Size and Power in Asset Pricing Tests

September 4, 2018

Abstract

We study the size-power tradeoff in commonly employed tests of return predictability. We provide conditions under which short-horizon dividend-growth tests and long-horizon return tests are asympotically more powerful than short-horizon return tests. Monte Carlo results show that the asymptotic power advantages carry over to small samples, although the reasons are different. Asymptotically, dividend-growth tests are close to uniformly most powerful. In small samples, dividend-growth tests and long-horizon tests have similar power.

JEL Codes: G12, G14, C32

Keywords: Power; Dividend-growth regressions; Long-horizon regressions; Return predictability; Market efficiency

Empirical tests of market efficiency—defined, in the context of stocks, as the proposition that asset prices equal the discounted value of expected dividends—have a long history. In his classic paper, Fama (1970) reported empirical tests of the implication of market efficiency that asset returns are unforecastable (Samuelson (1965)). He concluded in favor of market efficiency. However, analysts subsequently observed that asset price volatility appeared to exceed the volatility consistent, under market efficiency, with the volatility of dividends (Shiller (1981)). Also, analysts reversed Fama's original conclusion, finding that asset returns are forecastable (Fama and French (1988)). The return forecastability appeared to be negligible at short horizons, but was evaluated as being much stronger at longer horizons, like several years.[1]

The fact that different tests of the same null hypothesis appear to have different outcomes suggests that some tests have greater power than others: how likely is it that different tests will reject market efficiency under some alternative—power—holding constant the rejection probability if markets are efficient—size? Possible differences in power have been discussed in the asset pricing literature (LeRoy and Steigerwald (1995), for example), but until recently no formal results along these lines had been provided.[2]

Cochrane (2008) proposed addressing the size-power tradeoff head on. He concluded from simulations conducted using data generated by an estimated

[1] Statistical issues were raised about the conclusion that returns are forecastable. In particular, the apparently dramatic difference between the outcome of short-horizon and long-horizon return forecastability tests was convincingly criticized by Boudoukh et al. (2006), who showed that in efficient markets sampling variation alone should be expected to result in estimated forecasting coefficients that increase approximately in proportion as the horizon is more distant. This occurs because return forecasts over varying horizons are highly correlated when the explanatory variable (usually the dividend yield) is highly persistent.

[2] Campbell (2001) is an exception. That paper investigates the power advantages of long-horizon return regressions, over short-horizon regressions, using approximate slope, a measure of asymptotic power, as a guide. It finds that long-horizon regressions have higher approximate slope, suggestive of higher power. Monte Carlo simulations indicate that the higher approximate slope translates to power advantages in small samples.

model that the evidence against efficiency provided by returns forecastability alone is of marginal statistical significance. However, he noted that there exist alternative tests of market efficiency. To be consistent with the observed high level of price volatility either returns or dividend growth must be significantly forecastable. Under market efficiency returns are nonforecastable by assumption, implying that efficiency can be tested by determining whether dividend forecastability is sufficient to explain price volatility. Cochrane asserted that the test based on dividend forecastability has higher power than that based on return forecastability.

Although Cochrane's stated purpose was to argue for the greater power of one test relative to the other, his discussion centered almost exclusively on the behavior of test statistics under the null hypothesis of market efficiency. In this paper we address the same problem as Cochrane, but do so directly rather than taking a detour dealing with extensive analysis of estimator properties under the null hypothesis.

We find using Cochrane's model that his conclusion in favor of greater power of dividend forecastability tests in his model is correct. This is so because, based on a normal approximation, the estimated coefficient of return forecastability has greater variance than that of dividend growth forecastability. This explanation bears no close relation to Cochrane's discussion. Without the normal approximation the power advantage of dividend forecastability tests over return foreastability tests still occurs (in fact, appears to be even greater), suggesting that intuition based on the normal approximation carries over to settings where normality is not a particularly close assumption, as in Cochrane's setting.

1 The Model

We use the same log-linear model as Cochrane: log returns r_{t+1}, log dividend growth Δd_{t+1} and next-period log dividend yield $d_{t+1} - p_{t+1}$ depend linearly on the current log yield $d_t - p_t$:

$$r_{t+1} = a_r + \beta_r(d_t - p_t) + \varepsilon^r_{t+1} \tag{1}$$

$$\Delta d_{t+1} = a_d + \beta_d(d_t - p_t) + \varepsilon^d_{t+1} \tag{2}$$

$$d_{t+1} - p_{t+1} = a_{dp} + \varphi(d_t - p_t) + \varepsilon^{dp}_{t+1}. \tag{3}$$

The errors are normal with mean zero and are independent over time. At date t the variances and covariances of the error terms are denoted σ^2_r, σ_{rd}, etc. As Cochrane noted, the Campbell and Shiller (1988) log-linearization of the definition of the rate of return,

$$r_{t+1} = \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} - (p_t - d_t), \tag{4}$$

where ρ is the constant of linearization, implies that any one of the equations (1), (2) and (3) is redundant given the others. It follows that the coefficients obey the restriction

$$\beta_r = 1 - \rho\varphi + \beta_d, \tag{5}$$

and the errors obey the restriction

$$\varepsilon^r_{t+1} = \varepsilon^d_{t+1} - \rho\varepsilon^{dp}_{t+1}. \tag{6}$$

Both restrictions play a major role in the discussion below.

1.1 The Null and Alternative Hypotheses

The null hypothesis of market efficiency is generated by setting $\beta_r = 0$, so that returns are not forecastable in the population.

The alternative hypothesis is generated by setting β_r and β_d equal to the values Cochrane estimated from the real-world data. Thus returns have a forecastable component under the alternative. In going from the alternative to the null the parameter redundancy (5) implies that setting β_r equal to zero must be accompanied by a corresponding intervention on at least one of the other parameters. Until Section 4 we follow Cochrane in specifying the null hypothesis so that β_d is changed from the value estimated from the data to the value that satisfies (5). The parameter φ and the covariances of the errors are the same under the null and the alternative.

2 Asymptotic Properties

Market efficiency can be tested either directly by determining whether β_r equals 0 or indirectly by determining whether β_d equals $\rho\phi - 1$, the value implied by (5) when β_r is set equal to 0. These tests have the same null and alternative hypotheses, but generally not the same size-power tradeoff. The asymptotic size-power tradeoff serves as a useful benchmark when evaluating the finite sample size-power tradeoff of the two tests.

2.1 Short-Horizon Tests

The asymptotic properties of OLS estimators $\widehat{\beta}_r$, $\widehat{\beta}_d$ and $\widehat{\phi}$ follow from a version of the central limit theorem; see Proposition 11.1 in Hamilton (1994) for a detailed discussion. As the number of observations goes to infinity, $T \to \infty$, the distribution of each OLS estimate approaches a normal distribution with mean equal to the population value of the corresponding parameter. The asymptotic variances of the estimators follow the standard OLS formula. For $\widehat{\beta}_r$ and $\widehat{\beta}_d$, we have

$$T\,\mathrm{Var}(\widehat{\beta}_i) = \frac{\sigma_i^2(1-\phi^2)}{\sigma_{dp}^2}; \quad i \in \{r, d\} \tag{7}$$

where σ_i^2 is the date-t variance of the shock ε_t^i, and $\sigma_{dp}^2/(1-\phi^2)$ is the variance of log dividend yield. For the estimate of the dividend-yield autocorrelation, we have $T\,\mathrm{Var}(\widehat{\varphi}) = 1 - \varphi^2$. These expressions for the variances are derived in Appendix A.

How do the asymptotic distributions change as we move from the null to the alternative? The alternative hypothesis consists of a modification in the parameters β_r and β_d from their values under the null, with the remaining parameters remaining unchanged. Therefore, as we move from the null to the alternative, the asymptotic distribution of $\widehat{\beta}_r$ shifts to reflect the change in the mean from 0 to the value under the alternative, which we specified to equal the value estimated from real-world data. The covariances remain unchanged. Similarly, the distribution of $\widehat{\beta}_d$ shifts to reflect the value of the parameter under the alternative.

With knowledge of the asymptotic distributions of the OLS estimates under the null and the alternative, we can compare the size-power tradeoff of return and dividend forecastability tests of market efficiency. The following proposition provides a simple way to determine the relative power of the two tests.

Proposition 1 *For a given size, market efficiency tests based on dividend growth forecastability are asymptotically more powerful than return forecastability tests if and only if* $\sigma_d < \sigma_r$.

The size-power tradeoff depends on the relative asymptotic variances of the OLS estimators. Because, from (7), the asymptotic variances inherit the variances of the underlying error terms, the asymptotic power of the tests depends on the error variances. If $\sigma_d < \sigma_r$, the variance of $\widehat{\beta}_d$ is lower than that of $\widehat{\beta}_r$, implying that the dividend test is more powerful. The opposite is true if $\sigma_d > \sigma_r$. The two tests have identical size-power tradeoffs if $\sigma_d = \sigma_r$.

The intuition behind Proposition 1 is demonstrated in Figure 1. Because the estimates $\widehat{\beta}_r$ and $\widehat{\beta}_d$ are asymptotically bivariate normal, the level curves under the null and the alternative are ellipses. The alternative hypothesis is a 45 degree translation of the null in the $\widehat{\beta}_r - \widehat{\beta}_d$ space. Therefore the ellipse

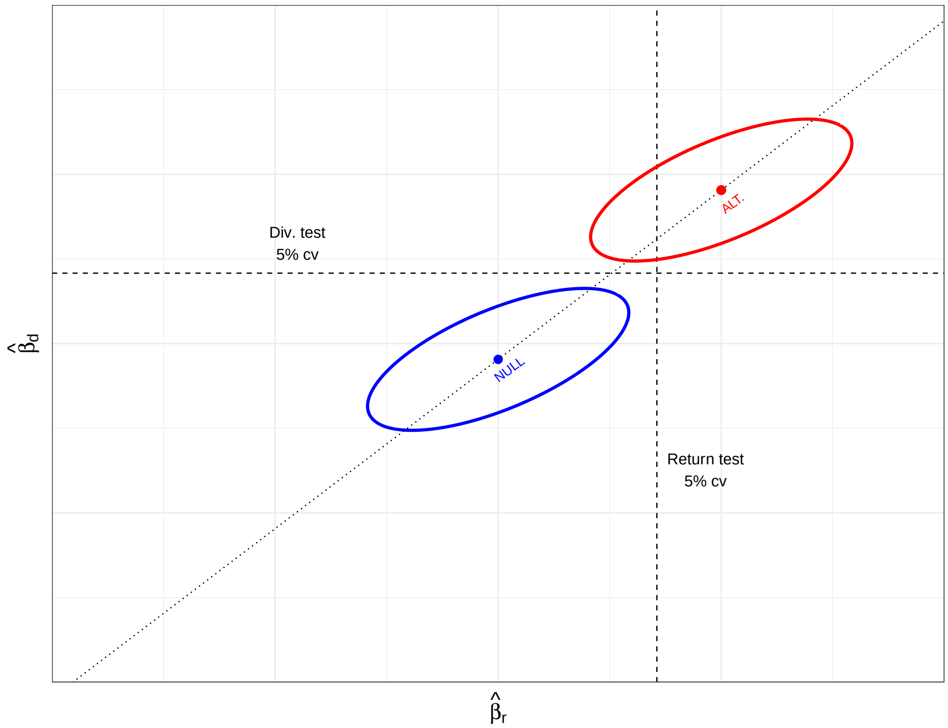

Figure 1: Level curves under the null and the alternative

representing the level curve under the alternative hypothesis is the same as that of the null hypothesis, but translated along a 45-degree line.

The ellipses are flat, reflecting the fact that the variance of $\widehat{\beta}_d$ is less than the variance of $\widehat{\beta}_r$. Together, the flatness and translation along the 45 degree line imply that the null and the alternative populations are more easily distinguished based on the $\widehat{\beta}_d$ coordinate than the $\widehat{\beta}_r$ coordinate. This fact implies that for given size the power of a market efficiency test based on $\widehat{\beta}_d$ is greater than one based on $\widehat{\beta}_r$. This result is demonstrated formally in Appendix B.

Numerical values for the size-power tradeoff are presented and discussed below.

2.2 Long-Horizon Tests

Cochrane also investigated the power of long-horizon forecastability of returns and dividend growth. He identified these with

$$\beta_r^{lh} = \beta_r/(1-\rho\phi) \tag{8}$$

and $\beta_d^{lh} = \beta_d/(1-\rho\phi)$. The coefficient $\beta_r/(1-\rho\phi)$ is relabeled β_r^{lh} because it is the forecastability coefficient associated with long-horizon returns $\sum_j \rho^j r_{t+j}$, and similarly for the dividend growth coefficient. Eq. (8) implies that $\beta_r^{lh} = 1 + \beta_d^{lh}$, from which it follows that the long-run return forecastability and dividend-growth forecastability tests are equivalent (and, in particular, have the same power for given size). Cochrane noted and discussed this result. Consequently it is sufficient to compare the size-power tradeoff for $\widehat{\beta}_r^{lh}$ with that for $\widehat{\beta}_r$, which we now do.

The size-power tradeoff for long-horizon tests depends on the asymptotic distributions of $\widehat{\beta}_r^{lh}$ under the null hypothesis and the alternative, so we begin by determining those distributions. This is done in Appendix A. As $T \to \infty$, the distribution of $\widehat{\beta}_r^{lh}$ converges to a normal distribution with mean β_r^{lh}. The asymptotic variance of $\widehat{\beta}_r^{lh}$ is given by

$$T\,\mathrm{Var}(\widehat{\beta}_r^{lh}) = \frac{1-\phi^2}{(1-\rho\phi)^2}\left(\frac{\sigma_r^2}{\sigma_{dp}^2} + \frac{2\rho\beta_r}{1-\rho\phi}\frac{\sigma_{r,dp}}{\sigma_{dp}^2} + \frac{\rho^2\beta_r^2}{(1-\rho\phi)^2}\right). \tag{9}$$

It is seen that the asymptotic variance of $\widehat{\beta}_r^{lh}$ depends on the population value of β_r. The presence of β_r in (9) implies that the asymptotic variance of $\widehat{\beta}_r^{lh}$ is generally different under the null and the alternative. Proposition 1 connecting the power of $\widehat{\beta}_r$ and $\widehat{\beta}_d$ tests with the volatilities of $\widehat{\beta}_r$ and $\widehat{\beta}_d$ depended on these variances being equal under the null and the alternative, implying that Proposition 1 does not apply directly to $\widehat{\beta}_r^{lh}$. Calculating the asymptotic size-power tradeoff, while still possible, is more involved.

Under the null hypothesis of unforecastable returns, $\beta_r = 0$, the variance of the long-horizon estimator is the variance of the short-horizon estimator scaled up. We have

$$T\,\mathrm{Var}(\widehat{\beta}_r^{lh})\Big|_{NULL} = \frac{1}{(1-\rho\phi)^2} T\,\mathrm{Var}(\widehat{\beta}_r), \tag{10}$$

from eqs. (7) and (9). The variance of the long-horizon estimator under the alternative hypothesis is greater than or less than its variance under the null, depending on the contribution of the terms in (9) involving β_r. The term $\rho^2\beta_r^2/(1-\rho\phi)^2$ is strictly positive. The sign of remaining term depends on the correlation between return shocks and dividend-yield shocks, denoted $\eta_{r,dp}$. If that correlation is positive, $\sigma_{r,dp}$ is positive and the variance under the alternative is greater than the variance under the null. If the correlation is strongly negative, the variance under the alternative is lower than that under the null. Formally, there exists a threshold $\underline{\eta}$, which is negative, such that the asymptotic variance of $\widehat{\beta}_r^{lh}$ under the null is equal to its variance under the alternative when $\eta_{r,dp} = \underline{\eta}$. We have

$$\underline{\eta} = -\rho \frac{\beta_{r,ALT}}{1-\rho\phi} \frac{\sigma_{dp}}{2\sigma_r}.$$

If $\eta_{r,dp} < \underline{\eta}$ (as will be the case in the data), the variance under the alternative is lower than the variance under the null. The opposite would be true if $\eta_{r,dp} > \underline{\eta}$.

The following two propositions show that the correlation $\eta_{r,dp}$ plays an important role in determining how the power of long-horizon tests compares to the power of $\widehat{\beta}_r$ tests.

Proposition 2 *For a given size, long-horizon tests and short-horizon return forecastability tests have equal asymptotic power if* $\eta_{r,dp} = \underline{\eta}$.

The intuition for Proposition 2 is simple: when $\eta_{r,dp} = \underline{\eta}$ the asymptotic distribution of $\widehat{\beta}_r^{lh}$ is the same as the distribution of a scalar multiple of $\widehat{\beta}_r$, so

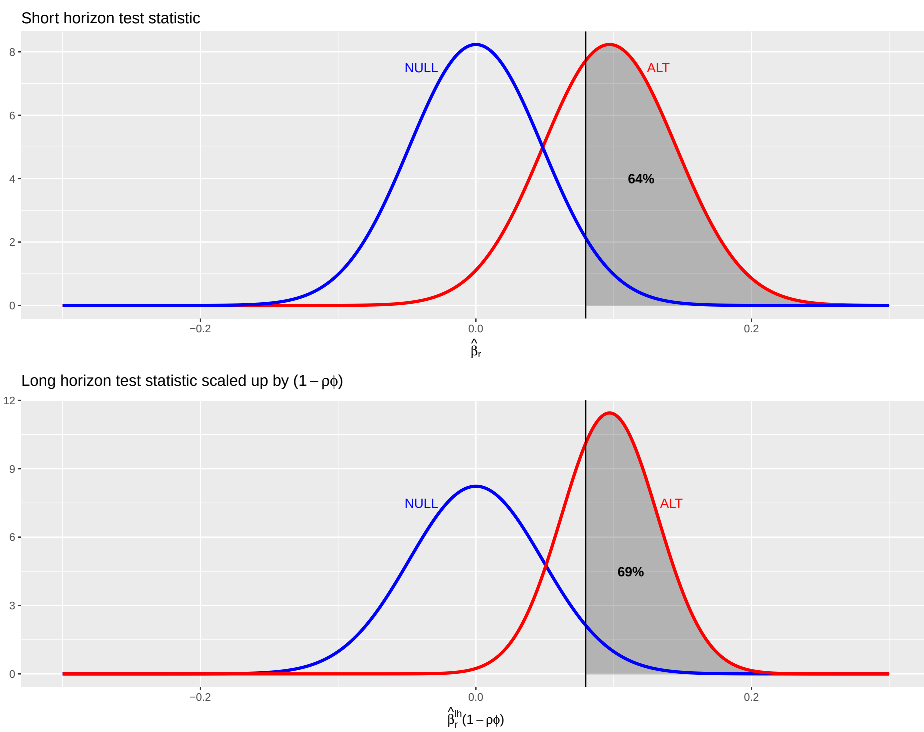

Figure 2: Why long-horizon tests have higher power than short-horizon tests

the two tests have equal power. Asymptotically, the means of the estimators $\widehat{\beta}_r^{lh}$ and $\widehat{\beta}_r$ equal their population values β_r^{lh} and β_r, respectively. Therefore, the mean of $\widehat{\beta}_r^{lh}$ is the mean of $\widehat{\beta}_r$ scaled up by $1/(1-\rho\phi)$, under both the null and the alternative. When $\eta_{r,dp} = \underline{\eta}$, the variance of $\widehat{\beta}_r^{lh}$ under the alternative equals its variance under the null. By (10), the standard deviation of $\widehat{\beta}_r^{lh}$, equal under the null and the alternative, is the standard deviation of $\widehat{\beta}_r$ scaled by $1/(1-\rho\phi)$. Because the two estimators are normally distributed under the null and the alternative, the asymptotic distribution of $\widehat{\beta}_r^{lh}$ is the same distribution as that of a scalar multiple of $\widehat{\beta}_r$. It follows that tests involving $\widehat{\beta}_r^{lh}$ and $\widehat{\beta}_r$ have equal power when $\eta_{r,dp} = \underline{\eta}$.

Now assume that $\eta_{r,dp} < \underline{\eta}$ as the data indicate; the conclusions reported below would be reversed for $\eta_{r,dp} > \underline{\eta}$. The asymptotic distribution of $\widehat{\beta}_r^{lh}$ under the null continues to be the same as the distribution of $\widehat{\beta}_r$ scaled by

$1/(1-\rho\phi)$, but the variance under the alternative is lower than under the null.

The following proposition shows that for $\eta_{r,dp} < \underline{\eta}$ the specification of the alternative hypothesis determines which test is more powerful.

Proposition 3 *Suppose $\eta_{r,dp} < \underline{\eta}$. For a given size, let $\overline{\beta}_r$ denote the critical value of the short-horizon return forecastability test. Given the size,*

1. *long-horizon tests are more powerful than short-horizon return tests if and only if $\beta_{r,ALT} > \overline{\beta}_r$. Short-horizon tests are more powerful if and only if the opposite inequality holds.*

2. *The two tests have equal power if $\beta_{r,ALT} = \overline{\beta}_r$.*

To develop intuition for Proposition 3 multiply $\widehat{\beta}_r^{lh}$ by $(1-\rho\phi)$ to undo the scale effect. The asymptotic distribution of $\widehat{\beta}_r^{lh}(1-\rho\phi)$ is identical to the distribution of $\widehat{\beta}_r$ under the null: normal with mean zero and variance given by (7). Under the alternative, the two estimators are normally distributed with equal means, but not equal variances. Which test is more powerful depends on whether it is easy to distinguish the alternative from the null; this is the case when the alternative is far away from the null. In that case, the test with the lower variance has greater power because it generates more values close to the alternative, and therefore far away from the null. Conversely, suppose that the alternative is close to the null, making it difficult to distinguish between the two. In that case, the test with the higher variance has greater power because it generates more values that are farther away from the null. To see this, consider the case in which the variance of $\widehat{\beta}_r^{lh}$ is extremely low. In that setting the probability of accepting the null if the data are generated under the alternative would be high, possibly higher than would occur if the data were generated under the null.

The threshold that determines whether or not its easy to distinguish the null and the alternative is the critical value $\overline{\beta}_r$. If $\beta_{r,ALT} > \overline{\beta}_r$ it is easy to distinguish the null and the alternative. Therefore the long-horizon test, having lower variance than the short-horizon test, provides the more powerful test. If $\beta_{r,ALT} < \overline{\beta}_r$ the short-horizon test is more powerful. If $\beta_{r,ALT} = \overline{\beta}_r$, we obtain a knife-edge result: the power of both the tests equals 0.5. The relative variance of the estimators does not matter in that case. The correlation $\eta_{r,dp}$ plays no role in determining which test is more powerful.

The fact that the relative power of short-horizon vs. long-horizon tests depends on the critical value of the short-horizon test means that the power ordering may be different depending on whether one is consider a 1%, 5% or 10% rejection probability. We will see below that the long-horizon test has higher power than the short-horizon test under 5% or 10% rejection probabilities, but lower power in the 1% case.

Figure 2 shows why the power of the long-horizon test relative to the short-horizon test depends on whether $\beta_{r,ALT}$ is greater than or less than $\overline{\beta}_r$.

2.3 Likelihood Ratio Tests

There is no reason to focus on efficiency tests based on either $\widehat{\beta}_r$ or $\widehat{\beta}_d$ to the exclusion of the other, as we have done. We can also conduct a likelihood ratio test, so that for any specified probability of Type I error the boundary between the acceptance and rejection regions consists of pairs $\widehat{\beta}_r$, $\widehat{\beta}_d$ that have equal likelihood ratios. In the case of the bivariate normal the boundaries separating these regions consist of straight lines traveling from northwest to southeast. The Neyman-Pearson lemma states that likelihood ratio tests have maximal power for given size.

3 Empirical Implementation

The asymptotic distribution may not be a good approximation of the finite sample distribution. In real-world data, the dividend yield is highly autocorrelated, which reduces the effective sample size for the estimation. Given that we only have about 80 years of real-world data, with the lower effective sample size the central limit theorem might turns out to yield only a coarse approximation of the finite sample distribution.

In this section we report the application of Monte Carlo simulations to compute both the size-power tradeoffs in both the asymptotic case and the finite-sample case. In the asymptotic case the simulations were used to estimate the parameters that determine power, as discussed in Section 2. In the finite-sample case we directly estimated power from the simulated regression coefficients. The two sets of results differ because in the latter case there is no appeal to normality, and the relevant distributions exhibit departures from normality. The computed power estimates are shown in Table 1.

We set the values of parameters used to generate the simulations to match estimates of their counterparts in the data. To facilitate comparison between our paper and Cochrane's, we take the empirical estimates from Cochrane (2008), instead of re-estimating the model using the latest available data. The Monte Carlo exercise consists of 50,000 draws with each draw consisting of 80 dates, agreeing with the length of the dataset for Cochrane's real world estimation.

The constant of log-linearization is $\rho = 0.9638$, calculated from mean log dividend yield in the data.[3] In each run, we draw two time series $\{\varepsilon_t^d\}$ and $\{\varepsilon_t^{dp}\}$ and impute the value of $\{\varepsilon_t^r\}$ using (6). The shocks are drawn from a bivariate normal distribution with mean zero and variance-covariance matrix given by

$$\begin{bmatrix} \sigma_d^2 & \\ \sigma_{dp,d} & \sigma_{dp}^2 \end{bmatrix} = \begin{bmatrix} 0.01960 & \\ 0.00161 & 0.02341 \end{bmatrix}. \tag{11}$$

Having obtained the time series of shocks, we created time series of log returns, log dividend growth and log dividend yield using (1)-(3). The autocorrelation ϕ of the log dividend yield is equal to 0.941 in both the null and the alternative hypotheses. As noted, the coefficients β_r and β_d depend

[3]We have $\rho = e^{E(p-d)}/(1 + e^{E(p-d)})$.

on whether the we assume null or the alternative hypothesis holds in the simulated data. If the null of market efficiency holds, then we have $\beta_r = 0$. Under the alternative, we have $\beta_r = 0.097$. The corresponding values of the dividend forecastability coefficient follow from (4). We have $\beta_d = -0.0931$ and $\beta_d = 0.0039$ in the null and the alternative, respectively.

For each run, we obtained the OLS estimates $\widehat{\beta}_r$ and $\widehat{\beta}_d$. For the $\widehat{\beta}_r$ test we computed the 95% critical value as the value $\overline{\beta}_r$ of $\widehat{\beta}_r$ such that 5% of the 50,000 simulated values of $\widehat{\beta}_r$ are greater than $\overline{\beta}_r$. Power was then computed as the proportion of draws under the alternative that are greater than $\overline{\beta}_r$. The calculation for $\widehat{\beta}_d$ was similar. We also computed 10% and 1% critical values and the corresponding figures for power.

The conclusions follow:

- We find that dividend-growth forecastability is more powerful than return forecastability in testing market efficiency. We already knew that this would be true in the asymptotic case from the fact that the volatility of return shocks exceeds the volatility of dividend-growth shocks. The fact that the finite-sample results support a similar conclusion suggests that the outcome does not depend critically on the normality assumption.

- The power advantage of the dividend growth tests is much more pronounced in the finite-sample case than in the asymptotic case. This probably reflects the fact that the simulated distribution of $\widehat{\beta}_r$ is skewed, whereas the simulated distribution of $\widehat{\beta}_d$ is not. Figure 3 shows these distributions under the null and the alternative. It also displays the 1%, 5% and 10% critical values. The figure makes clear why skewness results in a loss of power.

  As discussed in Stambaugh (1999), the finite-sample distribution of the OLS estimator depends on whether or not shocks to the predictor variable (dividend-yield) are correlated with shocks to the regressand (returns or dividend-growth). If the shocks are correlated, then the finite-sample distribution is skewed. In our setting, shocks to returns are highly negatively correlated with shocks to dividend yield, which generates the skewness observed in $\widehat{\beta}_r$. In contrast, shocks to dividend-growth are close to being uncorrelated, which is why the distribution of $\widehat{\beta}_d$ does not exhibit skewness.

- For the most part power is much higher in the asymptotic case than in the finite-sample case. This is as expected: finite-sample variability results in high critical values, implying higher probability of acceptance of the null when data are generated under the alternative. However, the long-horizon tests are an exception: power is about the same in the asymptotic and finite-sample cases. This may reflect the fact, shown in Figure 4, that skewness is much lower for long-horizon returns than short-horizon returns.

- As expected from the Neyman-Pearson lemma, the likelihood-ratio tests have greater power than any of the other tests, both asymptotically and in finite samples.

- The asymptotic dividend tests have power that is only slightly lower than the likelihood ratio tests. This results from the fact that the critical values of the two tests are very close, so the two tests have almost exactly the same rejection regions. Similarly, the finite-sample dividend growth forecastability tests have about the same power as the likelihood ratio tests (except in the case of the 1% test), again because the critical lines of the two tests are close. This is illustrated in Figure 5.

- As noted in Section 2, asymptotically the long-horizon test has higher power under the 5% and 10% tests, but lower power under the 1% test, compared to the returns test. The fact that these differences are not pronounced corresponds to the analysis of Boudoukh et al. (2006), also based on asymptotic distributions, that the two test statistics are highly correlated due to the high persistence of dividend yields. Proposition 3 helps us understand this relationship between size and power for the two tests. We have that the threshold $\underline{\eta}$ equals -0.39. Because $\eta_{r,dp} = -0.70$ in the benchmark parameterization, we have that $\eta_{r,dp} < \underline{\eta}$. As Proposition 3 showed, the asymptotic power of the long-horizon test depends on whether the critical value of the β_r test is greater than or less than $\beta_{r,ALT} = 0.097$. The critical value of the 10% and 5% β_r tests are 0.074 and 0.095. Therefore the long-horizon test has higher power than the β_r test. The critical value of the 1% β_r test is $0.135 > 0.097$, and therefore the 1% β_r has higher asymptotic power.

- The finite-sample long-horizon test has much higher power than the finite-sample short-horizon returns test. Cochrane reported the same result. Again, this appears to reflect the fact that the long-run coefficient has lower skewness than the short-run returns coefficient (see Figure 4).

The power differences between the asymptotic test of return forecastability and the corresponding long-horizon tests are small, so we conclude that the two tests have about the same power. This finding connects with the demonstration of Boudoukh et al. (2006) that the two tests are very close to being equivalent due to the high correlation of short-horizon and long-horizon test statistics when the explanatory variable is highly autocorrelated. However, in the finite-sample case the long-horizon tests have considerably higher power.

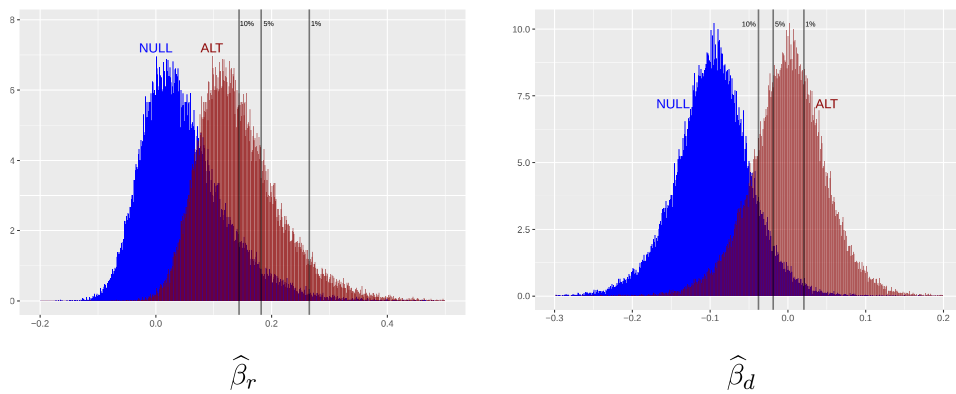

Figure 3: Simulated distributions of $\widehat{\beta}_r$ and $\widehat{\beta}_d$. The distribution under the null is shown in blue, while the distribution under the alternative is in red. The vertical black lines indicate 10%, 5% and 1% critical values.

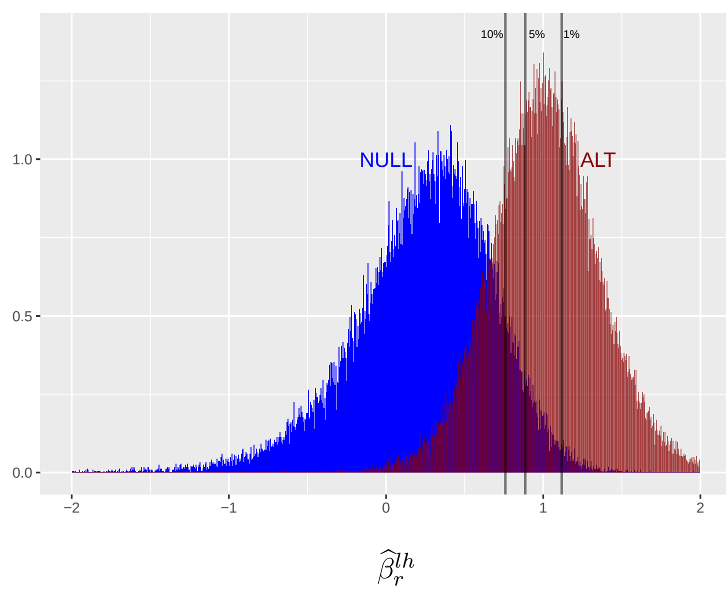

Figure 4: Simulated distribution of $\widehat{\beta}_r^{lh}$. The distribution under the null is shown in blue, while the distribution under the alternative is in red. The vertical black lines indicate 10%, 5% and 1% critical values.

	Asymptotic				Finite Sample			
Size	$\widehat{\beta}_r$	$\widehat{\beta}_d$	$\widehat{\beta}_r^{lh}$	LR test	$\widehat{\beta}_r$	$\widehat{\beta}_d$	$\widehat{\beta}_r^{lh}$	LR test
1%	0.378	0.686	0.326	0.689	0.065	0.328	0.371	0.437
5%	0.641	0.875	0.690	0.877	0.258	0.675	0.647	0.693
10%	0.766	0.935	0.841	0.936	0.444	0.805	0.775	0.807

Table 1: A comparison of statistical power of various asset pricing tests.

4 Tests Involving Price Volatility

It follows from (3) that the dividend yield (the reciprocal of which is the level of stock prices divided by dividends) is generated by the same equation with the same parameter values under the null and the alternative. This means that the assumed model and test specification imply that price volatility is the same whether or not markets are efficient. Thus the present setting can shed no light on the question of whether volatility tests have greater power than return forecastability tests in more general settings, as discussed in the introduction.

One way to alter the test so as to render it relevant to the size-power comparison of volatility and return forecastability tests would be to assume that the intervention setting β_r to the forecastable value is offset by an alteration in φ rather than β_d. Doing so implies that an efficiency test based on yield forecastability would replace the test based on dividend growth forecastability. However, there exists a critical problem with this proposed modification: a value setting β_r equal to zero is consistent with (5) only if φ is set equal to a value greater than 1, so that dividend yields are nonstationary. There is some empirical evidence for nonstationarity of dividend yields—see Craine (1993), and Welch and Goyal (2007), for example—but without stationarity the log-linearization underlying our size-power calculations is inaccurate. Thus the present framework does not apply.

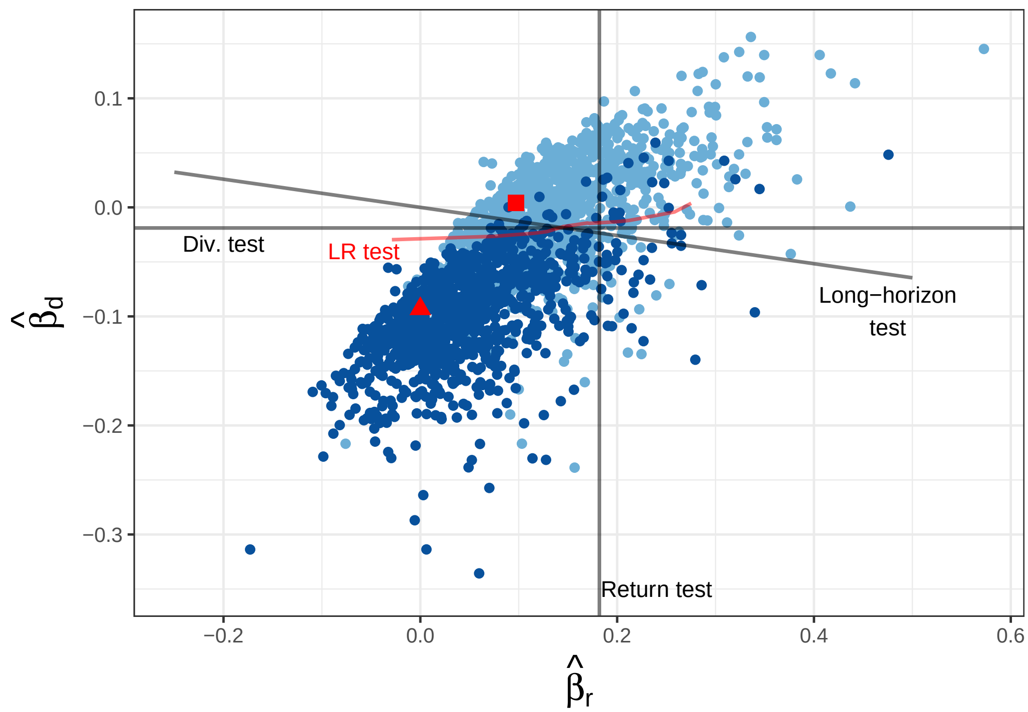

Figure 5: Scatter showing 1000 pairs of $\widehat{\beta}_d$ and $\widehat{\beta}_r$ with critical regions for the returns test, dividend growth test, long-horizon test and likelihood-ratio test. The critical regions were computed using all the 50,000 draws. The red triangle indicates the null hypothesis ($b_r = 0$). The red square indicates the alternative hypothesis ($b_r = 0.097$).

5 Conclusion

Cochrane concluded that tests based on dividend growth forecastability provide stronger evidence against market efficiency than those based on return forecastability. By this he meant that the test statistic for the dividend growth parameter was farther out on the tail of its distribution under market efficiency than that of returns. Our approach, in contrast, involves determining whether tests based on the dividend growth parameter are more likely than tests based on returns forecastability to detect departures from market efficiency if they exist. Determining this consists of ascertaining how the size-power tradeoffs of the two tests differ, and why. Our finding is that the size-power tradeoff favors the dividend growth test. Our most important contribution here is to provide a definitive explanation: dividend growth tests are more powerful asymptotically because returns have higher volatility than dividend growth. Our conclusion is the counterpart in our framework of Cochrane's conclusion that dividend growth tests provide stronger evidence than returns tests against market efficiency. To the extent that our conclusions correspond to those of Cochrane, our findings can be interpreted as supporting his.

A large literature is devoted to demonstrating that the appearance of return forecastability can occur even under the null due to a variety of econometric issues related to sampling variability. These issues apply as much to our simulations as to econometric tests based on real-world data. That being so, they are taken into account in simulations in determining critical values for market efficiency tests. It is, in fact, likely that the low power of finite-sample return forecastability tests reflects the fact that econometric biases result in high critical values for the return forecastability parameter. It appears that similar problems with the dividend-growth forecastability tests, if they exist at all, are less severe. However, our asymptotic tests, which by definition are not subject to the finite-sample issues under discussion, imply that the higher power of dividend growth forecastability tests cannot be attributed entirely to these issues.

We believe that our analysis gives strong support to Cochrane's contention that regressions of future returns on currently-observable variables do not provide the strongest evidence for return predictability. Rather, as Cochrane asserted, the best evidence comes from comparison of the behavior of dividends and stock prices: the volatility of stock prices dramatically exceeds the volatility level consistent with dividend behavior under efficient markets. This was exactly the conclusion reported in the variance bounds tests. However, there is a critical difference: the variance bounds tests took dividend volatility to be the determinant of stock price volatility, whereas in Cochrane's paper and here the relevant determinant is dividend growth predictability. This difference, if anything, greatly strengthens the evidence from the variance bounds literature for excessive stock price volatility relative to the efficient markets prediction: dividend growth forecastability is essentially zero, implying that price volatility should also be essentially zero under efficient markets.

References

Boudoukh, Jacob, Matthew Richardson, and Robert F Whitelaw (2006) 'The myth of long-horizon predictability.' *The Review of Financial Studies* 21(4), 1577–1605

Campbell, John Y (2001) 'Why long horizons? a study of power against persistent alternatives.' *Journal of Empirical Finance* 8(5), 459–491

Campbell, John Y, and Robert J Shiller (1988) 'The dividend-price ratio and expectations of future dividends and discount factors.' *The Review of Financial Studies* 1(3), 195–228

Cochrane, John H (2008) 'The dog that did not bark: A defense of return predictability.' *Review of Financial Studies* 21(4), 1533–1575

Craine, Roger (1993) 'Rational bubbles: A test.' *Journal of Economic Dynamics and Control* 17(5-6), 829–846

Fama, Eugene F (1970) 'Efficient capital markets: A review of theory and empirical work.' *Journal of Finance* 25(2), 383–417

Fama, Eugene F, and Kenneth R French (1988) 'Dividend yields and expected stock returns.' *Journal of Financial Economics* 22(1), 3–25

Hamilton, James D. (1994) *Time Series Analysis* (Princeton University Press)

LeRoy, Stephen F, and Douglas G Steigerwald (1995) 'Volatility.' *Handbook of Finance* 9, 411–433

Samuelson, Paul A (1965) 'Proof that properly anticipated prices fluctuate randomly.' *Industrial Management Review* 6(2), 41–49

Shiller, Robert J (1981) 'Do stock prices move too much to be justified by subsequent changes in dividends?' *American Economic Review* 71(3), 421–436

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

Welch, Ivo, and Amit Goyal (2007) 'A comprehensive look at the empirical performance of equity premium prediction.' *The Review of Financial Studies* 21(4), 1455–1508

A Appendix: Asymptotic Distribution of Estimators

In this section we derive the asymptotic distribution of $\widehat{b}_r$, $\widehat{b}_d$ and $\widehat{\phi}$. We do so using Proposition 11.1 in Hamilton (1994). For completeness, we restate the proposition below.

Proposition 4 *Let*

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\theta}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\theta}_2 \boldsymbol{y}_{t-2} + \cdots + \boldsymbol{\theta}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t \tag{12}$$

where $\boldsymbol{\varepsilon}_t$ is independent and identically distributed with mean $\mathbf{0}$, variance $\boldsymbol{\Omega}$, and $\mathbb{E}(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{lt}\varepsilon_{mt}) < \infty$ for all i, j, l, m and where the roots of

$$\left| \boldsymbol{I}_n - \boldsymbol{\theta}_1 z - \boldsymbol{\theta}_2 z^2 - \cdots - \boldsymbol{\theta}_p z^p = 0 \right|$$

lie outside the unit circle. Let $k \equiv np + 1$ and let $\boldsymbol{x}_t^{'}$ be the $1 \times k$ vector

$$\boldsymbol{x}_t^{'} \equiv [1 \quad \boldsymbol{y}_{t-1}^{'} \quad \boldsymbol{y}_{t-2}^{'} \quad \ldots \quad \boldsymbol{y}_{t-p}^{'}].$$

Let $\widehat{\boldsymbol{\pi}}_T = vec(\widehat{\boldsymbol{\Pi}}_T)$ denote the $(nk \times 1)$ vector of coefficients resulting from OLS regressions of each of the elements of $\boldsymbol{y}_t$ on $\boldsymbol{x}_t$ for a sample of size T:

$$\widehat{\boldsymbol{\pi}}_T = \begin{bmatrix} \widehat{\boldsymbol{\pi}}_{1,T} \\ \widehat{\boldsymbol{\pi}}_{2,T} \\ \vdots \\ \widehat{\boldsymbol{\pi}}_{n,T} \end{bmatrix}$$

where

$$\widehat{\boldsymbol{\pi}}_{i,T} = \left[\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t^{'} \right]^{-1} \left[\sum_{t=1}^{T} \boldsymbol{x}_t y_{it} \right];$$

and let $\boldsymbol{\pi}$ *denote the* $(nk \times 1)$ *vector of corresponding population coefficients. Finally, let*

$$\widehat{\boldsymbol{\Omega}}_T = \frac{1}{T} \sum_{t=1}^{T} \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t^{'},$$

where

$$\begin{aligned} \widehat{\boldsymbol{\varepsilon}}_t^{'} &= [\widehat{\varepsilon}_{1t} \quad \widehat{\varepsilon}_{2t} \quad \ldots \quad \widehat{\varepsilon}_{nt}] \\ \widehat{\varepsilon}_{it} &= y_{it} - \boldsymbol{x}_t^{'} \widehat{\boldsymbol{\pi}}_{i,T}. \end{aligned}$$

Then

1. $(1/T) \sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t^{'} \xrightarrow{p} \boldsymbol{Q}$ *where* $\boldsymbol{Q} = \mathbb{E}(\boldsymbol{x}_t \boldsymbol{x}_t^{'})$*;*
2. $\widehat{\boldsymbol{\pi}}_T \xrightarrow{p} \boldsymbol{\pi}$*;*
3. $\widehat{\boldsymbol{\Omega}}_T \xrightarrow{p} \boldsymbol{\Omega}$*;*
4. $\sqrt{T}(\widehat{\boldsymbol{\pi}}_T - \boldsymbol{\pi}) \xrightarrow{L} N(\mathbf{0}, (\boldsymbol{\Omega} \otimes \boldsymbol{Q}^{-1}))$*, where* $\otimes$ *denotes the Kronecker product.*

We verify that the conditions required for Proposition 4 hold in the model presented here. We have

$$\boldsymbol{y}_t' = [r_t \quad \Delta d_t \quad dp_t]$$

It follows that $n = 3$. The error terms are drawn from a multivariate normal distribution with variance-covariance matrix (11). The vector of coefficients $\boldsymbol{\theta}_1$ is given by

$$\boldsymbol{\theta}_1 = \begin{bmatrix} 0 & 0 & b_r \\ 0 & 0 & b_d \\ 0 & 0 & \phi \end{bmatrix}$$

The first two columns have zeros because the regressand is dp_{t-1}. The coefficients on all lags greater than one are zero.

The condition $\mathbb{E}(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{lt}\varepsilon_{mt}) < \infty$ for all i, j, l, m is satisfied because the error terms are normally distributed. The condition that the roots lie outside the unit circle is satisfied because b_r, b_d and ϕ are less than unity.

Because we are not regressing on past values of r_t and Δd_t, we redefine k and x_t accordingly. We have $k = 1 + 1 = 2$. The vector $\boldsymbol{x}_t'$ is given by

$$\boldsymbol{x}_t' = [1 \quad dp_t]$$

The matrix $\mathbf{\Pi}_T$ is given by

$$\mathbf{\Pi}_T = \begin{bmatrix} a_r & b_r \\ a_d & b_d \\ a_{dp} & \phi \end{bmatrix}$$

The vector $\boldsymbol{\pi}_T \equiv vec(\mathbf{\Pi}_T)$ follows.

We can now calculate $Q \equiv \mathbb{E}(x_t x_t')$. We have

$$\boldsymbol{Q} = \mathbb{E}(\boldsymbol{x}_t \boldsymbol{x}_t') = \begin{bmatrix} 1 & 0 \\ 0 & \sigma_{dp}^2/(1-\phi^2) \end{bmatrix}$$

where, WLOG, we have assumed dp_t is mean-zero. Because $\boldsymbol{Q}$ is a diagonal matrix, it is easy to invert. We have

$$\boldsymbol{Q}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & (1-\phi^2)/\sigma_{dp}^2 \end{bmatrix}$$

We can calculate $\mathbf{\Omega} = \mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$. We have

$$\mathbf{\Omega} = \begin{bmatrix} \sigma_r^2 & \sigma_{r,d} & \sigma_{r,dp} \\ \sigma_{r,d} & \sigma_d^2 & \sigma_{d,dp} \\ \sigma_{r,dp} & \sigma_{d,dp} & \sigma_{dp}^2 \end{bmatrix}$$

The Kronecker product $\boldsymbol{\Omega} \otimes \boldsymbol{Q}^{-1}$ follows. In particular, we note the asymptotic variance of $\sqrt{T}(\widehat{b}_r - b_r)$ is given by $\sigma_r^2/\sigma_{dp}^2(1-\phi^2)$. It follows that

$$avar(\widehat{b}_r) = \frac{1}{T}\frac{\sigma_r^2(1-\phi^2)}{\sigma_{dp}^2}$$

Similarly,

$$avar(\widehat{b}_d) = \frac{1}{T}\frac{\sigma_d^2(1-\phi^2)}{\sigma_{dp}^2}$$

$$avar(\widehat{\phi}) = \frac{1}{T}(1-\phi^2)$$

It should be noted that the expressions for the asymptotic variance are identical to the standard OLS expressions.

A.1 Asymptotic variance of $\widehat{\beta}_r^{lh}$

We find the asymptotic variance of $\widehat{\beta}_r^{lh}$ using the multivariate delta method. Let $h(\widehat{\beta}_r, \widehat{\phi}) \equiv \widehat{\beta}_r^{lh} = \widehat{\beta}_r/(1-\rho\widehat{\phi})$. The delta method uses the Taylor approximation to derive the asymptotic distribution of $h(\widehat{\beta}_r, \widehat{\phi})$ from knowledge of the asymptotic joint distribution of $\widehat{\beta}_r$ and $\widehat{\phi}$. We have

$$\sqrt{T}\begin{bmatrix}\widehat{\beta}_r - \beta_r \\ \widehat{\phi} - \phi\end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \Sigma\right) \tag{13}$$

where Σ denotes the 2×2 variance-covariance matrix. The delta method implies

$$\sqrt{T}[h(\widehat{\beta}_r, \widehat{\phi}) - h(\beta_r, \phi)] \xrightarrow{d} \mathcal{N}\left(0, [\nabla h(\beta_r, \phi)]' \, \Sigma \, \nabla h(\beta_r, \phi)\right) \tag{14}$$

where $\nabla h(\beta_r, \phi)$ denotes the gradient of h evaluated at the mean (β_r, ϕ), and prime denotes transpose. We have

$$\nabla h(\beta_r, \phi) = \begin{bmatrix} \frac{1}{1-\rho\phi} \\ \frac{\rho\beta_r}{(1-\rho\phi)^2} \end{bmatrix} \tag{15}$$

It follows that

$$\begin{aligned} &[\nabla h(\beta_r, \phi)]' \, \Sigma \, \nabla h(\beta_r, \phi) \\ &= \begin{bmatrix} \frac{1}{1-\rho\phi} & \frac{\rho\beta_r}{(1-\rho\phi)^2} \end{bmatrix} \begin{bmatrix} \gamma_r^2 & \gamma_{r,dp} \\ \gamma_{r,dp} & \gamma_{dp}^2 \end{bmatrix} \begin{bmatrix} \frac{1}{1-\rho\phi} \\ \frac{\rho\beta_r}{(1-\rho\phi)^2} \end{bmatrix} \\ &= \frac{1-\phi^2}{(1-\rho\phi)^2} \left(\frac{\sigma_r^2}{\sigma_{dp}^2} + \frac{2\rho\beta_r}{1-\rho\phi} \frac{\sigma_{r,dp}}{\sigma_{dp}^2} + \frac{\rho^2\beta_r^2}{(1-\rho\phi)^2} \right) \end{aligned} \tag{16}$$

The asymptotic variance of the long-horizon estimator depends on the value of β_r. Therefore it is different under the null and the alternative hypotheses.

Weighted cumulative return regressions. The delta method can also be used to derive the asymptotic distribution of cumulative return regressions that are commonly employed in the literature.

Geometrically declining weights.—Consider the regression in which future returns are weighted using ρ, with the weights declining geometrically. We have

$$\sum_{j=0}^{\tau} \rho^j r_{t+1+j} = a_r^{(\tau)} + \beta_r^{(\tau)}(d_t - p_t) + \varepsilon_{t+1,t+1+\tau}^r \tag{17}$$

Let $g(\widehat{\beta}_r, \widehat{\varphi}) \equiv \widehat{\beta}_r^{(\tau)}$. We have that

$$g(\widehat{\beta}_r, \widehat{\varphi}) = \widehat{\beta}_r(1 + \widehat{\varphi} + \widehat{\varphi}^2 + \cdots + \widehat{\varphi}^\tau) = \widehat{\beta}_r \left(\frac{1-(\rho\widehat{\varphi})^{\tau+1}}{1-\rho\widehat{\varphi}} \right)$$

We compute $\nabla g(\widehat{\beta}_r, \widehat{\varphi})$, and evaluate the resulting gradient at the asymptotic means of the estimates. We have

$$\nabla g(\beta_r, \phi) = \begin{bmatrix} \frac{1-(\rho\varphi)^{\tau+1}}{1-\rho\phi} \\ \frac{\rho\beta_r}{(1-\rho\phi)^2}\left(1-(\rho\varphi)^{\tau}(1+\tau(1-\rho\varphi))\right) \end{bmatrix} \tag{18}$$

It follows that

$$\begin{aligned} &[\nabla g(\beta_r, \varphi)]' \, \Sigma \, \nabla g(\beta_r, \varphi) \\ &\quad = \frac{1-\varphi^2}{(1-\rho\varphi)^2}\Bigg[\left(1-(\rho\varphi)^{\tau+1}\right)\left(\frac{\sigma_r^2}{\sigma_{dp}^2} + \frac{2\sigma_{r,dp}}{\sigma_{dp}^2}\frac{\rho\beta_r}{1-\rho\varphi}\big[1-(\rho\varphi)^{\tau}(1+(1-\rho\varphi)\tau)\big]\right) \\ &\qquad + \frac{\rho^2\beta_r^2}{(1-\rho\varphi)^2}\big[1-(\rho\varphi)^{\tau}(1+(1-\rho\varphi)\tau)\big]^2\Bigg] \end{aligned} \tag{19}$$

The long-horizon regression corresponds the limiting case $\tau \to \infty$. In that case, the variance expression in (19) reduces to (16).

Equal weights.— We can also consider regressions with equal weights on future returns

$$\sum_{j=0}^{\tau} r_{t+1+j} = \widetilde{a}_r^{(\tau)} + \widetilde{\beta}_r^{(\tau)}(d_t - p_t) + \widetilde{\varepsilon}_{t+1,t+1+\tau}^{r} \tag{20}$$

In that case, the asymptotic variance of the $\sqrt{T}(\widehat{\beta}_r^{(\tau)} - \beta_r^{\tau})$ is given by,

$$\begin{aligned} &\frac{1-\varphi^2}{(1-\varphi)^2}\Bigg(\left(1-\varphi^{\tau+1}\right)\left(\frac{\sigma_r^2}{\sigma_{dp}^2} + \frac{2\sigma_{r,dp}}{\sigma_{dp}^2}\frac{\beta_r}{1-\varphi}[1-\varphi^{\tau}(1+(1-\varphi)\tau)]\right) \\ &\qquad + \frac{\beta_r^2}{(1-\varphi)^2}\big[1-\varphi^{\tau}(1+(1-\varphi)\tau)\big]^2\Bigg) \end{aligned} \tag{21}$$

B Appendix: Derivation of Power

Consider two estimators $\widehat{\beta}_1$ and $\widehat{\beta}_2$ that are asymptotically distributed as bivariate normal. Suppose that they satisfy

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} \sim BN\left(\begin{bmatrix} \mu_{1,N} \\ \mu_{2,N} \end{bmatrix}, \begin{bmatrix} \gamma_{1,N}^2 & \gamma_{12,N} \\ \gamma_{12,N} & \gamma_{2,N}^2 \end{bmatrix}\right). \tag{22}$$

under the null hypothesis, and

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} \sim BN\left(\begin{bmatrix} \mu_{1,A} \\ \mu_{2,A} \end{bmatrix}, \begin{bmatrix} \gamma_{1,A}^2 & \gamma_{12,A} \\ \gamma_{12,A} & \gamma_{2,A}^2 \end{bmatrix}\right). \tag{23}$$

under the alternative. Define $\overline{\beta}_i$ as the critical value of $\widehat{\beta}_i$ for a one-tailed test of the null hypothesis with probability of Type-I error 0.05. Under the null hypothesis we have

$$\mathcal{N}\left(\frac{\overline{\beta}_i - \mu_{i,N}}{\gamma_{i,N}}\right) = 0.95, \quad \text{for } i = 1,2 \tag{24}$$

where $\mathcal{N}$ is the CDF of the standard univariate normal distribution. The power of test i equals the probability that $\widehat{\beta}_i \geq \overline{\beta}_i$ under the alternative, or

$$\begin{aligned} \text{power of } \widehat{\beta}_i \text{ test} &= 1 - \mathcal{N}\left(\frac{\overline{\beta}_i - \mu_{i,A}}{\gamma_{i,A}}\right) \\ &= 1 - \mathcal{N}\left(\mathcal{N}^{-1}(0.95)\frac{\gamma_{i,N}}{\gamma_{i,A}} + \frac{\mu_{i,N} - \mu_{i,A}}{\gamma_{i,A}}\right). \end{aligned} \tag{25}$$

Here $\mathcal{N}^{-1}$ is the inverse of $\mathcal{N}$ (note that for $\mu_{i,A} = \mu_{i,N}$ and $\gamma_{i,N} = \gamma_{1,A}$ we have power $= 0.05$, as expected). It follows that test 2 is more powerful than test 1 if and only if

$$\mathcal{N}^{-1}(0.95)\frac{\gamma_{2,N}}{\gamma_{2,A}} + \frac{\mu_{2,N} - \mu_{2,A}}{\gamma_{2,A}} \leq \mathcal{N}^{-1}(0.95)\frac{\gamma_{1,N}}{\gamma_{1,A}} + \frac{\mu_{1,N} - \mu_{1,A}}{\gamma_{1,A}} \tag{26}$$

We can use (26) to compare the power of various asset pricing tests. Consider the tests involving $\widehat{\beta}_r$ and $\widehat{\beta}_d$. Under the null hypothesis, the mean of $[\widehat{\beta}_r;\ \widehat{\beta}_d]$ is equal to $[0;\ \mu_{d,A} - \mu_{r,A}]$. Under the alternative hypothesis, the mean is $[\mu_{r,A};\ \mu_{d,A}]$. The variance-covariance matrix of the estimators $\widehat{\beta}_r$ and $\widehat{\beta}_d$ is the same under the null and the alternative, so we do not distinguish between the two cases and denote the estimator variances as γ_r^2 and γ_d^2. From (26), the $\widehat{\beta}_d$ test has greater power than the $\widehat{\beta}_r$ test if and only if $\gamma_d < \gamma_r$.

This result holds for any values of the parameters in (22) and (23); in the text we specified the parameter values to be those estimated by Cochrane. Under those parameter values, we have $\gamma_d < \gamma_r$. Therefore the power of the test based on $\widehat{\beta}_d$ is higher than that based on $\widehat{\beta}_r$.

Now consider tests involving $\widehat{\beta}_r$ and $\widehat{\beta}_r^{lh}$. Under the null hypothesis, the mean of $[\widehat{\beta}_r;\ \widehat{\beta}_r^{lh}]$ is equal to $[0; 0]$. Under the alternative hypothesis, the mean is $[\mu_{r,A};\ \mu_{r,A}^{lh}]$, where $\mu_{r,A}^{lh} = \mu_{r,A}/(1-\rho\phi)$. The variance of $\widehat{\beta}_r$ does not change as we go from the null to the alternative. However, the variance of $\widehat{\beta}_r^{lh}$ changes as we move from the null to alternative. We denote the variances of the long-horizon test under the null and the alternative as $\gamma_{lh,N}^2$ and $\gamma_{lh,A}^2$. From (25), we have that the long-horizon test has greater power than the $\widehat{\beta}_r$ test if and only if,

$$\mathcal{N}^{-1}(0.95)\left(\frac{\gamma_{lh,N}}{\gamma_{lh,A}} - 1\right) + \frac{0 - \mu_{r,A}^{lh}}{\gamma_{lh,A}} \leq \frac{0 - \mu_{r,A}}{\gamma_r} \tag{27}$$

This inequality holds under the benchmark parameter values, so the long-horizon test is more powerful than the $\widehat{\beta}_r$ test.