold up well in finite samples. However, our main purpose for stating this result is to motivate the standard t-statistics considered below, and show that their standard formulation leads to an equivalence between the return based and the dividend-growth based tests, analogous to that for the actual coefficients seen in equation (24) above.

Let $\hat{\epsilon} = \left(\hat{\epsilon}^r, \hat{\epsilon}^d, \hat{\epsilon}^{dp}\right)$ be the matrix of fitted residuals. A standard estimator of the asymptotic covariance matrix in (26) is given by

$$\hat{\Omega} = \left(\frac{1}{T}\hat{\epsilon}'\hat{\epsilon}\right)\left(\frac{1}{T}X_{-1}'X_{-1}\right)^{-1} = \begin{bmatrix} \hat{\epsilon}_r'\hat{\epsilon}_r & \hat{\epsilon}_r'\hat{\epsilon}_d & \hat{\epsilon}_r'\hat{\epsilon}_{dp} \\ \hat{\epsilon}_r'\hat{\epsilon}_d & \hat{\epsilon}_d'\hat{\epsilon}_d & \hat{\epsilon}_d'\hat{\epsilon}_{dp} \\ \hat{\epsilon}_r'\hat{\epsilon}_{dp} & \hat{\epsilon}_d'\hat{\epsilon}_{dp} & \hat{\epsilon}_{dp}'\hat{\epsilon}_{dp} \end{bmatrix} \left(\sum_{t=2}^{T} x_{t-1}^2\right)^{-1}, \tag{27}$$

where the first diagonal element of $\hat{\Omega}$ corresponds to the asymptotic variance of $\hat{\beta}_r$. The standard t-statistic for testing the null hypothesis of $\beta_r = 0$ is thus given by,

$$t_{\beta_r} = \frac{\hat{\beta}_{r,LS}}{\sqrt{\frac{1}{T}\left(\hat{\epsilon}^{r\prime}\hat{\epsilon}^r\right)\left(X_{-1}'X_{-1}\right)^{-1}}} = \frac{\hat{\beta}_{r,LS}}{\sqrt{\frac{1}{T}\left(\hat{\epsilon}^{d\prime}\hat{\epsilon}^d + \rho^2\hat{\epsilon}^{dp\prime}\hat{\epsilon}^{dp} - 2\rho\hat{\epsilon}^{d\prime}\hat{\epsilon}^{dp}\right)\left(X_{-1}'X_{-1}\right)^{-1}}}, \tag{28}$$

where the last equality follows from the result in (25). To formulate the dividend-growth based t-statistic of the null of no return predictability, let $c = (0, 1, -\rho)$. The t-statistic

[22] The stationarity condition ($\phi < 1$) is by no means necessary for our main analysis, but it enables us to frame the main properties of the OLS estimator of equations (E1)-(E3) in terms of classical asymptotic results for stationary models.

is then given by

$$
\begin{aligned}
t_{\beta_d+1-\rho\phi} &= \frac{\hat{\beta}_{d,LS} + \left(1-\rho\hat{\phi}_{LS}\right)}{\sqrt{\frac{1}{T}\left(c'\left(\hat{\epsilon}'\hat{\epsilon}\right)c\right)\left(X_{-1}'X_{-1}\right)^{-1}}} \\
&= \frac{\hat{\beta}_{r,LS}}{\sqrt{\frac{1}{T}\left(\hat{\epsilon}^{d\prime}\hat{\epsilon}^{d} + \rho^2\hat{\epsilon}^{dp\prime}\hat{\epsilon}^{dp} - 2\rho\hat{\epsilon}^{d\prime}\hat{\epsilon}^{dp}\right)\left(X_{-1}'X_{-1}\right)^{-1}}} = t_{\beta_r}. \qquad (29)
\end{aligned}
$$

The standard t-statistics, t_{β_r} and $t_{\beta_d+1-\rho\phi}$, are thus numerically identical, provided the restrictions in (R1) and (R2) hold.

B The simulation-based test parametrized with ϕ^{Max}

Before analyzing the simulation-based test, recall first the "conservative" ML estimator in equation (6),

$$\hat{\beta}_{r,ML}\left(\phi^{Max}\right) = \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}}\left(\phi^{Max} - \hat{\phi}_{LS}\right). \qquad (30)$$

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 ϕ^{Max}. Provided the true value of the AR parameter in the data is less than or equal to ϕ^{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 ϕ^{Max}.[23] Let $\hat{\beta}^{i,sim}_{r,ML}\left(\phi^{Max}\right)$ be the ML estimator of β_r in the simulations, and let $b^{Data}_{r,ML}\left(\phi^{Max}\right) \equiv \hat{\beta}^{Data}_{r,ML}\left(\phi^{Max}\right)$ be the conservative ML estimate of β_r in the actual data. In the simulated data $\hat{\beta}^{i,sim}_{r,ML}\left(\phi^{Max}\right)$ is the "correct" ML estimator, since the simulated model has an AR parameter equal to ϕ^{Max}, whereas in the original data the estimator is conservative (provided the true AR parameter is less than or equal

[23] This formulation is without loss of generality, as the actual derivations make no use of the assumption that ϕ^{Max} is indeed larger than the true value for the AR parameter in the data.

to ϕ^{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

$$\Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \approx \Pr\left(\hat{\beta}_{r,ML}^{i,sim}\left(\phi^{Max}\right) \geq b_{r,ML}^{Data}\left(\phi^{Max}\right)\right). \tag{31}$$

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

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

$$\hat{\beta}_{r,ML}^{i,sim}\left(\phi^{Max}\right) = \hat{\beta}_{r,LS}^{i,sim} + \frac{\sigma_{13}}{\sigma_{33}}\left(\phi^{Max} - \hat{\phi}_{LS}^{i,sim}\right) \approx \hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Max} - \hat{\phi}_{LS}^{i,sim}\right). \tag{32}$$

As in Section 3.2, let $b_r^{Data} \equiv \hat{\beta}_{r,LS}^{Data}$ and $b_d^{Data} \equiv \hat{\beta}_{d,LS}^{Data}$. By restriction (R1),

$$b_d^{Data} = b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{Data}\right) = b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) + \rho\left(\hat{\phi}_{LS}^{Data} - \hat{\phi}_{LS}^{i,sim}\right), \tag{33}$$

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

$$b_{r,ML}^{Data}\left(\phi^{Max}\right) \approx b_r^{Data} - \rho\left(\phi^{Max} - \hat{\phi}_{LS}^{Data}\right). \tag{34}$$

It follows that

$$\begin{aligned}
& \Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \\
= \ & \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) + \rho\left(\hat{\phi}_{LS}^{Data} - \hat{\phi}_{LS}^{i,sim}\right)\right) \\
= \ & \Pr\left(\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data} + \rho\left(\hat{\phi}_{LS}^{Data} - \hat{\phi}_{LS}^{i,sim}\right)\right) \\
= \ & \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Max} - \hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data} - \rho\left(\phi^{Max} - \hat{\phi}_{LS}^{Data}\right)\right) \\
\approx \ & \Pr\left(\hat{\beta}_{r,ML}^{i,sim}\left(\phi^{Max}\right) \geq b_{r,ML}^{Data}\left(\phi^{Max}\right)\right). \qquad (35)
\end{aligned}$$

C Variance of the estimators

Using the joint convergence for $\hat{\beta}_{LS} = \left(\hat{\beta}_{r,LS}, \hat{\beta}_{d,LS}, \hat{\phi}_{LS}\right)$ in equation (26), along with the shape of Σ given in (3), the asymptotic variance of the ML estimator of β_r can be expressed as follows,

$$\begin{aligned}
Var\left(\hat{\beta}_{r,ML}\right) &= Var\left(\hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}}\left(\phi - \hat{\phi}_{LS}\right)\right) \\
&= Var\left(\hat{\beta}_{r,LS}\right) + \frac{\sigma_{13}^2}{\sigma_{33}^2} Var\left(\hat{\phi}_{LS}\right) - 2\frac{\sigma_{13}}{\sigma_{33}} Cov\left(\hat{\beta}_{r,LS}, \hat{\phi}_{LS}\right) \\
&= Var\left(\hat{\beta}_{r,LS}\right) + \frac{\sigma_{13}^2}{\sigma_{33}^2}\frac{\sigma_{33}}{\sigma_{11}} Var\left(\hat{\beta}_{r,LS}\right) - 2\frac{\sigma_{13}}{\sigma_{33}}\frac{\sigma_{13}}{\sigma_{11}} Var\left(\hat{\beta}_{r,LS}\right) \\
&= Var\left(\hat{\beta}_{r,LS}\right)\left(1 - \frac{\sigma_{13}^2}{\sigma_{33}\sigma_{11}}\right) \\
&= Var\left(\hat{\beta}_{r,LS}\right)\left(1 - \delta_{13}^2\right). \qquad (36)
\end{aligned}$$

Similarly, using $\hat{\beta}_{d,LS} = \hat{\beta}_{r,LS} - \left(1 - \rho\hat{\phi}_{LS}\right)$,

$$\begin{aligned}
Var\left(\hat{\beta}_{d,LS}\right) &= Var\left(\hat{\beta}_{r,LS} - \left(1 - \rho\hat{\phi}_{LS}\right)\right) \\
&= Var\left(\hat{\beta}_{r,LS}\right) + \rho^2 Var\left(\hat{\phi}_{LS}\right) + 2\rho Cov\left(\hat{\beta}_{r,LS}, \hat{\phi}_{LS}\right) \\
&= Var\left(\hat{\beta}_{r,LS}\right)\left(1 + \rho^2\frac{\sigma_{33}}{\sigma_{11}} + 2\rho\frac{\sigma_{13}}{\sigma_{11}}\right). \qquad (37)
\end{aligned}$$

Thus,

$$\frac{Var(\hat{\beta}_{d,LS})}{Var(\hat{\beta}_{r,ML})} = \frac{1+\rho^2\frac{\sigma_{33}}{\sigma_{11}}+2\rho\frac{\sigma_{13}}{\sigma_{11}}}{1-\delta_{13}^2} = \frac{1+\rho^2\frac{\sigma_{33}}{\sigma_{11}}+2\rho\delta_{13}\sqrt{\frac{\sigma_{33}}{\sigma_{11}}}}{1-\delta_{13}^2}, \tag{38}$$

and

$$\begin{aligned}
\frac{Var(\hat{\beta}_{d,LS})}{Var(\hat{\beta}_{r,ML})}-1 &= \frac{1+\rho^2\frac{\sigma_{33}}{\sigma_{11}}+2\rho\delta_{13}\sqrt{\frac{\sigma_{33}}{\sigma_{11}}}-(1-\delta_{13}^2)}{1-\delta_{13}^2} \\
&= \frac{\rho^2\frac{\sigma_{33}}{\sigma_{11}}+2\rho\sqrt{\frac{\sigma_{33}}{\sigma_{11}}}\delta_{13}+\delta_{13}^2}{1-\delta_{13}^2} \\
&= \frac{\left(\rho\sqrt{\frac{\sigma_{33}}{\sigma_{11}}}+\delta_{13}\right)^2}{1-\delta_{13}^2} \\
&\geq 0.
\end{aligned} \tag{39}$$

D The Binsbergen-Koijen model

In the Binsbergen and Koijen (2010) model, expected returns (μ_t) and expected dividend growth rates (g_t) are both viewed as latent state variables, satisfying the transition equations,

$$\hat{g}_t = \gamma_1 \hat{g}_{t-1} + \epsilon_t^g, \tag{40}$$

$$\hat{\mu}_t = \delta_1 \hat{\mu}_{t-1} + \epsilon_t^\mu, \tag{41}$$

where $\hat{g}_t = g_t - \gamma_0$ is the demeaned expected dividend growth and $\hat{\mu}_t = \mu_t - \delta_0$ the demeaned expected returns. Observed dividend growth is thus given by

$$\Delta d_t = \gamma_0 + \hat{g}_{t-1} + \epsilon_t^d. \tag{42}$$

Combined with the Campbell and Shiller (1988) present value identity, the observation equation for the price-dividend ratio is given by

$$pd_t = A - B_1\hat{\mu}_t + B_2\hat{g}_t, \tag{43}$$

where A, B_1, and B_2 are combinations of underlying parameters. The expected return $\hat{\mu}_t$ can be substituted out from (43), giving the final state space system,

$$\Delta d_t = \gamma_0 + \hat{g}_{t-1} + \epsilon_t^d, \tag{44}$$

$$\hat{g}_t = \gamma_1\hat{g}_{t-1} + \epsilon_t^g, \tag{45}$$

$$pd_t = (1-\delta_1)A + B_2(\gamma_1 - \delta_1)\hat{g}_{t-1} + \delta_1 pd_{t-1} - B_1\epsilon_t^\mu + B_2\epsilon_t^g. \tag{46}$$

The system is identified by imposing $Corr(\epsilon_t^d, \epsilon_t^g) = 0$ and is estimated using maximum likelihood via the Kalman filter. Full details of the model and the estimation procedure are found in Binsbergen and Koijen (2010).

Testing for return predictability in this system amounts to testing the following joint coefficient restrictions,

$$\delta_1 = \sigma_\mu = \rho_{\mu g} = \rho_{\mu d} = 0, \tag{47}$$

where $\sigma_\mu^2 = Var\left(\epsilon_t^\mu\right)$, $\rho_{\mu g} = Corr\left(\epsilon_t^\mu, \epsilon_t^g\right)$, and $\rho_{\mu d} = Corr\left(\epsilon_t^\mu, \epsilon_t^d\right)$. The test statistic used by Binsbergen and Koijen (2010) is the likelihood ratio test $LR = 2(\mathcal{L}_1 - \mathcal{L}_0)$, where $\mathcal{L}_1$ is the log-likelihood of the unconstrained model and $\mathcal{L}_0$ is the log-likelihood of the constrained model. Under standard assumptions, the LR test statistic is asymptotically distributed as a χ_4^2 distribution.

To evaluate the actual size of this test, we simulate the system under the null of no return predictability. The parameter values under the null hypothesis are set to the restricted point estimates of Binsbergen and Koijen (2010), where the restriction (47) is imposed. These values are reported in the first row of table IA.D in Binsbergen and Koijen (2010), and reproduced here in Table A1.[24] The sample size in each simulation draw is set to $T = 62$, corresponding to the empirical sample size in the original study.

[Insert Table A1 here]

Using the parametrization under the null of no return predictability, $10,000$ samples are simulated, and the maximum likelihood estimates of the parameters of both the restricted and unrestricted models are obtained. The likelihood-ratio test statistic is computed for each sample and its empirical distribution is shown in Figure A1, along with the asymptotic χ^2_4 distribution. As is seen, the actual distribution of the test statistic is substantially to the right of the asymptotic distribution, and using asymptotic critical values results in a substantial over-rejection of the null hypothesis. Specifically, in the simulations, $\Pr(LR > \chi^2_{4,0.05}) = 0.34$, such that the actual size of the test using a five percent nominal size is 34 percent. As a comparison, the plain OLS t-test, simply calculated from a standard regression of returns on the lagged dividend-price ratio using the simulated data from the Binsbergen and Koijen (2010) model, results in a rejection rate of 22 percent for a one-sided test (and 34 percent for a two-sided test) with a five percent nominal size.

[Insert Figure A1 here]

E Implementation of Monte Carlo simulations

We improve the speed of the simulations by re-using the empirical distributions we obtain in the "second layer "of the simulation (steps (3) and (4)). In order to do so, we define a matrix CV, with a generic column CV_j. CV_j contains the parametrization of a simulation

[24]This model formulation and the corresponding parameter values refer to the case with cash-reinvested dividends.

in step (3) along with the critical values for $\hat{\beta}^{sim}_{r,LS}$ and $\hat{\beta}^{sim}_{d,LS}$, which correspond to the 95th percentiles of the empirical distributions in the given simulation. That is $CV_j = \left(\hat{\phi}^{j,sim}_{LS}, \hat{\delta}^{j,sim}_{13}, \hat{\sigma}^{j,sim}_{33}, \hat{\sigma}^{j,sim}_{11}, \hat{\beta}^{j,critical}_{r,LS}, \hat{\beta}^{j,critical}_{d,LS}\right)'$.[25]

In each round i for any parameter specification, before we run the simulation step (3), we check whether we have already stored a close parameter combination in the CV matrix. Two parameter combinations are close to each other if the conditions

$$(\hat{\phi}^{j,sim}_{LS} - \hat{\phi}^{i,sim}_{LS})^2 < 0.005^2$$
$$(\hat{\delta}^{j,sim}_{13} - \hat{\delta}^{i,sim}_{13})^2 < 0.01^2$$
$$(\hat{\sigma}^{j,sim}_{33} - \hat{\sigma}^{i,sim}_{33})^2 < 0.002^2$$
$$(\hat{\sigma}^{j,sim}_{11} - \hat{\sigma}^{i,sim}_{11})^2 < 0.002^2$$

are *jointly* satisfied. The tolerance levels allow for approximately one percent deviation for ϕ and δ_{13}, and slightly more than five percent for the variances. These choices of tolerance levels reflect the fact that the autoregressive parameter and the correlation are the key quantities determining the shape of the distribution (and thus the critical values). Diagnostic checks strongly indicate that using these tolerance levels lead to critical values that are very close to the "exact" simulation-based critical values, obtained for a given set of parameters, and that this speed-improvement does not affect the results of the Monte Carlo simulations.

If there is a column CV_j that includes a parameter combination close to the one we obtained in step (2), we skip step (3) and go directly to step (4). We classify as rejection/non-rejection depending on whether the estimates $\hat{\beta}^{sim}_{r,LS}$ and $\hat{\beta}^{sim}_{d,LS}$ are less or greater than the respective critical values stored in CV_j. If there is no such CV_j, then we

[25]Note that the four parameters $(\phi, \delta_{13}, \sigma_{33}, \sigma_{11})$ and the restrictions imposed by the Campbell and Shiller (1988) decomposition completely characterize a given parametrization under the null hypothesis of no return predictability.

proceed by step (3). At the end of step (3), we store the parameters and the critical values in a new column in CV. The procedure can be summarized as the following modification of the steps described in the main text.

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 Σ. Denote the OLS estimates of β_r, β_d, and ϕ as $\hat{\beta}^{sim}_{r,LS}$, $\hat{\beta}^{sim}_{d,LS}$, and $\hat{\phi}^{sim}_{LS}$, respectively. Check if there is a CV_j column in the matrix CV, whose parameter values are close to the estimates obtained. If yes, critical values $\hat{\beta}^{j,critical}_{r,LS}$ and $\hat{\beta}^{j,critical}_{d,LS}$ are used, and one proceeds directly to step (4'). If no, proceed to step (3').

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, $\beta_r = 0$, $\beta_d = -\left(1 - \rho\hat{\phi}^{sim}_{LS}\right)$, 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 β_r and β_d.

   Choose the 95th percentiles of the empirical distributions in the simulation as critical values. Store the parameters obtained in (2'), along with the critical values, as a new column in the matrix CV.

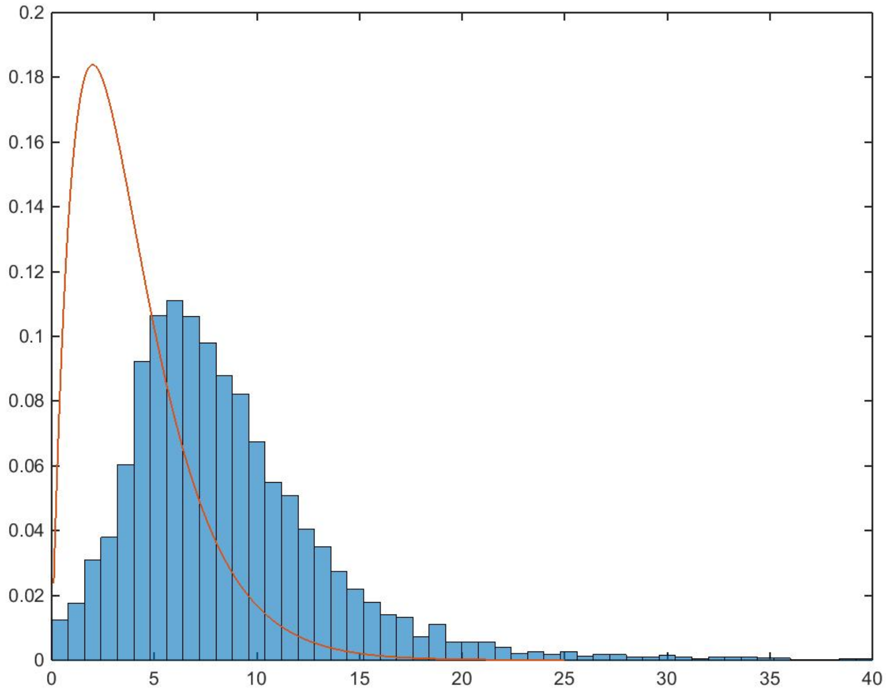
4'. Compare the coefficient estimates $\hat{\beta}^{sim}_{r,LS}$ and $\hat{\beta}^{sim}_{d,LS}$, obtained from the initial simulated sample in steps (1') and (2') to the critical values obtained in either step (2') or step (3'). Classify as rejection/non-rejection depending on whether the respective estimates are less than or greater than the corresponding critical values.

Table A1: **Parameter values for the simulated Binsbergen and Koijen (2010) model**

The table shows parameter estimates of the restricted state space model, imposing the null of no return predictability. The values are taken from Table IA.D in Binsbergen and Koijen (2010) and correspond to the model with cash-reinvested dividends.

Parameter	Estimate (restricted model)
γ_0	0.0637
γ_1	0.9900
δ_0	0.0936
δ_1	0
σ_g	0.0065
σ_d	0.0659
σ_μ	0
$\rho_{\mu g}$	0
$\rho_{\mu d}$	0

Figure A1: **Empirical distribution of the likelihood-ratio test of return predictability in Binsbergen and Koijen (2010)**



Notes: The figure shows the empirical distribution of the likelihood-ratio test for the null hypothesis of no return predictability in the state space model proposed by Binsbergen and Koijen (2010). The solid line represents the χ^2_4 distribution. The empirical distribution is obtained from the Monte Carlo simulation of the model under the null of no return predictability, using $10,000$ repetitions.

References

Amihud, Y. and Hurvich, C. M. (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 39(04):813–841.

Amihud, Y., Hurvich, C. M., and Wang, Y. (2008). Multiple-predictor regressions: Hypothesis testing. *Review of Financial Studies*, 22(1):413–434.

Amihud, Y., Hurvich, C. M., and Wang, Y. (2010). Predictive regression with order-p autoregressive predictors. *Journal of Empirical Finance*, 17(3):513–525.

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.

Boudoukh, J., Michaely, R., Richardson, M., and Roberts, M. R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *The Journal of Finance*, 62(2):877–915.

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.

Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric Theory*, 11(05):1131–1147.

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.

Detzel, A. L. and Strauss, J. (2016). The dog has barked for a long time: Dividend growth is predictable. Technical report.

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. (2017). Four Centuries of Return Predictability. *Journal of Financial Economics*, Forthcoming.

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.

Lacerda, F. and Santa-Clara, P. (2010). Forecasting dividend growth to better predict returns. *Manuscript Universidade Nova de Lisboa.*

Lettau, M. and Ludvigson, S. (2010). Measuring and Modelling Variation in the Risk-Return Trade-off. CEPR Discussion Papers 3105, C.E.P.R. Discussion Papers.

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

Lettau, M. and Van Nieuwerburgh, S. (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.

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.

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.

Table 1: **Relative efficiency of the ML and dividend-growth based procedures**

The table shows the ratio of the (asymptotic) variances of the OLS estimator in the dividend-growth equation ($\hat{\beta}_{d,LS}$) and the maximum likelihood estimator in the return equation ($\hat{\beta}_{r,ML}$). The ratios of variances are tabulated as a function of the correlation $\delta_{13} = \sigma_{13}/\sqrt{\sigma_{11}\sigma_{33}}$.

Correlation δ_{13}	$\frac{Var(\hat{\beta}_{d,LS})}{Var(\hat{\beta}_{r,ML})}$
-0.5	1.083
-0.6	1.035
-0.7	1.005
-0.8	1.007
-0.9	1.119
-0.95	1.413
-0.99	3.909

Table 2: **Empirical estimates of the model**

The table shows the empirical estimates of the predictive model, based on annual data. Panels A1 and A2 show the results for real returns, with samples dating from 1927 to 2004 and 2016, respectively. Panels B1 and B2 show the results for excess returns, with samples dating from 1927 to 2004 and 2016, respectively. The first column in each panel shows the least squares estimates of the coefficients, and the second column shows the heteroscedasticity robust standard errors. The "implied" values are calculated from the other two least squares coefficient estimates and the restriction $\beta_r = \beta_d + 1 - \rho\phi$. The diagonal of the last three columns are the volatilities of the error terms, while the off-diagonal elements are their correlations.

Panel A1: Real returns, 1927-2004

	$\hat{\beta}, \hat{\phi}$	$\sigma(\hat{\beta})$	**implied**	**r**	Δd	**dp**
r	0.098	0.051	0.102	19.6	65.7	-69.9
Δd	0.008	0.045	0.004	65.7	14.1	8
dp	0.94	0.047	0.944	-69.9	8	15.4

Panel A2: Real returns, 1927-2016

	$\hat{\beta}, \hat{\phi}$	$\sigma(\hat{\beta})$	**implied**	**r**	Δd	**dp**
r	0.089	0.046	0.09	19.4	67.9	-68.2
Δd	-0.002	0.039	-0.003	67.9	14.3	7.3
dp	0.939	0.04	0.941	-68.2	7.3	14.8

Panel B1: Excess returns, 1927-2004

	$\hat{\beta}, \hat{\phi}$	$\sigma(\hat{\beta})$	**implied**	**r**	Δd	**dp**
r	0.111	0.051	0.102	19.5	65	-67.5
Δd	0.008	0.045	0.017	65	14.1	8
dp	0.94	0.047	0.931	-67.5	8	15.4

Panel B2: Excess returns, 1927-2016

	$\hat{\beta}, \hat{\phi}$	$\sigma(\hat{\beta})$	**implied**	**r**	Δd	**dp**
r	0.094	0.047	0.09	19.5	67.5	-65.8
Δd	-0.002	0.039	0.002	67.5	14.3	7.3
dp	0.939	0.04	0.935	-65.8	7.3	14.8

Table 3: **Empirical significance of the predictability tests**

The table shows p-values from different tests of return predictability. Panels A1 and A2 show the results for real returns, with samples dating from 1927 to 2004 and 2016, respectively. Panels B1 and B2 show the results for excess returns, with samples dating from 1927 to 2004 and 2016, respectively. The first row in each panel presents the p-value for the simulation-based test using the dividend-growth coefficient, with the AR parameter set to the OLS estimate. The second row uses the standard OLS t-test in the return equation (based on heteroscedasticity robust standard errors), and the third row corresponds to the conservative simulation-based test using the dividend-growth coefficient (with the AR parameter set to $\phi^{Max} = 0.995$). The fourth row shows the p-value for the conservative ML-based test, again using $\phi^{Max} = 0.995$. The fifth and final row in each panel shows the p-value for the simulation-based test using the return coefficient, with the AR parameter in the simulations set to the OLS estimate. The p-values for the OLS t-test and the conservative ML-test are based on the normal distribution.

Panel A1: Real returns, 1927-2004

	p-value
Simulation β_d-test	0.017
OLS t-test	0.028
Conservative Simulation β_d-test	0.074
Conservative ML-test	0.102
Simulation β_r-test	0.218

Panel A2: Real returns, 1927-2016

	p-value
Simulation β_d-test	0.023
OLS t-test	0.027
Conservative Simulation β_d-test	0.111
Conservative ML-test	0.130
Simulation β_r-test	0.216

Panel B1: Excess returns, 1927-2004

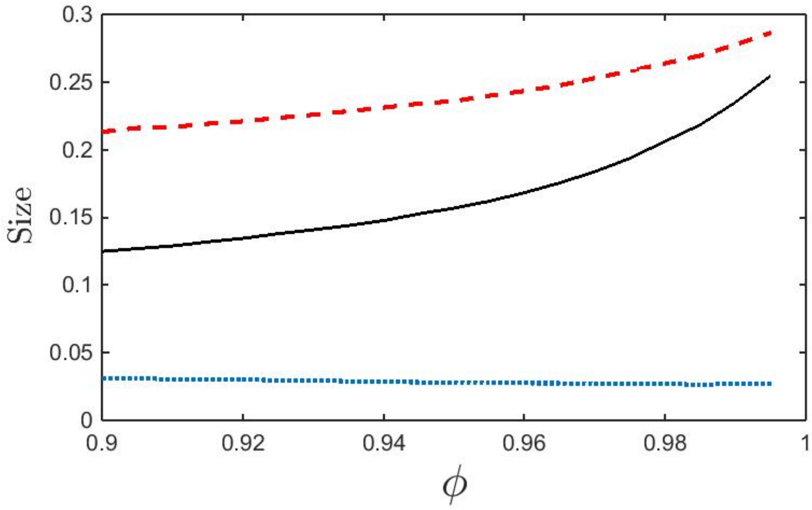
	p-value
Simulation β_d-test	0.017
OLS t-test	0.015
Conservative Simulation β_d-test	0.073
Conservative ML-test	0.055
Simulation β_r-test	0.181

Panel B2: Excess returns, 1927-2016

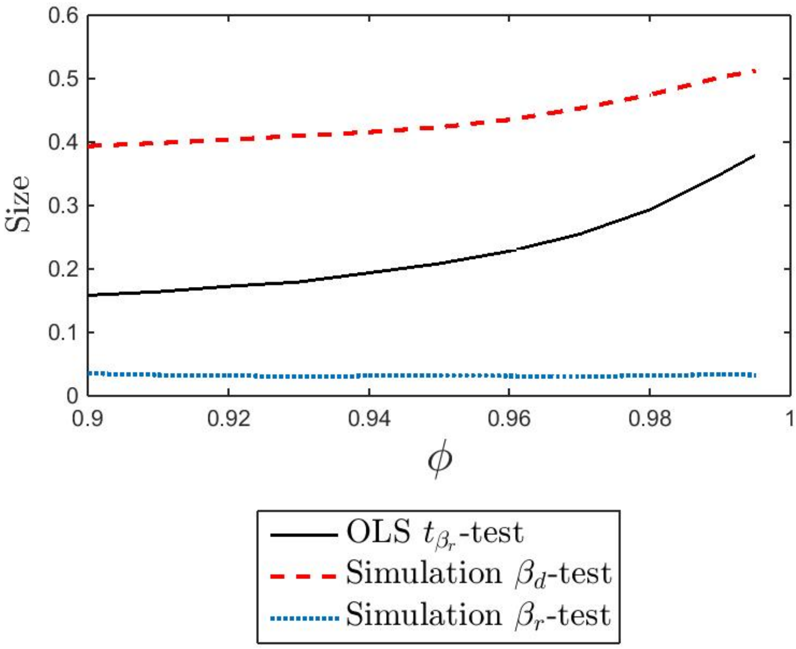
	p-value
Simulation β_d-test	0.022
OLS t-test	0.022
Conservative Simulation β_d-test	0.108
Conservative ML-test	0.103
Simulation β_r-test	0.197

Figure 1: **Size distortions using $\hat{\phi}_{LS}$**

Panel A: $\delta_{13} = -0.7$

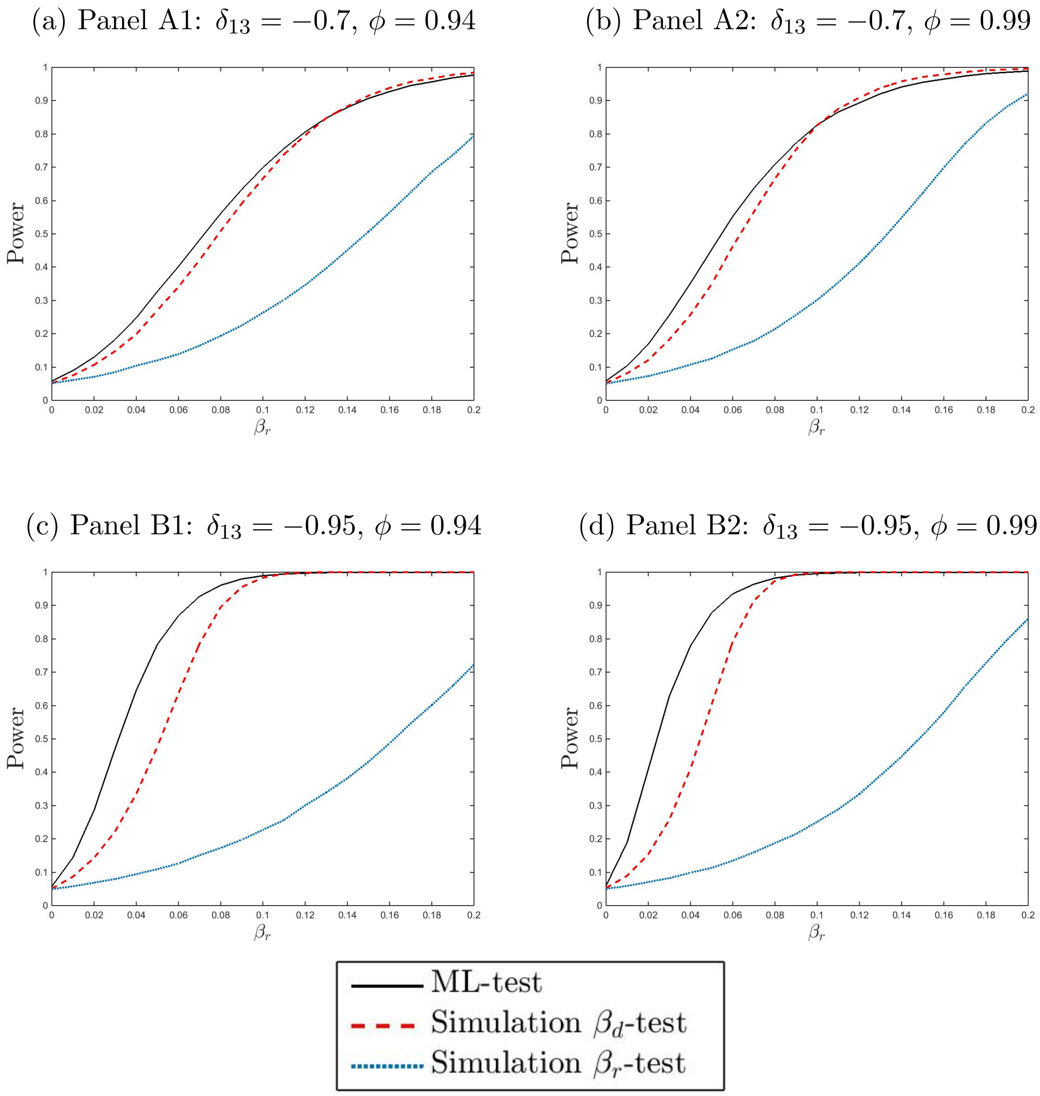


Panel B: $\delta_{13} = -0.95$



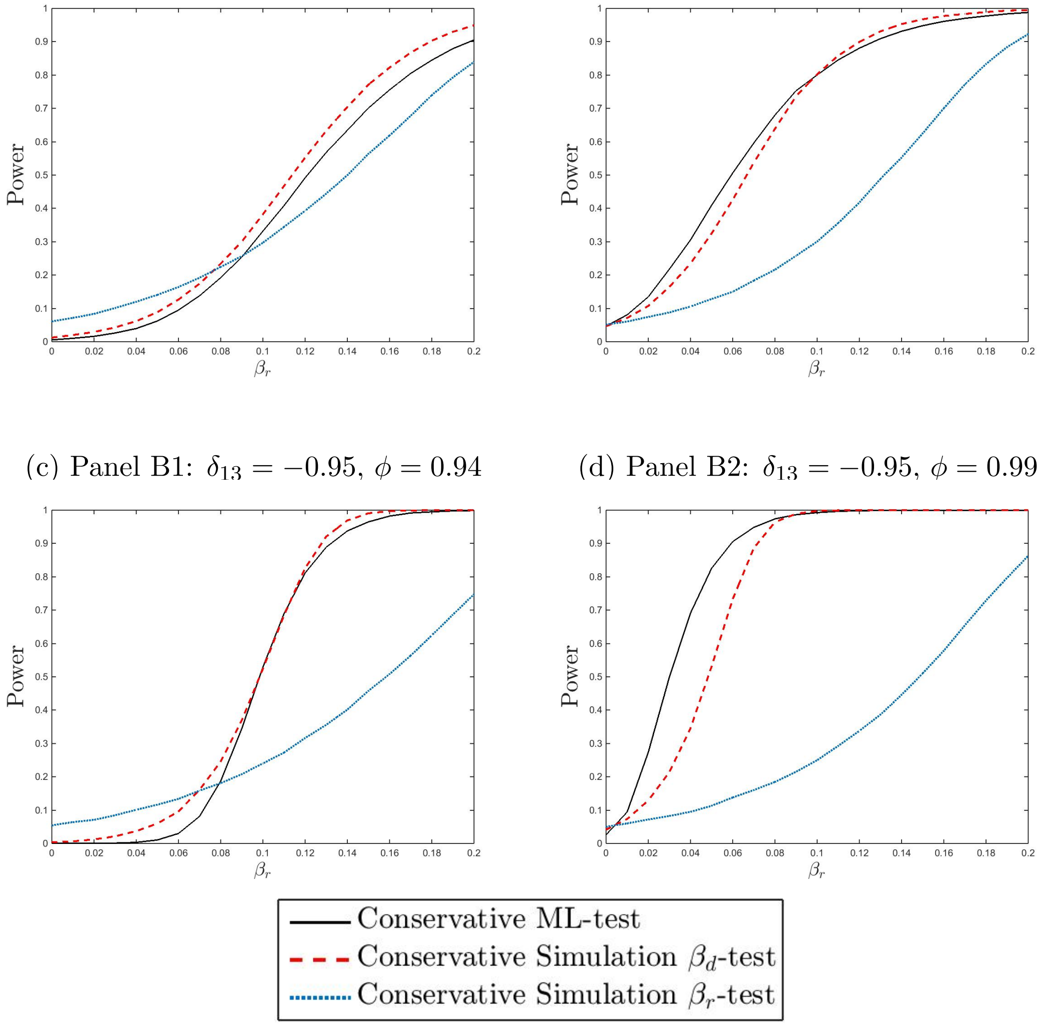
Notes: The figure shows the actual size of one-sided five percent tests of the null hypothesis of $\beta_r = 0$ against a positive alternative. That is, the graphs show the average rejection rates, for the corresponding tests, under the null hypothesis of no return predictability. The horizontal axis shows the true value of ϕ in the return system. The dashed lines represent the simulation-based test using the dividend-growth coefficient (β_d), and the dotted lines represent the simulation-based test using the return coefficient (β_r), with both tests using the OLS estimate $\hat{\phi}_{LS}$ as the value for the AR parameter. The solid lines represent the rejection rates of the standard OLS t-test in the return equation (t_{β_r}), using critical values from the standard normal distribution (i.e., reject for $t_{\beta_r} > 1.65$). The results are based on the Monte Carlo simulation described in the main text, and the average rejection rates are calculated over 50,000 repetitions.

Figure 2: **Power of infeasible tests with known ϕ**



Notes: The figure shows the power curves of tests of return predictability, with known ϕ. Specifically, the graphs show the average rejection rates for one-sided five percent tests of the null hypothesis of $\beta_r = 0$ against a positive alternative. The x-axis shows the true value of the parameter β_r, and the y-axis indicates the average rejection rate (i.e., power). Panels A1 and B1 (A2 and B2) show results for return systems where the true value of ϕ is equal to 0.94 (0.99). The solid lines represent the ML-based test (using critical values from the standard normal distribution), the dashed lines represent the simulation-based test using the dividend-growth coefficient (β_d), and the dotted lines represent the simulation-based test using the return coefficient (β_r). All three tests use knowledge of the true value of ϕ. The results are based on the Monte Carlo simulation described in the main text, and the power is calculated as the average rejection rates over 50,000 repetitions.

Figure 3: **Power of feasible conservative tests**



(a) Panel A1: $\delta_{13} = -0.7$, $\phi = 0.94$

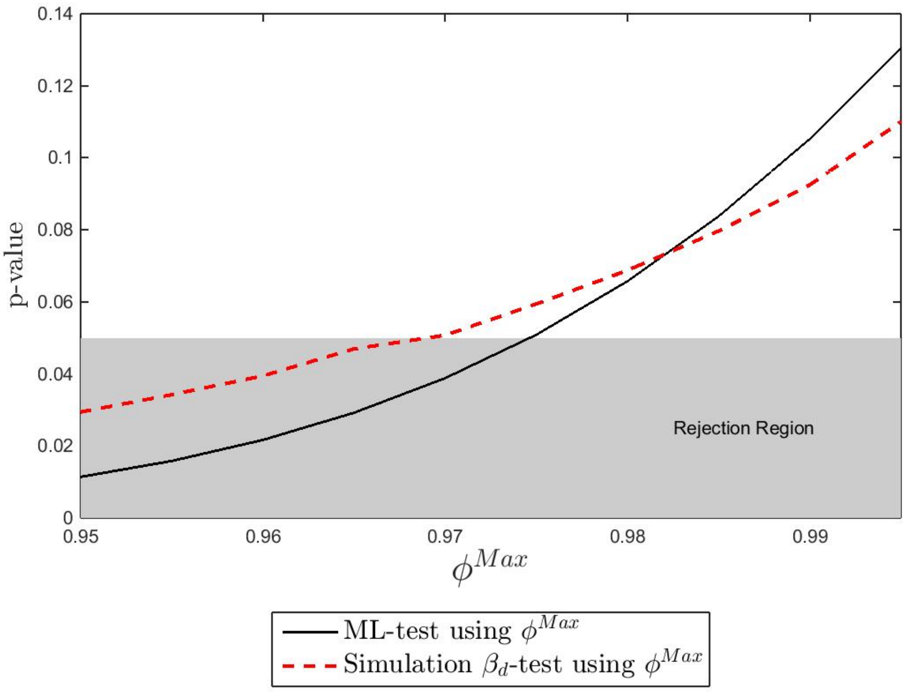
(b) Panel A2: $\delta_{13} = -0.7$, $\phi = 0.99$

(c) Panel B1: $\delta_{13} = -0.95$, $\phi = 0.94$

(d) Panel B2: $\delta_{13} = -0.95$, $\phi = 0.99$

Notes: The figure shows the power curves of conservative tests of return predictability, with $\phi^{Max} = 0.995$. Specifically, the graphs show the average rejection rates for one-sided five percent tests of the null hypothesis of $\beta_r = 0$ against a positive alternative. The x-axis shows the true value of the parameter β_r, and the y-axis indicates the average rejection rate (i.e., power). Panels A1 and B1 (A2 and B2) show results for return systems where the true value of ϕ is equal to 0.94 (0.99). The solid lines represent the ML-based test (using critical values from the standard normal distribution), the dashed lines represent the simulation-based test using the dividend-growth coefficient (β_d), and the dotted lines represent the simulation-based test using the return coefficient β_r. All three tests specify a conservative value for the autoregressive parameter, set equal to $\phi^{Max} = 0.995$. The results are based on the Monte Carlo simulation described in the main text, and the power is calculated as the average rejection rates over 50,000 repetitions.

Figure 4: **Empirical p-values as a function of ϕ^{Max}**



Notes: The graph shows the p-values of tests of return predictability, using the full sample (1927-2016) and real returns. The p-values are plotted as functions of the value specified for the upper bound, ϕ^{Max}, on the AR parameter. The solid line represents the ML-based test (using critical values from the standard normal distribution), and the dashed line represents the simulation-based test using β_d. Both tests specify that the value for the autoregressive parameter is set to the corresponding value for ϕ^{Max}.