The Dog That Did Not Bark: A Defense of Return Predictability

John H. Cochrane
University of Chicago GSB and NBER

If returns are not predictable, dividend growth must be predictable, to generate the observed variation in divided yields. I find that the *absence* of dividend growth predictability gives stronger evidence than does the *presence* of return predictability. Long-horizon return forecasts give the same strong evidence. These tests exploit the negative correlation of return forecasts with dividend-yield autocorrelation across samples, together with sensible upper bounds on dividend-yield autocorrelation, to deliver more powerful statistics. I reconcile my findings with the literature that finds poor power in long-horizon return forecasts, and with the literature that notes the poor out-of-sample R^2 of return-forecasting regressions. (*JEL* G12, G14, C22)

Are stock returns predictable? Table 1 presents regressions of the real and excess value-weighted stock return on its dividend-price ratio, in annual data. In contrast to the simple random walk view, stock returns do seem predictable. Similar or stronger forecasts result from many variations of the variables and data sets.

Economic significance

The estimates in Table 1 have very large *economic* significance. The standard deviation of expected returns in the last column of Table 1 is about five percentage points, almost as large as the 7.7% level of the equity premium in this sample. The equity premium apparently varies over time by as much as its unconditional mean. The 4–7% R^2 do not look that impressive, but the R^2 rises with horizon, reaching values between 30 and 60%, depending on time period and estimation details, as emphasized by Fama and French (1988). The slope coefficient of over three in the top two rows means that when dividend yields rise one percentage point, prices *rise* another two percentage points on average, rather than declining one percentage point to offset the extra dividends and render returns unpredictable. Finally, the regressions of Table 1 imply that *all* variation in market price-dividend ratios corresponds to changes in expected excess

I acknowledge research support from CRSP and from a NSF grant administered by the NBER. I thank Alan Bester, John Campbell, John Heaton, Lars Hansen, Anil Kashyap, Sydney Ludvigson, Lubos Pastor, Ivo Welch, and an anonymous referee for very helpful comments. Address correspondence to John H. Cochrane, Graduate School of Business, 5807 S. Woodlawn, Chicago IL 60637, 773 702 3059, or e-mail: john.cochrane@chicagogsb.edu.

doi:10.1093/rfs/hhm046 Advance Access publication September 22, 2007

Table 1
Forecasting regressions

Regression	b	t	R^2(%)	$\sigma(bx)$(%)
$R_{t+1} = a + b(D_t/P_t) + \varepsilon_{t+1}$	3.39	2.28	5.8	4.9
$R_{t+1} - R_t^f = a + b(D_t/P_t) + \varepsilon_{t+1}$	3.83	2.61	7.4	5.6
$D_{t+1}/D_t = a + b(D_t/P_t) + \varepsilon_{t+1}$	0.07	0.06	0.0001	0.001
$r_{t+1} = a_r + b_r(d_t - p_t) + \varepsilon_{t+1}^r$	0.097	1.92	4.0	4.0
$\Delta d_{t+1} = a_d + b_d(d_t - p_t) + \varepsilon_{t+1}^{dp}$	0.008	0.18	0.00	0.003

R_{t+1} is the real return, deflated by the CPI, D_{t+1}/D_t is real dividend growth, and D_t/P_t is the dividend-price ratio of the CRSP value-weighted portfolio. R_{t+1}^f is the real return on 3-month Treasury-Bills. Small letters are logs of corresponding capital letters. Annual data, 1926–2004. $\sigma(bx)$ gives the standard deviation of the fitted value of the regression.

returns—risk premiums—and *none* corresponds to news about future dividend growth. I present this calculation below.

Statistical significance

The *statistical* significance of the return forecast in Table 1 is marginal, however, with a t-statistic only a little above two. And the ink was hardly dry on the first studies[1] to run regressions like those of Table 1 before a large literature sprang up examining their econometric properties and questioning that statistical significance. The right-hand variable (dividend yield) is very persistent, and return shocks are negatively correlated with dividend-yield shocks. As a result, the return-forecast regression inherits the near-unit-root properties of the dividend yield. The coefficient is biased upward, and the t-statistic is biased toward rejection. Stambaugh (1986, 1999) derived the finite-sample distribution of the return-forecasting regression. In monthly regressions, Stambaugh found that in place of OLS p-values of 6% (1927–1996) and 2% (1952–1996), the correct p-values are 17 and 15%. The regressions are far from statistically significant at conventional levels.[2]

Does this evidence mean return forecastability is dead? No, because there are more powerful tests, and these tests give stronger evidence against the null.

First, we can examine dividend growth. In the regressions of Table 1, dividend growth is clearly not forecastable at all. In fact, the small point

[1] Rozeff (1984), Shiller (1984), Keim and Stambaugh (1986), Campbell and Shiller (1988), and Fama and French (1988).

[2] Precursors include Goetzmann and Jorion (1993) and Nelson and Kim (1993) who found the distribution of the return-forecasting coefficient by simulation and also did not reject the null. Mankiw and Shapiro (1986) show the bias point by simulation, though not applied to the dividend-yield/return case. Additional contributions include Kothari and Shanken (1997), Paye and Timmermann (2003), Torous, Valkanov, and Yan (2004), Campbell and Yogo (2006), and Ang and Bekaert (2007).

estimates have the wrong sign—a high dividend yield means a low price, which should signal lower, not higher, future dividend growth.

If *both* returns and dividend growth are unforecastable, then present value logic implies that the price/dividend ratio is constant, which it obviously is not. Alternatively, in the language of cointegration, since the dividend yield is stationary, *one* of dividend growth or price growth must be forecastable to bring the dividend yield back following a shock. We cannot just ask, "Are returns forecastable?" and "Is dividend growth forecastable?" We must ask, "*Which* of dividend growth or returns is forecastable?" (Or really, "How much of each?") A null hypothesis in which returns are not forecastable must also specify that dividend growth *is* forecastable, and the statistical evaluation of that null must also confront the *lack* of dividend-growth forecastability in the data.

I set up such a null, and I evaluate the *joint* distribution of return and dividend-growth forecasting coefficients. I confirm that the return-forecasting coefficient, taken alone, is not significant: Under the unforecastable-return null, we see return forecast coefficients as large or larger than those in the data about 20% of the time and a t-statistic as large as that seen in the data about 10% of the time. However, I find that the *absence* of dividend–growth forecastability offers much more significant evidence against the null. The best overall number is a 1–2% probability value (last row of Table 5)—dividend growth *fails* to be forecastable in only 1–2% of the samples generated under the null. The important evidence, as in Sherlock Holmes's famous case, is the dog that does not bark.[3]

Second, we can examine the long-horizon return forecast implied by one-year regressions. It turns out to be most convenient to look at $b_r^{lr} \equiv b_r/(1-\rho\phi)$ where ϕ is the dividend-yield autocorrelation, $\rho \approx 0.96$ is a constant related to the typical level of the dividend yield, and b_r is the return-forecast coefficient as defined in Table 1. The "long horizon" label applies because b_r^{lr} is the implied coefficient of weighted long-horizon returns $\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$ on dividend yields. The null hypothesis produces a long-horizon return regression coefficient b_r^{lr} larger than its sample value only about 1–2% of the time, again delivering much stronger evidence against the null than the one-period return coefficient b_r.

Why are these tests more powerful? They exploit a feature of the data and a feature of the null hypothesis that the conventional b_r test ignores.

The feature of the *data* is that return shocks ε^r and dividend-yield shocks ε^{dp} are strongly and negatively correlated. Briefly, a price rise raises returns and lowers dividend yields. This correlation means that regression

[3] Inspector Gregory: "Is there any other point to which you would wish to draw my attention?"
Holmes: "To the curious incident of the dog in the night-time."
"The dog did nothing in the night time."
"That was the curious incident."
(From "The Adventure of Silver Blaze" by Arthur Conan Doyle.)

estimates b_r and ϕ are strongly and negatively correlated across samples. A large long-run coefficient $b_r^{lr} = b_r/(1 - \rho\phi)$ requires both a large b_r *and* a large ϕ, so that the coefficients can build with horizon as they do in our data. But since b_r and ϕ are negatively correlated, samples with unusually large b_r tend to come with unusually low ϕ, so it is much harder for the null to generate large long-run coefficients. The dividend-growth test works the same way.

The feature of the *null* is that we know something about the dividend-yield autocorrelation ϕ. The Wald test on b_r uses no information about other parameters. It is the appropriate test of the null $\{b_r = 0, \phi = \text{anything}\}$. But we know ϕ cannot be too big. If $\phi > 1/\rho \approx 1.04$, the present value relation explodes and the price-dividend ratio is infinite, which it also is obviously not. If $\phi \geq 1.0$, the dividend yield has a unit or larger root, meaning that its variance explodes with horizon. Economics, statistics, and common sense mean that if our null is to describe a coherent world, it should contain *some* upper bound on ϕ as well as $b_r = 0$, something like $\left\{b_r = 0, \|\phi\| < \overline{\phi}\right\}$. A good test uses information on *both* $\hat{b}_r$ and $\hat{\phi}$ to evaluate such a null, drawing regions in $\{b_r, \phi\}$ space around the null $\left\{b_r = 0, \|\phi\| < \overline{\phi}\right\}$, and exploiting the fact that under the null $\hat{b}_r$ should not be too big *and* $\hat{\phi}$ should not be too big. The test regions in $\{b_r, \phi\}$ described by the long-run return coefficient $b_r^{lr} = b_r/(1 - \rho\phi)$ and by the dividend-growth coefficient slope downward in $\{b_r, \phi\}$ space in just this way.

The long-run return forecasting coefficients also describe a more *economically* interesting test region. In economic terms, we want our test region to contain draws "more extreme" than the observed sample. Many of the draws that produce a one-period return forecast coefficient b_r larger than the sample value also have forecastable dividend growth, and dividend-yield variation is partially due to changing dividend-growth forecasts—their dogs do bark; volatility tests are in them a half-success rather than the total failure they are in our data. It makes great economic sense to consider such draws "closer to the null" than our sample, even though the one-year return-forecast coefficient b_r is greater than it is in our sample. This is how the long-run coefficients count such events, resulting in small probability values for events that really are, by this measure, "more extreme" than our data. The long-run return and dividend-growth forecast coefficients are also linked by an identity $b_{lr}^r - b_d^{lr} = 1$, so the test is exactly the same whether one focuses on returns or dividend growth, removing the ambiguity in short-horizon coefficients.

Powerful long-horizon regressions?

The success of long-horizon regression tests leads us to another econometric controversy. Fama and French (1988) found that return-forecast t-statistics rise with horizon, suggesting that long-horizon return

regressions give greater statistical evidence for return forecastability. This finding has also been subject to great statistical scrutiny. Much of this literature concludes that long-horizon estimates do not, in fact, have better statistical power than one-period regressions. Boudoukh, Richardson, and Whitelaw (2006) are the most recent examples and they survey the literature. Their Table 5, top row, gives probability values for return forecasts from dividend-price ratios at 1 to 5 year horizons, based on simulations similar to mine. They report 15, 14, 13, 12 and 17% values. In short, they find no advantage to long-horizon regressions.

How do I find such large power advantages for long-horizon regression coefficients? The main answer is that typical long-horizon estimates, going out to 5-year or even 10-year horizons, do not weight ϕ enough to see the power benefits. For example, the 2 year return coefficient is $b_r^{(2)} = b_r(1+\phi)$. Since $b_r \approx 0.1$, this coefficient weights variation in ϕ by 0.1 times as much as it weights variation in b_r. But ϕ and b_r estimates vary about one-for-one across samples, so a powerful test needs to construct a region in $\{b_r, \phi\}$ space with about that slope, which the implied infinite-horizon coefficient $b_r^{lr} = b_r/(1-\rho\phi)$ does.

This finding does not imply that one should construct 30-year returns and regress them on dividend yields or other forecasting variables. I calculate "long-horizon" coefficients implied from the one-year regression coefficients, and they are here just a convenient way of combining those one-year regression coefficients b_r, ϕ to generate a test region in $\{b_r, \phi\}$ space that has good power and strong economic intuition.

We therefore obtain a nice resolution of this long-running statistical controversy. I reproduce results such as Boudoukh, Richardson, and Whitelaw's (2006), that direct regressions at 1-year to 5-year horizons have little or no power advantages over 1-year regressions, but I also agree with results such as Campbell (2001) and Valkanov (2003), that there are strong power advantages to long-horizon regressions, advantages that are maximized at very long horizons and, to some extent, by calculating long-horizon statistics implied by VARs rather than direct estimates.

Out-of-sample R^2

Goyal and Welch (2003, 2005) found that return forecasts based on dividend yields and a number of other variables do not work out of sample. They compared forecasts of returns at time $t+1$ formed by estimating the regression using data up to time t, with forecasts that use the sample mean in the same period. They found that the sample mean produces a better out-of-sample prediction than do the return-forecasting regressions.

I confirm Goyal and Welch's observation that out-of-sample return forecasts are poor, but I show that this result is to be expected. Setting up a null in which *return* forecasts account for all dividend-yield volatility, I find

out-of-sample performance as bad or worse than that in the data 30–40% of the time. Thus, the Goyal–Welch calculations do not provide a statistical rejection of forecastable returns. Out-of-sample R^2 is not a *test*; it is not a statistic that somehow gives us better power to distinguish alternatives than conventional full-sample hypothesis tests. Instead, Goyal and Welch's findings are an important caution about the practical usefulness of return forecasts in forming aggressive real-time market-timing portfolios given the persistence of forecasting variables and the short span of available data.

Common misunderstandings

First, one should not conclude that "returns are not forecastable, but we can somehow infer their forecastability from dividend evidence." The issue is *hypothesis tests*, not *point estimates*. The point estimates are, as in Table 1, that returns are *very* forecastable, where the adjective "very" means by any economic metric. The point estimate (possibly with a bias adjustment) remains anyone's best guess. Hypothesis tests ask, "What is the probability that we see something as large as Table 1 by chance, if returns are truly not forecastable?" Stambaugh (1999) answer is about 15%. Even 15% is still not 50 or 90%, so zero return forecastability is still not a very likely summary of the data. "Failing to reject the null" does not mean that we wholeheartedly *accept* the i.i.d. worldview. Lots of nulls cannot be rejected.

In this context, I point out that the unforecastable-return null has other implications that one can also test—the implication that we should see a large dividend-growth forecast, a low dividend-yield autocorrelation, and a small "long-run" return forecast. Looking at these other statistics, we can say that there is in fact less than a 5% chance that our data or something more extreme is generated by a coherent world with unpredictable returns. But this evidence, like the return-based evidence, also does *nothing* to change the point estimate.

Second, this paper is about the *statistics* of return forecastability, not "how best to forecast returns." Simple dividend-yield regressions do not provide the strongest estimates or the best representation of return forecastability. If one really wants to forecast returns, additional variables are important, and one should pick variables and specifications that reflect repurchases, dividend smoothing, and possible changes in dividend payment behavior.

I use the simplest environment in order to make the statistical points most transparently. Better specifications can only increase the evidence for forecastable returns. For this reason, the point of this article is not to vary the specification until the magic 5% barrier is crossed. The point of this article is to see how different and more comprehensive statistical analysis yields different results, and in particular how tests based on

dividend growth and long-run regressions yield stronger evidence. Those points carry over to more complex and hence more powerful forecasting environments, and it is there that the real search for 5% values (or 1%, or the appropriate Bayesian evaluation) should occur.

1. Null Hypothesis

To keep the analysis simple, I consider a first-order VAR representation of log returns, log dividend yields, and log dividend growth,

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

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

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

Returns and dividend growth do not add much forecast power, nor do further lags of dividend yields. Of course, adding more variables can only make returns more forecastable.

The Campbell-Shiller (1988) linearization of the definition of a return[4] gives the approximate identity

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

where $\rho = PD/(1 + PD)$, PD is the price-dividend ratio about which one linearizes, and lowercase letters are demeaned logarithms of corresponding capital letters.

This identity applies to each data point, so it links the regression coefficients and errors of the VAR (1)–(3). First, projecting on $d_t - p_t$,

[4] Start with the identity

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{\left(1 + \frac{P_{t+1}}{D_{t+1}}\right)\frac{D_{t+1}}{D_t}}{\frac{P_t}{D_t}}$$

Loglinearizing,

$$\begin{aligned} r_{t+1} &= \log\left[1 + e^{(p_{t+1} - d_{t+1})}\right] + \Delta d_{t+1} - (p_t - d_t) \\ &\approx k + \frac{P/D}{1 + P/D}(p_{t+1} - d_{t+1}) + \Delta d_{t+1} - (p_t - d_t) \end{aligned}$$

where P/D is the point of linearization. Ignoring means, and defining $\rho = \frac{P/D}{1+P/D}$, we obtain Equation (4).

identity (4) implies that the regression coefficients obey the approximate identity

$$b_r = 1 - \rho\phi + b_d \tag{5}$$

Second, the identity (4) links the errors in (1)–(3) by

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

Thus, the three equations (1)–(3) are redundant. One can infer the data, coefficients, and error of any one equation from those of the other two.

The identity (5) shows clearly how we cannot form a null by taking $b_r = 0$ without changing the dividend-growth forecast b_d or the dividend-yield autocorrelation ϕ. In particular, as long as ϕ is nonexplosive, $\phi < 1/\rho \approx 1.04$, we cannot choose a null in which *both* dividend growth and returns are unforecastable—$b_r = 0$ and $b_d = 0$. To generate a coherent null with $b_r = 0$, we must assume a negative b_d, and then we must address the *absence* of this coefficient in the data.

By subtracting inflation from both sides, Equations (4)–(6) can apply to real returns and real dividend growth. Subtracting the risk-free rate from both sides, we can relate the excess log return $(r_{t+1} - r^f_t)$ on the left-hand side of Equation (4) to dividend growth less the interest rate $(\Delta d_{t+1} - r^f_t)$ on the right-hand side. One can either introduce an extra interest rate term or simply understand "dividend growth" to include both terms. I follow the latter convention in the excess return results below. One can form similar identities and decompositions with other variables. For example, starting with the price/earnings ratio, we form a similar identity that also includes the earnings/dividend ratio.

To form a null hypothesis, then, I start with estimates of Equations (1)–(3) formed from regressions of log real returns, log real dividend growth and the log dividend yield in annual Center for Research in Security Prices (CRSP) data, 1927–2004, displayed in Table 2. The coefficients are worth keeping in mind. The return-forecasting coefficient is $b_r \approx 0.10$, the dividend-growth forecasting coefficient is $b_d \approx 0$, and the OLS estimate of the dividend-yield autocorrelation is $\phi \approx 0.94$. The standard errors are about the same, 0.05 in each case.

Alas, the identity (5) is not exact. The "implied" column of Table 2 gives each coefficient implied by the other two equations and the identity, in which I calculate ρ from the mean log dividend yield as

$$\rho = \frac{e^{E(p-d)}}{1 + e^{E(p-d)}} = 0.9638$$

The difference is small, about 0.005 in each case, but large enough to make a visible difference in the results. For example, the t-statistic calculated from the implied b_r coefficient is $0.101/0.050 = 2.02$ rather

Table 2
Forecasting regressions and null hypothesis

	Estimates			ε s. d. (diagonal) and correlation.			Null 1	Null 2
	$\hat{b}, \hat{\phi}$	$\sigma(\hat{b})$	implied	r	Δd	dp	b, ϕ	b, ϕ
r	0.097	0.050	0.101	19.6	66	−70	0	0
Δd	0.008	0.044	0.004	66	14.0	7.5	−0.0931	−0.046
dp	0.941	0.047	0.945	−70	7.5	15.3	0.941	0.99

Each row represents an OLS forecasting regression on the log dividend yield in annual CRSP data 1927–2004. For example, the first row presents the regression $r_{t+1} = a_r + b_r(d_t - p_t) + \varepsilon^r_{t+1}$. Standard errors $\sigma(\hat{b})$ include a GMM correction for heteroskedasticity. The "implied" column calculates each coefficient based on the other two coefficients and the identity $b_r = 1 - \rho\phi + b_d$, using $\rho = 0.9638$. The diagonals of the "ε s. d." matrix give the standard deviation of the regression errors in percent; the off-diagonals give the correlation between errors in percent. The "Null" columns describes coefficients used to simulate data under the null hypothesis that returns are not predictable.

than $0.097/0.05 = 1.94$, and we will see as much as two to three percentage point differences in probability values to follow.

The middle three columns of Table 2 present the error standard deviations on the diagonal and correlations on the off-diagonal. Returns have almost 20% standard deviation. Dividend growth has a large 14% standard deviation. In part, this number comes from large variability in dividends in the prewar data. In part, the standard method for recovering dividends from the CRSP returns[5] means that dividends paid early in the year are reinvested at the market return to the end of the year. In part, aggregate dividends, which include all cash payouts, are in fact quite volatile. Most importantly for the joint distributions that follow, return and dividend-yield shocks are strongly negatively correlated (−70%), in contrast to the nearly zero correlation between dividend-growth shocks and dividend-yield shocks (7.5%).

The final columns of Table 2 present the coefficients of the null hypotheses I use to simulate distributions. I set $b_r = 0$. I start by choosing ϕ at its sample estimate $\phi = 0.941$. I consider alternative values of ϕ below.

[5] CRSP gives total returns R and returns without dividends Rx. I find dividend yields by

$$\frac{D_{t+1}}{P_{t+1}} = \frac{R_{t+1}}{Rx_{t+1}} - 1 = \frac{P_{t+1} + D_{t+1}}{P_t}\frac{P_t}{P_{t+1}} - 1$$

I then can find dividend growth by

$$\frac{D_{t+1}}{D_t} = \frac{(D_{t+1}/P_{t+1})}{(D_t/P_t)} Rx_{t+1} = \frac{D_{t+1}}{P_{t+1}}\frac{P_t}{D_t}\frac{P_{t+1}}{P_t}$$

Cochrane (1991) shows that this procedure implies that dividends paid early in the year are reinvested at the return R to the end of the year. Accumulating dividends at a different rate is an attractive and frequently followed alternative, but then returns, prices, and dividends no longer obey the identity $R_{t+1} = (P_{t+1} + D_{t+1})/P_t$ with end-of-year prices.

Given $b_r = 0$ and ϕ, the necessary dividend forecast coefficient b_d follows from the identity $b_d = \rho\phi - 1 + b_r \approx -0.1$.

We have to choose two variables to simulate and then let the third follow from the identity (4). I simulate the dividend-growth and dividend-yield system. This is a particularly nice system, since we can interpret the errors as essentially uncorrelated "shocks to expected dividend growth" and "shocks to actual dividend growth" respectively. (Formally, the VAR (7) can be derived from a model in which expected dividend growth follows an AR(1), $E_t(\Delta d_{t+1}) = x_t = \phi x_{t-1} + \delta_t^x$, returns are not forecastable, and dividend yields are generated from the present value identity (9)). However, the identity (4) holds well enough that this choice has almost no effect on the results.

In sum, the null hypotheses takes the form

$$\begin{bmatrix} d_{t+1} - p_{t+1} \\ \Delta d_{t+1} \\ r_{t+1} \end{bmatrix} = \begin{bmatrix} \phi \\ \rho\phi - 1 \\ 0 \end{bmatrix} (d_t - p_t) + \begin{bmatrix} \varepsilon_{t+1}^{dp} \\ \varepsilon_{t+1}^{d} \\ \varepsilon_{t+1}^{d} - \rho\varepsilon_{t+1}^{dp} \end{bmatrix} \tag{7}$$

I use the sample estimate of the covariance matrix of ε^{dp} and ε^{d}. I simulate 50,000 artificial data sets from each null. For $\phi < 1$, I draw the first observation $d_0 - p_0$ from the unconditional density $d_0 - p_0 \sim \mathcal{N}\left[0, \sigma^2\left(\varepsilon^{dp}\right)/(1-\phi^2)\right]$. For $\phi \geq 1$, I start at $d_0 - p_0 = 0$. I then draw ε_t^d and ε_t^{dp} as random normals and simulate the system forward.

2. Distribution of Regression Coefficients and *t*-statistics

2.1 Return and dividend-growth forecasts

In each Monte Carlo draw I run regressions (1)–(3). Figure 1 plots the joint distribution of the return b_r and dividend-growth b_d coefficients, and the joint distribution of their t-statistics. Table 3 collects probabilities.

The *marginal* distribution of the return-forecast coefficient b_r gives quite weak evidence against the unforecastable-return null. The Monte Carlo draw produces a coefficient larger than the sample estimate 22% of the time, and a larger t-statistic than the sample about 10% of the time (points to the right of the vertical line in the top panels of Figure 1, top left entries of Table 3). Taken on its own, we cannot reject the hypothesis that the return-forecasting coefficient b_r is zero at the conventional 5% level. This finding confirms the results of Goetzmann and Jorion (1993), Nelson and Kim (1993), and Stambaugh (1999).

However, the null must assume that dividend growth *is* forecastable. As a result, almost all simulations give a large negative dividend-growth forecast coefficient b_d. The null and cloud of estimates in Figure 1 are vertically centered a good deal below zero and below the horizontal line of the sample estimate $\hat{b}_d$. *Dividend*-growth forecasting coefficients larger

Table 3
Percent probability values under the $\phi = 0.941$ null

	b_r	t_r	b_d	t_d
Real	22.3	10.3	1.77	1.67
Excess	17.4	6.32	1.11	0.87

Each column gives the probability that the indicated coefficients are greater than their sample values, under the null. Monte Carlo simulation of the null described in Table 2 with 50,000 draws.

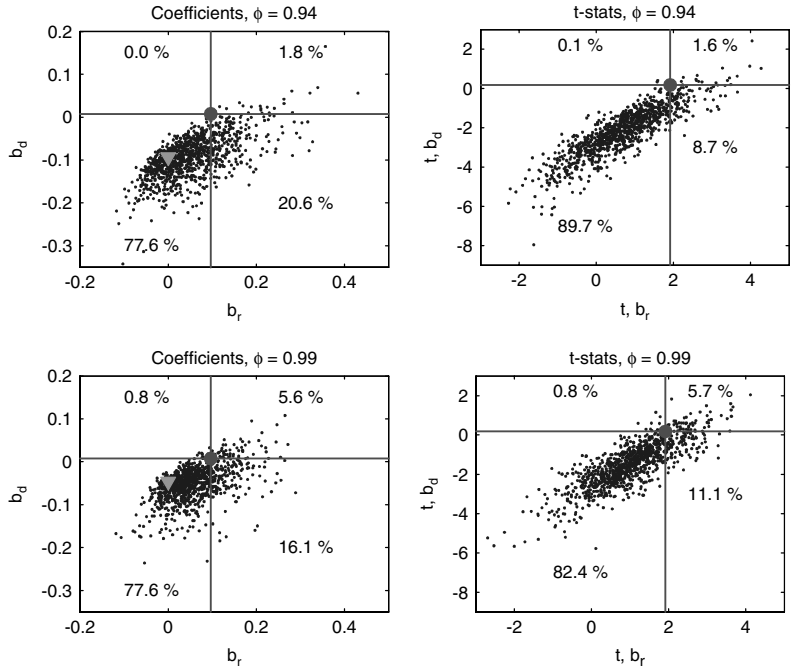

Figure 1
Joint distribution of return and dividend-growth forecasting coefficients (left) and t-statistics (right). The lines and large dots give the sample estimates. The triangle gives the null. One thousand simulations are plotted for clarity; each point represents 1/10% probability. Percentages are the fraction of 50,000 simulations that fall in the indicated quadrants.

than the roughly zero values observed in sample are seen only 1.77% of the time, and the dividend-growth t-statistic is only greater than its roughly zero sample value 1.67% of the time (points above the horizontal lines in Figure 1, b_d and t_d columns of Table 3). Results are even stronger for excess returns, for which $b_d > \hat{b}_d$ is observed only 1.11% of the time and the t-statistic only 0.87% of the time (Table 3).

In sum, the *lack* of dividend forecastability in the data gives far stronger statistical evidence against the null than does the *presence* of return

forecastability, lowering probability values from the 10–20% range to the 1–2% range. (I discuss the $\phi = 0.99$ results seen in Figure 1 below.)

2.2 Long-run coefficients

If we divide the identity (5) $b_r - b_d = 1 - \rho\phi$ by $1 - \rho\phi$, we obtain the identity

$$\frac{b_r}{1-\rho\phi} - \frac{b_d}{1-\rho\phi} = 1 \tag{8}$$
$$b_r^{lr} - b_d^{lr} = 1$$

The second row defines notation.

The terms of identity (8) have useful interpretations. First, b_r^{lr} is the regression coefficient of *long-run* returns $\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$ on dividend yields $d_t - p_t$, and similarly for b_d^{lr}, hence the lr superscript. Second, b_r^{lr} and $-b_d^{lr}$ represent the fraction of the variance of dividend yields that can be attributed to time-varying expected returns and to time-varying expected dividend growth, respectively.

To see these interpretations, iterate the return identity (4) forward, giving the Campbell–Shiller (1988) present value identity

$$d_t - p_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} - E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} \tag{9}$$

Multiply by $(d_t - p_t) - E(d_t - p_t)$ and take expectations, giving

$$var(d_t - p_t) = cov\left(\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}, d_t - p_t\right)$$
$$-cov\left(\sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}, d_t - p_t\right) \tag{10}$$

This equation states that all variation in the dividend-price (or price-dividend) ratio must be accounted for by its covariance with, and thus ability to forecast, future returns or future dividend growth. Dividing by $var(d_t - p_t)$ we can express the variance decomposition in terms of regression coefficients

$$\beta\left(\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}, d_t - p_t\right) - \beta\left(\sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}, d_t - p_t\right) = 1 \tag{11}$$

where $\beta(y, x)$ denotes the regression coefficient of y on x. In the context of our VAR(1) representation we have

$$\beta\left(\sum_{j=1}^{\infty}\rho^{j-1}r_{t+j}, d_t - p_t\right) = \sum_{j=1}^{\infty}\rho^{j-1}\beta\left(r_{t+j}, d_t - p_t\right)$$
$$= \sum_{j=1}^{\infty}\rho^{j-1}\phi^{j-1}b_r = \frac{b_r}{1-\rho\phi} = b_r^{lr} \tag{12}$$

and similarly for dividend growth. This sort of calculation is the standard way to adapt the ideas of Shiller (1981) and LeRoy and Porter (1981) volatility tests to the fact that dividend yields rather than prices are stationary. See Campbell and Shiller (1988) and Cochrane (1991, 1992, 2004) for more details.

2.3 Long-run estimates and tests

Table 4 presents estimates of the long-horizon regression coefficients. These are not new estimates, they are simply calculations based on the OLS estimates $\hat{b}_r, \hat{b}_d, \hat{\phi}$ in Table 2. I calculate asymptotic standard errors using the delta-method and the heteroskedasticity-corrected OLS standard errors from Table 2.

Table 4 shows that dividend-yield volatility is almost exactly accounted for by return forecasts, $\hat{b}_r^{lr} \approx 1$, with essentially no contribution from dividend-growth forecasts $\hat{b}_d^{lr} \approx 0$. This is another sense in which return forecastability is economically significant. This finding is a simple consequence of the familiar one-year estimates. $\hat{b}_d \approx 0$ means $\hat{b}_d^{lr} \approx 0$, of course, and

$$\hat{b}_r^{lr} = \frac{\hat{b}_r}{1-\rho\hat{\phi}} \approx \frac{0.10}{1-0.96\times 0.94} \approx 1.0$$

Table 4
Long-run regression coefficients

Variable	$\hat{b}^{lr}$	s. e.	t	% p value
r	1.09	0.44	2.48	1.39–1.83
Δd	0.09	0.44	2.48	1.39–1.83
Excess r	1.23	0.47	2.62	0.47–0.69

The long-run return forecast coefficient $\hat{b}_r^{lr}$ is computed as $\hat{b}_r^{lr} = \hat{b}_r/(1-\rho\hat{\phi})$, where $\hat{b}_r$ is the regression coefficient of one-year returns r_{t+1} on $d_t - p_t$, $\hat{\phi}$ is the autocorrelation of $d_t - p_t$, $\rho = 0.961$, and similarly for the long-run dividend-growth forecast coefficient $\hat{b}_d^{lr}$. The standard error is calculated from standard errors for $\hat{b}_r$ and $\hat{\phi}$ by the delta method. The t-statistic for Δd is the statistic for the hypothesis $\hat{b}_d^{lr} = -1$. Percent probability values (% p value) are generated by Monte Carlo under the $\phi = 0.941$ null. The range of probability values is given over the three choices of which coefficient ($\hat{b}_r, \hat{\phi}, \hat{b}_d$) is implied from the other two.

In fact, the point estimates in Table 4 show slightly more than 100% of dividend-yield volatility coming from returns, since the point estimate of dividend-growth forecasts go slightly the wrong way. The decomposition (10) is not a decomposition into orthogonal components, so elements can be greater than 100% or less than 0% in this way. Excess returns in the last row of Table 4 show slightly stronger results. In the point estimates, high price-dividend ratios actually signal slightly higher interest rates, so they signal even lower excess returns.

The first two rows of Table 4 drive home the fact that, by the identity $b_r^{lr} - b_d^{lr} = 1$, the long-horizon dividend-growth regression gives *exactly* the same results as the long-horizon return regression.[6] The standard errors are also exactly the same, and the t-statistic for $b_r^{lr} = 0$ is exactly the same as the $t-$statistic for $b_d^{lr} = -1$.

One great advantage of using long-horizon regression coefficients is that we do not need to choose between return and dividend-growth tests, as they give precisely the same results. As a result, we can tabulate the small-sample distribution of the test in a conventional histogram, rather than a two-dimensional plot.

Figure 2 tabulates the small-sample distribution of the long-run return-forecast coefficients, and Table 4 includes the probability values—how many long-run return forecasts are greater than the sample value under the unforecastable-return null $b_{lr}^r = 0$. There is about a 1.5% probability value of seeing a long-run forecast larger than seen in the data. (The range of probability values in Table 4 derives from the fact that the identities are only approximate, so the result depends on which of the three parameters (b_r, ϕ, b_d) is implied from the other two.) The long-run return (or dividend-growth) regressions give essentially the same strong rejections as the short-run dividend-growth regression.

The last row of Table 4 shows the results for excess returns. Again, excess returns paint a stronger picture. The probability values of 0.38–0.64% are lower and the evidence against the null even stronger.

3. Power, Correlation, and the ϕ View

Where does the greater power of dividend-growth and long-run return tests come from? How do we relate these results to the usual analysis of the $\{b_r, \phi\}$ coefficients in a two-variable VAR consisting of the return and the forecasting variable?

[6] The identities are only approximate, so to display estimates that obey the identities one must estimate two of b_r, b_d, and ϕ, and imply the other using the identity $b_r - b_d = 1 - \rho\phi$. In the top two lines of Table 4, I use the direct $\hat{b}_r$ and $\hat{b}_d$ estimates from Table 2. I then use $\rho\hat{\phi}_{impl} = 1 - \hat{b}_r + \hat{b}_d$ and I construct long-run estimates by $\hat{b}_r^{lr} = \hat{b}_r/(1 - \rho\hat{\phi}_{impl})$. Since $\hat{\phi} = 0.94$ and $\hat{\phi}_{impl} = 0.95$, the difference between these estimates and those that use $\hat{\phi}$ is very small. Using the direct estimate $\hat{\phi}$ rather than $\hat{\phi}_{impl}$, we have $\hat{b}_r^{lr} = 1.04$ ($s.e. = 0.42$) and $b_d^{lr} = 0.08$ ($s.e. = 0.42$).

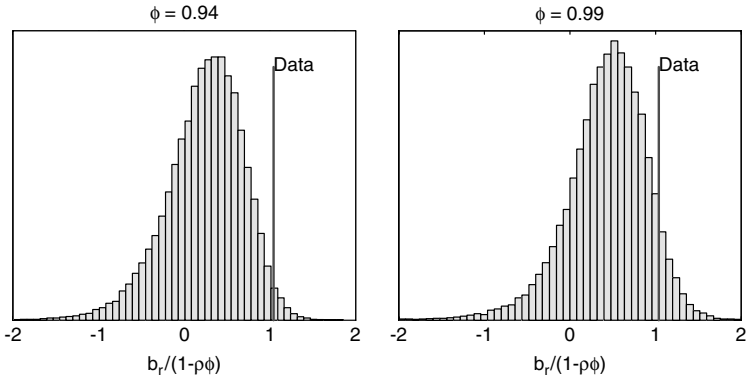

Figure 2
Distribution of $b_r^{lr} = b_r/(1-\rho\phi)$. The vertical bar gives the corresponding value in the data.

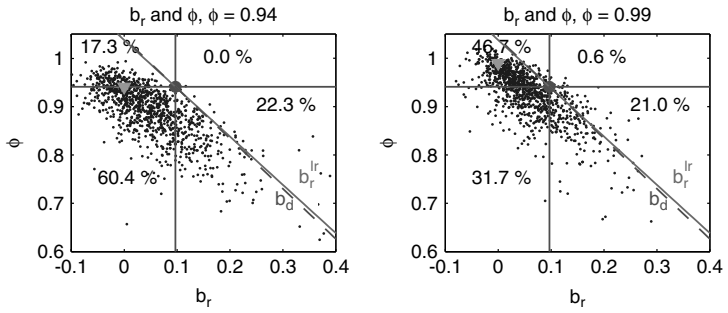

Figure 3
Joint distribution of return and dividend yield forecasting coefficients b_r, ϕ. In each graph the triangle marks the null hypothesis used to generate the data and the circle marks the estimated coefficients $\hat{b}_r, \hat{\phi}$. The diagonal dashed line marked "b_d" marks the line $b_r = 1 - \rho\phi + \hat{b}_d$; points above and to the right are draws in which b_d exceeds its sample value. The solid diagonal line marked "b_r^{lr}" marks the line defined by $b_r/(1-\rho\phi) = \hat{b}_r/(1-\rho\hat{\phi})$; points above and to the right are draws in which b_r^{lr} exceeds its sample value. Numbers are the percentage of the draws that fall in the indicated quadrants.

Figure 3 addresses these questions by plotting the joint distribution of estimates $\{b_r, \phi\}$ across simulations. We see again that a high return coefficient b_r by itself is not so unusual, occurring about 22% of the time (area to the right of the vertical line). We learn, however, that b_r and ϕ estimates are negatively correlated across samples. Though we often see large b_r and large ϕ, we almost never see b_r larger than in the data *together* with ϕ larger than in the data. (Figure 3 is the same as Lewellen (2004) Figure 1, Panel B, except Lewellen calibrates to monthly postwar data. Lewellen focuses on a different distributional calculation.)

This observation on its own is not a good way to form a test statistic. Though the northeast quadrant of the plot is suggestively empty, we would

not want to commit to accepting the null for ϕ just below a rectangular rejection region and arbitrarily large b_r.

The $\{b_r, \phi\}$ plot is more important to help us to digest why the dividend-growth test b_d and the long-horizon regression test b_r^{lr} give so many fewer rejections under then null than the usual one-period return b_r test. The diagonal dashed line marked b_d in Figure 3 uses the identity $b_r = 1 - \rho\phi + b_d$ to mark the region $b_d > \hat{b}_d$ in this $\{b_r, \phi\}$ space. Points above and to the right of this dashed line are exactly the points above $b_d > \hat{b}_d$ in Figure 1. The similar diagonal solid line marked b_r^{lr} uses the definition $b_r^{lr} = b_r/(1 - \rho\phi)$ to mark the region $b_r^{lr} > \hat{b}_r^{lr}$ in this $\{b_r, \phi\}$ space. Points above and to the right of this line are exactly the points above $b_r^{lr} > \hat{b}_r^{lr}$ in the histogram of Figure 2.

Viewed in $\{b_r, \phi\}$ space, dividend-growth b_d and long-run regression b^{lr} tests capture in a single number and in a sensible test region the fact that samples with high b_r typically come with low ϕ, and they exploit that negative correlation to produce more powerful tests. By the definition $b_r/(1 - \rho\phi)$, samples with high b_r but low ϕ produce a low long-run return forecast b_r^{lr}. Generating a large long-run return forecast requires both a large one-year return forecast and a large autocorrelation, so forecasts can build with horizon. Because most large return forecasts come with low autocorrelation, it is much harder for the null to deliver a large long-run return forecast. By the identity $b_d = b_r + \rho\phi - 1$, dividend growth works the same way.

3.1 The source of negative correlation

The strong negative correlation of *estimates* b_r and ϕ across samples, which underlies the power of long-horizon return and dividend-growth tests, stems from the strong negative correlation of the *shocks* ε_{t+1}^r and ε_{t+1}^{dp} in the underlying VAR, (1)–(3). If shocks are negatively correlated in two regressions with the same right-hand variable, then a draw of shocks that produces an unusually large coefficient in the first regression corresponds to a draw of shocks that produces an unusually small coefficient in the second regression.

It is important to understand this correlation. We do not want the power of long-run or dividend-growth tests to hinge on some arbitrary and inessential feature of the null hypothesis, and strong correlations of shocks are usually not central parts of a specification.

From the identity

$$\varepsilon_{t+1}^r = \varepsilon_{t+1}^d - \rho\varepsilon_{t+1}^{dp} \tag{13}$$

the fact that return shocks and dividend-yield shocks *are* strongly and negatively correlated is equivalent to the fact that dividend-yield shocks and dividend-growth shocks *are not* correlated. Intuitively, we can see this

fact by looking at the definition of return

$$R_{t+1} = \frac{(1 + P_{t+1}/D_{t+1})}{P_t/D_t}\frac{D_{t+1}}{D_t}$$

that underlies (13): a decline in dividend yield D_{t+1}/P_{t+1} is a rise in prices, which raises returns, but only so long as there is no offsetting change in dividend growth D_{t+1}/D_t. More precisely, multiply both sides of (13) by ε^{dp}_{t+1} and take expectations, yielding

$$cov\left(\varepsilon^{r}_{t+1}, \varepsilon^{dp}_{t+1}\right) = cov\left(\varepsilon^{dp}_{t+1}, \varepsilon^{d}_{t+1}\right) - \rho\sigma^2\left(\varepsilon^{dp}_{t+1}\right) \tag{14}$$

When dividend growth and dividend yields are uncorrelated $cov(\varepsilon^{dp}_{t+1}, \varepsilon^{d}_{t+1}) = 0$, we obtain a strong negative correlation between returns and dividend yields $cov(\varepsilon^{r}_{t+1}, \varepsilon^{dp}_{t+1})$.

Dividend yields move on news of expected returns (in the point estimates) or news of expected dividend growth (in the null) (see (9)). Thus, the central fact in our data is that shocks to *expected* returns (data) or *expected* dividend growth (null) are uncorrelated with shocks to *ex post* dividend growth. The strong negative correlation of dividend-yield shocks with return shocks follows from the definition of a return.

It seems that we can easily imagine other structures, however. For example, in typical time-series processes, like an AR(1), shocks to *ex post* dividend growth are correlated with shocks to expected dividend growth; only rather special cases do not display this correlation. In economic models, it is not inconceivable that a negative shock to current dividends would raise risk premia, raising expected returns and thus dividend yields.

However, identity (13) makes it hard to construct plausible alternatives. There are only three degrees of freedom in the variance-covariance matrix of the three shocks, since any one variable can be completely determined from the other two. As a result, changing one correlation forces us to change the rest of the covariance matrix in deeply counterfactual ways.

For example, let us try to construct a covariance matrix in which return and dividend-yield shocks are uncorrelated, $cov(\varepsilon^{r}_{t+1}, \varepsilon^{dp}_{t+1}) = 0$. Let us continue to match the volatility of returns $\sigma(\varepsilon^{r}_{t+1}) = 0.2$ and dividend yields $\sigma(\varepsilon^{dp}_{t+1}) = 0.15$. We cannot, however, match the volatility of dividend growth. Writing the identity (13) as

$$\varepsilon^{d}_{t+1} = \varepsilon^{r}_{t+1} + \rho\varepsilon^{dp}_{t+1} \tag{15}$$

we see that in order to produce $cov(\varepsilon^{r}_{t+1}, \varepsilon^{dp}_{t+1}) = 0$, we must specify dividend-growth shocks that are more volatile than returns! We need to specify 25% dividend-growth volatility, rather than the 14% volatility in

our data:

$$\sigma(\varepsilon^d) = \sqrt{\sigma^2(\varepsilon^r_{t+1}) + \rho^2\sigma^2(\varepsilon^{dp}_{t+1})}$$
$$= \sqrt{0.20^2 + 0.96^2 \times 0.15^2} = 0.25$$

It is a quite robust fact that return variation is dominated by variation of *prices* or *valuations* with little change in cashflows, and more so at high frequencies. The variance of returns far exceeds the variance of dividend growth. By (15) that fact alone implies that positive innovations to current returns ε^r_{t+1} must come with negative innovations to dividend yields.

Continuing the example, we can find the required correlation of dividend-growth shocks and dividend-yield shocks by multiplying (15) by ε^{dp}_{t+1} giving

$$cov\left(\varepsilon^d, \varepsilon^{dp}\right) = cov(\varepsilon^r, \varepsilon^{dp}) + \rho\sigma^2\left(\varepsilon^{dp}\right)$$
$$corr\left(\varepsilon^d, \varepsilon^{dp}\right) = \rho\frac{\sigma\left(\varepsilon^{dp}\right)}{\sigma(\varepsilon^d)} = 0.96 \times \frac{0.15}{0.25} = 0.58$$

rather than 0.07, essentially zero, in the data (Table 2). In this alternative world, good news about dividend growth is frequently accompanied by an increase in dividend yield, meaning prices do not move that much. In turn, that means the good news about dividend growth comes either with news that future dividend growth will be low—dividends have a large mean-reverting component—or with news that future expected returns will be high. In our data, dividend-yield shocks typically raise prices proportionally, leading to no correlation with dividend yields.

Needless to say, the changes to the covariance matrix required to generate a *positive* correlation between return and dividend-yield shocks are even more extreme. In sum, the negative correlation of estimates $\{b_r, \phi\}$, which ultimately derives from the negative correlation of shocks $\varepsilon^r_{t+1}, \varepsilon^{dp}_{t+1}$ or equivalently from the near-zero correlation of shocks $\varepsilon^{dp}_{t+1}, \varepsilon^d_{t+1}$ is a deep and essential feature of the data, not an easily changed auxiliary to the null.

3.2 Which is the right region?—economics

We now have three tests: the one-period regression coefficients b_r and b_d, and the long-horizon regression coefficient b^{lr}. Which is the right one to look at? Should we test $b_r > \hat{b}_r$, or should we test $b_d > \hat{b}_d$, or $b^{lr}_r > \hat{b}^{lr}_r$? Or perhaps we should test some other subset of the $\{b_r, \phi\}$ region?

The central underlying question is, how should we form a single test statistic from the joint distribution of many parameters? We have three parameters, b_r, b_d, ϕ. The identity $b_r = 1 - \rho\phi + b_d$ means we can reduce the issue to a two-dimensional space, but we still have two dimensions to think about.

In *economic* terms, we want the most interesting test. The issue comes down to defining what is the "event" we have seen, and what other events we would consider "more extreme," or "further from the null" than the event we have seen. If we focus on the one-year return regression, we think of the "event" as the return forecast coefficient seen in the data $b_r = \hat{b}_r \approx 0.1$, and "more extreme" events as those with greater one-year return-forecast coefficients, $b_r > \hat{b}_r$. But, as the joint distributions point out, most of the events with $b_r > \hat{b}_r$, have dividend-growth forecast coefficients larger (more negative) than seen in the data, $b_d < \hat{b}_d$, they have dividend-yield autocorrelations lower than seen in the data $\phi < \hat{\phi}$, they have long-run return coefficients less than seen in the data $b_r^{lr} < \hat{b}_r^{lr}$, and thus (by identity) they have long-run dividend-growth coefficients larger (more negative) than seen in the data, $b_d^{lr} < \hat{b}_d^{lr} \approx 0$. In these events, dividend-growth *is* forecastable, prices *are* moving to some extent on forecasts of future dividend growth, and in the right direction. Volatility tests are a half-success, rather than the total failure that they are in our data. The long-run coefficients count these draws as "closer to the null" than our data, despite the larger values of b_r. From this point of view, the test on the long-run coefficient is the economically interesting test. If we want to view that test in the $\{b_r, \phi\}$ space of one-period regression coefficients, diagonal test regions as marked by b_r^{lr} in Figure 3 are the right ones to look at.

The dividend-growth coefficient tests $b_d > \hat{b}_d$ give almost exactly the same answers as the long-run coefficient tests, as can be seen both in the tables and by the fact that the dividend b_d and long-run b_r^{lr} regions of Figure 3 are nearly the same. In fact, these tests are different conceptually and slightly different in this sample. The long-run return coefficient test $b_r^{lr} > \hat{b}_r^{lr}$ means $b_d^{lr} > \hat{b}_d^{lr}$ which means $b_d/(1-\rho\phi) > \hat{b}_d/(1-\rho\hat{\phi})$. If we had $\hat{b}_d = 0$ exactly, this would mean $b_d > \hat{b}_d = 0$ and the two regions would be exactly the same. With $\hat{b}_d \neq 0$, a different sample ϕ can affect the long-run dividend-growth coefficient $b_d^{lr} = b_d/(1-\rho\phi)$ for a given value of b_d, perhaps pushing it across a boundary. In a sample with $\hat{b}_d$ further from zero, the two test statistics could give substantially different answers.

When there is a difference, I find the long-run coefficients more economically attractive than the dividend-growth coefficients. As an important practical example, think about the specification of the null. In long-run coefficient terms, we specify the null as $b_d^{lr} = -1, b_r^{lr} = 0$, that is, all variation in dividend yields is due to time-varying expected dividend growth and none to time-varying expected returns. In short-run coefficient terms, this specification is equivalent to $b_d = 1/(1-\rho\phi)$. At the sample $\phi = \hat{\phi} \approx 0.96$, we have $b_d \approx -0.1$. As we vary ϕ, however, we vary b_d to keep $b_d^{lr} = -1$. This is exactly how I specify the $\phi = 0.99$ null above and how I specify the null for different ϕ values below.

Suppose instead that we specify the null in short-run coefficient terms as $b_d = -0.1$ for any value of ϕ. Now, different values of ϕ give us specifications in which expected returns do explain nonzero fractions of dividend yield and in which returns are predictable. For example, at $\phi = 0.99$, we would have $b_d^{lr} = -0.1/(1 - 0.96 \times 0.99) \approx -2$ and thus $b_r^{lr} \approx -1$, with $b_r \approx (1 - 0.96 \times 0.99) \times (-1) \approx 0.05$. In this null, a rise in prices signals so much higher dividend growth that it must also signal much higher future returns. Obviously, this is not a very interesting way to express the null hypothesis that returns are unpredictable. The same sorts of things happen to test regions if $b_d \neq 0$.

If one accepts that the *null* should be expressed this way, in terms of long-horizon coefficients to accommodate variation in ϕ, it seems almost inescapable that the economically interesting *test region* should be specified in the same way.

3.3 Which is the right region?—statistics

In *statistical* terms we want the most powerful test. It is clear that the dividend growth b_d and long-run b_r^{lr} tests, implying a test of a diagonal region in $\{b_r, \phi\}$ space, are more powerful. It is important to understand the source of that power.

"Power," of course, is not the probability under the null of finding more extreme statistics that I have calculated. To document power, I should set up regions based on b_r, b_d, and b^{lr} that reject at a given level, say 5% of the time, under the null. Then I should evaluate the probability that draws enter those rejection regions under alternatives, in particular generating data from the estimated parameters $\hat{b}_r$ and $\hat{\phi}$. I should document that draws do enter the long-run or dividend-growth rejection regions more frequently. I do not perform these calculations in the interest of brevity, since it is clear from the graphs how they work out. Since the b_r vertical line in Figure 3 demarks a 22% probability value now, the boundary of the 5% region under the null is farther to the right. Since the b_d and b^{lr} diagonal lines demark 1–2% probability values now, the boundaries of the 5% regions under the null are a bit to the left of the current lines. Under the alternative, the cloud of points in Figure 3 moves to the right—drag the triangle to the circle and move the cloud of points with it. Because of the negative correlation between b_r and ϕ estimates, that operation will drag roughly half of the simulated data points across the diagonal lines, but it will still leave the bulk of the data points shy of the 5% vertical b_r region. The long-run and dividend-growth tests do have more power.

Therefore, one key to the extra power is the negative correlation between b_r and ϕ coefficients, documented by Figure 3. If the cloud sloped the other way, there would be no power advantage.

The other key to extra power is a limitation on ϕ in the null. So far, I have calculated test statistics from a *point* null, specifying $\{b_r = 0, \phi = 0.941\}$.

One wants obviously to think about other, and particularly larger, values of ϕ. As we raise ϕ in the null, the cloud of points in Figure 3 rises. The right-hand panel of Figure 3 shows this rise for $\phi = 0.99$. If we were to raise ϕ arbitrarily, say to 2 or 3, the cloud of points would rise so much that all of them would be above the diagonal lines. The null would generate greater long-run and dividend predictability than the sample 100% of the time, and the power would decline to zero. The b_r test, based on a vertical rejection region, would still have something like the same probability of rejection, and we would reverse the power advantages.

But of course $\phi = 2$ or $\phi = 3$ are ridiculous null hypotheses. The null should be a coherent, economically sensible view of the world, and explosive price-dividend ratios do not qualify. I argue in detail below for upper limits between $\phi = 0.99$ and $\phi = 1.04$. For the moment, it is enough to accept that there is *some* upper limit $\overline{\phi}$ that characterizes coherent and economically sensible null hypotheses.

This observation solves a statistical mystery. How can it be that a powerful test of a simple null hypothesis like $b_r = 0$ involves other parameters, or a joint region in $\{b_r, \phi\}$ space? The answer is that the null hypothesis is not $\{b_r = 0, \phi = \text{anything}\}$, but it is $\left\{b_r = 0, \|\phi\| < \overline{\phi}\right\}$ (Going further, one-sided tests make more sense in this context.) Given such a *joint* null hypothesis, any sensible test will set up a region surrounding the null hypothesis, and thus use information on both parameters. For example, if the upper bound is $\overline{\phi} = 1.04$, the observation $\hat{\phi} = 3$ would reject the null even if the estimate is $\hat{b}_r = 0$. A test region surrounding the null $\left\{b_r = 0, \|\phi\| < \overline{\phi}\right\}$ will be downward sloping in $\{b_r, \phi\}$ space in the region of our sample estimates, with large $\hat{b}_r$ and large $\hat{\phi}$, exactly as the long-run regression test or dividend-growth coefficient test give downward sloping regions in $\{b_r, \phi\}$ space.

For example, for $\|\phi\| < 1$ and in large samples, we can write the likelihood ratio test statistic as[7]

$$LR \approx (T-1)\ln\left[1 + \frac{\hat{b}^2}{1-\hat{\phi}^2}\frac{\sigma^2(\varepsilon^{dp}_{t+1})}{\sigma^2(\varepsilon^{r}_{t+1})}\right]$$

[7] The likelihood ratio is the ratio of constrained ($b_r = 0$) to unconstrained (OLS) residual variances. We can write

$$LR = (T-1)\ln\frac{\sigma^2(r_{t+1})}{\sigma^2(\varepsilon^r_{t+1})} = (T-1)\ln\frac{\sigma^2(\hat{b}(d_t - p_t) + \varepsilon^r_{t+1})}{\sigma^2(\varepsilon^r_{t+1})}$$

$$= (T-1)\ln\left[\hat{b}^2\frac{\sigma^2(d_t - p_t)}{\sigma^2(\varepsilon^r_{t+1})} + 1\right] = (T-1)\ln\left[\frac{\hat{b}^2}{1-\hat{\phi}^2}\frac{\sigma^2(\varepsilon^{dp}_{t+1})}{\sigma^2(\varepsilon^r_{t+1})} + 1\right]$$

where $\sigma^2(\varepsilon^r_{t+1})$ and $\sigma^2(\varepsilon^{dp}_{t+1})$ are the unconstrained, OLS regression errors. The corresponding test regions in $\{b_r, \phi\}$ space are ellipses surrounding the line $\{b_r = 0, \|\phi\| < 1\}$, and slope downward through the data point $\left\{\hat{b}_r, \hat{\phi}\right\}$ with, it turns out, almost exactly the same slope as the long-run b^{lr}_r or dividend-growth b_d regions. If we expand the null to $\{b_r = 0, \phi = \text{anything}\}$, then of course the likelihood ratio test becomes asymptotically the same as the Wald test on b_r alone. (It turns out that the likelihood ratio test is not as powerful as the long-horizon or dividend-growth regressions in this data set, in part because it is a two-sided test and in part because it does not exploit the correlation of errors, so I do not pursue it further.)

4. Autocorrelation ϕ, Unit Roots, Bubbles, and Priors

Higher values of ϕ lower the power of the dividend-growth and long-run test statistics. Thus, we have to ask, first, how large a value of ϕ should we consider in our null hypothesis, and second, how large, quantitatively, is the loss of power as we move toward sensible upper limits for ϕ?

4.1 How large can ϕ be?

We can start by ruling out $\phi > 1/\rho \approx 1.04$, since this case implies an infinite price-dividend ratio, and we observe finite values. Iterating forward the return identity (4), we obtain the present value identity

$$p_t - d_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} + \lim_{k\to\infty} \rho^k E_t \left(p_{t+k} - d_{t+k}\right) \tag{16}$$

In our VAR(1) model, the last term is $\rho^k \phi^k \left(p_t - d_t\right)$, and it explodes if $\phi > 1/\rho$.

If we have $\phi = 1/\rho \approx 1.04$, then it seems we *can* adopt a null with both $b_r = 0$ and $b_d = 0$, and respect the identity $b_r = 1 - \rho\phi + b_d$. In fact, in this case we *must* have $b_r = b_d = 0$, otherwise terms such as $E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} = \sum_{j=1}^{\infty} \rho^{j-1} \phi^{j-1} b_d (d_t - p_t)$ in Equation (16) do not converge. This is a "rational bubble." If $\phi = 1/\rho$ exactly, then price-dividend ratios can vary on changing expectations of future price-dividend ratios, the last term of Equation (16), with no news at all about dividends or expected returns. This view is hard to hold as a matter of economic theory, so I rule it out on that basis. Since I will argue against any $\phi \geq 1$, it does not make sense to spend a lot of time on a review of the rational bubbles literature to rule out $\phi = 1.04$.

At $\phi = 1$, the dividend yield follows a random walk. $\phi = 1$ still implies some predictability of returns or dividend growth, $b_r + b_d = 1 - \rho\phi \approx 0.04$. If prices and dividends are not expected to move after a

dividend-yield rise, the higher dividend yield still means more dividends and thus a higher return. $\phi = 1$ does not cause trouble for the present value model; $\phi = 1$ is the point at which the *statistical* model explodes to an infinite unconditional variance. $\phi = 1$ does cause trouble for loglinear approximation of course, since that approximation is only valid near the expansion point. To take $\phi = 1$ seriously, one really has to move past local approximations.

Can we consider a unit root in dividend yields? The dividend yield does pass standard unit root tests (Craine (1993)), but with $\hat{\phi} = 0.941$ that statistical evidence will naturally be tenuous. In my simulations with $\phi = 1$, the observed $\hat{\phi} = 0.941$ is almost exactly the median value, so I do not reject $\phi = 1$ on that basis.

However, we do not need a continuous data set to evaluate statistical questions, and evidence over very long horizons argues better against a random walk for the dividend yield. Stocks have been trading since the 1600s, giving spotty observations of prices and dividends, and privately held businesses and partnerships have been valued for a millennium. A random walk in dividend yields generates far more variation than we have seen in that time. Using the measured 15% innovation variance of the dividend yield, and starting at a price/dividend ratio of 25 (1/0.04), the one-century one-standard deviation band—looking backwards as well as forwards—is a price-dividend ratio between[8] 5.6 and 112, and the ± 2 standard deviation band is between[9] 1.24 and 502. In 300 years, the bands are $\pm 1\sigma = 1.9$ to 336, and $\pm 2\sigma = 0.14$ to 4514. If dividend yields really follow a random walk, we should have seen observations of this sort. But market-wide price-dividend ratios of two or three hundred have never been approached, let alone price-dividend ratios below one or over a thousand.

Looking forward, and as a matter of economics, do we really believe that dividend yields will wander *arbitrarily* far in either the positive or negative direction? Are we likely to see a market price-dividend ratio of one, or one thousand, in the next century or two? Is the unconditional variance of the dividend yield really *infinite?*

In addition, the present value relation (9) means that a unit root in the dividend yield requires a unit root in stock returns or dividend growth: if r and Δd are stationary, then $p_t - d_t = E_t \sum \rho^{j-1} \left(\Delta d_{t+j} - r_{t+j}\right)$ is stationary as well, and conversely. Almost all economic models describe stationary returns and stationary dividend-growth rates, and the same sort of long-run volatility calculations give compelling intuition for that specification.

[8] That is between $e^{\ln(25)-0.15\sqrt{100}} = 5.6$ and $e^{\ln(25)+0.15\sqrt{100}} = 112$.

[9] That is $e^{\ln(25)-2\times 0.15\sqrt{100}} = 1.24$ and $e^{\ln(25)+2\times 0.15\sqrt{100}} = 502$.

Having argued against $\phi = 1$, how close to one should we seriously consider as a null for ϕ? Neither the statistical nor the economic arguments against $\phi = 1$ rest on an exact random walk in dividend yields. Both arguments center on the conditional variance of the price-dividend ratio, returns, and dividend-growth rates over long horizons, and $\phi = 0.999$ or $\phi = 1.001$ generate about the same magnitudes as $\phi = 1.000$. Thus, if $\phi = 1.00$ is too large to swallow, there is some range of $\phi < 1$ that is also too large to swallow.

4.2 Results for different ϕ values

Table 5 collects probability values for various events as a function of ϕ. The previous figures include the case $\phi = 0.99$. As ϕ rises, Figure 3 shows that more points cross the b_d and b^{lr} boundaries, giving higher probabilities. We see the same point in the $\{b_r, b_d\}$ region of Figure 1: as ϕ rises, $b_d = 1 - \rho\phi + 0$ also rises, so the cloud of points rises, and more of them cross the $b_d > \hat{b}_d$ line. However, probability values are not so simple as a vertical translation of the sampling draws. The small-sample biases increase as ϕ rises, so the clouds do not rise quite as far as the null hypothesis, and this attenuates the loss of power somewhat. A quantitative evaluation of the effects of higher ϕ is important.

Table 5 verifies the conjecture that as ϕ rises, the probability of the one-period return coefficient b_r exceeding its sample value is little changed, at about 20–22% for all values of ϕ.

Looking down the b_d column of Table 5, the $b_d > \hat{b}_d$ probability for real returns rises with ϕ. It crosses the 5% mark a bit above $\phi = 0.98$ and is still below 10% at $\phi = 1$. Excess returns give stronger results as usual, with the b_d probability value still below 5% at $\phi = 1$. The probability values of the long-run coefficients b^{lr} are nearly the same as those of the dividend-growth coefficients b_d, which we expect since the dividend-growth and long-run regression regions are nearly the same in our data. (This would not be true of a data set with an estimate $\hat{b}_d$ not so close to zero.)

Since the dividend-yield regression is a very simplified specification, and since 5% is an arbitrary cutoff, the main point is not exactly where a test crosses the 5% region. The main point is that in all cases, the dividend-growth b_d and long-run regression tests b^{lr} still have a great deal more power than the one-period return regressions b_r for any reasonable value of ϕ.

To get additional insight on upper limits for ϕ, the final columns of Table 5 include the unconditional variance of dividend yields and the half-life of dividend yields implied by various values of ϕ. The sample estimate $\hat{\phi} = 0.941$ is consistent with the sample standard deviation of $\sigma(dp) = 0.45$, and a 11.4-year half-life of dividend-yield fluctuations. In the $\phi = 0.99$ null, the standard deviation of log dividend yields is actually 1.14, more than twice the volatility that has caused so much consternation

Table 5
The effects of dividend-yield autocorrelation ϕ

	Percent probability values								Other	
Null	Real returns				Excess returns				Statistics	
ϕ	b_r	b_d	$b^{lr}_{\min}$	$b^{lr}_{\max}$	b_r	b_d	$b^{lr}_{\min}$	$b^{lr}_{\max}$	$\sigma\,(dp)$	1/2 life
0.90	24	0.6	0.3	0.6	19	0.4	0.1	0.2	0.35	6.6
0.941	22	1.6	1.2	1.7	17	1.1	0.5	0.7	0.45	11.4
0.96	22	2.6	2.0	2.8	17	1.6	0.8	1.2	0.55	17.0
0.98	21	4.9	4.3	5.5	17	2.7	1.8	2.5	0.77	34.3
0.99	21	6.3	5.9	7.4	17	3.6	2.7	3.6	1.09	69.0
1.00	22	8.7	8.1	10	16	4.4	3.7	4.8	∞	∞
1.01	19	11	11	13	14	5.1	5.1	6.3	∞	∞
Draw ϕ	23	1.6	1.4	1.7	18	1.1	0.6	0.8		

The first column gives the assumed value of ϕ. "Draw ϕ" draws ϕ from the concentrated unconditional likelihood function displayed in Figure 4. "Percent probability values" give the percent chance of seeing each statistic larger than the sample value. b_r is the return forecasting coefficient, b_d is the dividend-growth forecasting coefficient. b^{lr} is the long-run regression coefficient, $b^{lr}_r = b_r/(1-\rho\phi)$. $b^{lr}_{\min}$ and $b^{lr}_{\max}$ are the smallest and largest values across the three ways of calculating the sample value of $b_r/(1-\rho\phi)$, depending on which coefficient is implied by the identity $b_r = 1 - \rho\phi + b_d$. $\sigma\,(dp)$ gives the implied standard deviation of the dividend yield $\sigma\,(dp) = \sigma\left(\varepsilon^{dp}\right)/\sqrt{1-\phi^2}$. Half life is the value of τ such that $\phi^\tau = 1/2$.

in our sample, and the half-life of market swings is in reality 69 years; two generations rather than two business cycles. These numbers seems to me an upper bound on a sensible view of the world.

Nothing dramatic happens as ϕ rises from 0.98 to 1.01. In particular, none of the statistics explode as ϕ passes through 1, so one may take any upper limit in this range without changing the conclusions dramatically. And that conclusion remains much stronger evidence against the null that returns are unpredictable.

What about higher values of ρ? From the identity $b_r = 1 - \rho\phi + b_d$, a higher ρ works just like a higher ϕ in allowing us to consider low values for both b_r and b_d. In the long-run coefficient $b^{lr}_r = b_r/(1-\rho\phi)$, a higher ρ allows us to generate a larger long-run coefficient with a lower ϕ. Of course ρ cannot exceed one, and in fact ρ must stay somewhat below one, as $\rho = 1$ implies an infinite level of the dividend yield. Still, ρ is estimated from the mean dividend yield, which it is not perfectly measured, so it's natural to ask how much a larger ρ would change the picture.

In my sample, the mean log dividend yield is -3.28, corresponding to a price-dividend ratio of $PD = e^{3.28} = 26.6$ and $\rho = PD/(1+PD) = 0.963$. The standard error of the mean log price-dividend ratio

is[10] 0.415. A one-standard error increase in the mean dividend yield, gives $PD = e^{3.28+0.41} = 34.85$ and $\rho = 0.972$—a roughly one-percentage-point increase in ρ. In simulations (not reported), changing ρ in this way has about the same effect as a one-percentage-point change in ϕ. Higher ρ also lowers our upper limits for ϕ, which may offset some power losses. For example, we imposed $\phi \geq 1/\rho \approx 1.04$ because that value would produce an infinite price-dividend ratio.

4.3 An overall number

It would be nice to present a single number, rather than a table of values that depend on assumed values for the dividend-yield autocorrelation ϕ. One way to do this is by integrating over ϕ with a prior distribution. The last row of Table 5 presents this calculation, using the unconditional likelihood of ϕ as the integrating distribution.

Figure 4 presents the likelihood function for ϕ. This is the likelihood function of an AR(1) process fit to the dividend yield, with the intercept and error variance parameters maximized out. The conditional likelihood takes the first data point as fixed. The unconditional likelihood adds the log probability of the first data point, using its unconditional density. As Figure 4 shows, the conditional and unconditional likelihoods have pretty much the same shape. The unconditional likelihood goes to zero at $\phi = 1$, which is the boundary of stationarity in levels. The maximum unconditional likelihood is only very slightly below the maximum conditional likelihood and OLS estimate of ϕ.

I repeat the simulation, but this time drawing ϕ from the unconditional likelihood plotted in Figure 4 before drawing a sample of errors ε_t^{dp} and ε_t^d. I use the unconditional likelihood in order to impose the view that dividend yields are stationary with a finite variance, $\phi < 1$, and to avoid any draws in the region $\phi > 1/\rho \approx 1.04$ in which present value formulas blow up.

The last row of Table 5 summarizes the results. The results are quite similar to the $\phi = 0.941$ case. This happens because the likelihood function is reasonably symmetric around the maximum likelihood estimate, and our statistics are not strongly nonlinear functions of ϕ. If something blew up as $\phi \to 1$, for example, then we could see an important difference between results for a fixed $\phi = 0.941$ and this calculation.

Most importantly, rather than a 23% chance of seeing a return-forecasting coefficient $b_r > \hat{b}_r$, we can reject the null based on the 1.4–1.7%

[10] I compute this standard error with a correction for serial correlation as

$$\frac{\sigma(d_t - p_t)}{\sqrt{T-1}}\sqrt{\frac{1+\phi}{1-\phi}}.$$

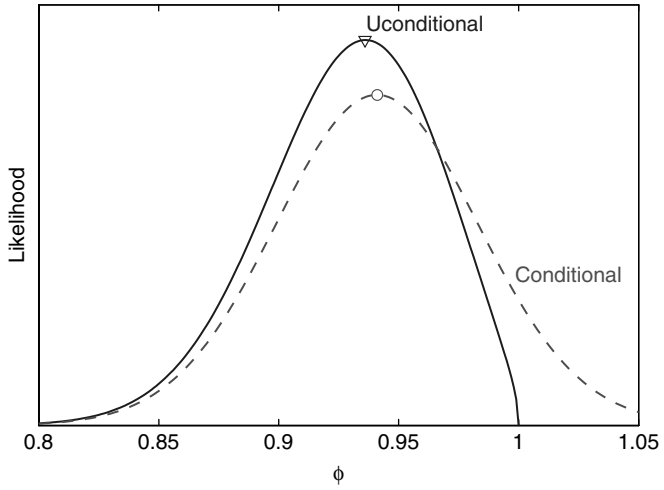

Figure 4
Likelihood function for ϕ, the autoregressive parameter for dividend yields. The likelihood is based on an autoregressive model, $d_{t+1} - p_{t+1} = a_{dp} + \phi(d_t - p_t) + \varepsilon^{dp}_{t+1}$. The intercept a_{dp} and innovation variance $\sigma^2(\varepsilon^{dp})$ are maximized out.

chance of seeing the dividend growth b_d or long-run regression coefficients b^{lr} greater than their sample values. As usual, excess returns give even stronger rejections, with probability values of 0.6–1.1%. (Lewellen (2004) presents a similar and more formally Bayesian calculation that also delivers small probability values.)

5. Power in Long-run Regression Coefficients?

I find much greater ability to reject the unforecastable-return null in the long-horizon coefficient $b^{lr}_r = b_r/(1 - \rho\phi)$ than in the one-year coefficient b_r. A large literature, most recently exemplified by Boudoukh, Richardson, and Whitelaw (2006), finds no power advantage in long-horizon regressions.[11] How can we reconcile these findings?

5.1 Long-horizon regressions compared

There are three main differences between the coefficients that I have calculated and typical long-horizon regressions. First, b^{lr}_r is an *infinite*-horizon coefficient. It corresponds to the regression of $\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$ on $d_t - p_t$. Most studies examine instead the power of *finite*-horizon

[11] Boudoukh, Richardson, and Whitelaw focus much of their discussion on the high correlation of short-horizon and long-horizon regression coefficients. This is an interesting but tangential point. Short- and long-horizon coefficients are not *perfectly* correlated, so long-horizon regressions add *some* information. The only issue is how *much* information they add, and whether the extra information overcomes additional small sample biases—whether long-run regressions have more or less power than one-year regressions.

regression coefficients, $\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$ on $d_t - p_t$. Second, I calculate $b_r^{lr} = b_r/(1-\rho\phi)$ as an *implication* of the first-order VAR coefficients. Most studies examine instead *direct* regression coefficients—they actually construct $\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$ and explicitly run it on $d_t - p_t$. Third, b_r^{lr} is *weighted* by ρ where most studies examine instead *unweighted* returns, that is, they run $\sum_{j=1}^{k} r_{t+j}$ on $d_t - p_t$.

Table 6 investigates which of these three differences in technique accounts for the difference in results. The first row of Table 6 presents the familiar one-year return forecast. We see the usual sample coefficient of $\hat{b}_r = 0.10$, with 22% probability value of observing a larger coefficient under the null.

Increasing to a 5-year horizon, we see that the sample regression coefficient rises substantially, to 0.35–0.43, depending on which method one uses. In the direct estimates, the probability values get slightly worse, rising to 28–29% of seeing a larger value. I therefore confirm findings such as those of Boudoukh, Richardson, and Whitelaw's (2006) that directly-estimated 5-year regressions have slightly worse power than 1-year regressions. The *implied* 5-year regression coefficients do a little bit better, with probability values declining to 16-18%. The improvement is small, however, and looking only at 1-year to 5-year horizons, one might well conclude that long-horizon regressions have at best very little additional power.

As we increase the horizon, the probability values decrease substantially. The implied long-horizon regression coefficients reach 5% probability values at horizons between 15 and 20 years under $\phi = 0.94$, and there are still important gains in power going past the 20-year horizon. Excess returns, as usual, show stronger results (not shown).

Table 6 shows that the central question is the horizon: conventional 5-year and even 10-year horizons do not go far enough out to see the power advantages of long horizon regressions. All of the methods show much better power at 20-year horizons than at 1 year horizons.

To understand why long horizons help and why we need such long horizons, consider two-year and three-year return regressions

$$r_{t+1} + \rho r_{t+2} = a_r^{(2)} + b_r^{(2)} x_t + \delta_{t+2}$$
$$r_{t+1} + \rho r_{t+2} + \rho^2 r_{t+3} = a_r^{(3)} + b_r^{(3)} x_t + \delta_{t+3}$$

The coefficients are (in population, or in the indirect estimate)

$$b_r^{(3)} = b_r(1 + \rho\phi) \tag{17}$$
$$b_r^{(3)} = b_r(1 + \rho\phi + \rho^2\phi^2) \tag{18}$$

Table 6
Long-horizon forecasting regressions

	Weighted						Unweighted					
	$\sum_{j=1}^{k} \rho^{j-1} r_{t+j} = a + b_r^{(k)}(d_t - p_t) + \delta_{t+k}$						$\sum_{j=1}^{k} r_{t+j} = a + b_r^{(k)}(d_t - p_t) + \delta_{t+k}$					
	Direct			Implied			Direct			Implied		
k	coeff. $\hat{b}_r^{(k)}$	p-value, 0.94	$\phi =$ 0.99	coeff. $\hat{b}_r^{(k)}$	p-value, 0.94	$\phi =$ 0.99	coeff. $\hat{b}_r^{(k)}$	p-value, 0.94	$\phi =$ 0.99	coeff. $\hat{b}_r^{(k)}$	p-value, 0.94	$\phi =$ 0.99
1	0.10	22	22	0.10	22	22	0.10	22	22	0.10	22	22
5	0.35	28	29	0.40	17	19	0.37	29	29	0.43	16	18
10	0.80	16	16	0.65	10	15	0.92	16	16	1.02	9.0	14
15	1.38	4.4	4.7	0.80	6.2	12	1.68	4.8	5.0	1.26	4.3	10
20	1.49	4.7	5.2	0.89	4.1	9.8	1.78	7.8	8.3	1.41	2.2	7.6
∞				1.04	1.8	7.3				1.64	0.5	8.9

In each case $\hat{b}_r^{(k)}$ gives the point estimate in the data. The column labeled "p-value" gives the percent probability value—the percentage of simulations in which the long-horizon regression coefficient $\hat{b}_r^{(k)}$ exceeded the sample value $\hat{b}_r^{(k)}$. $\phi = 0.94, 0.99$ indicates the assumed dividend-yield autocorrelation ϕ in the null hypothesis. "Direct" constructs long-horizon returns and explicitly runs them on dividend yields. "Implied" calculates the indicated long-horizon regression coefficient from one-period regression coefficients. For example, the 5-year weighted implied coefficient is calculated as $b_r^{(5)} = \sum_{j=1}^{5} \rho^{j-1}\phi^{j-1} b_r = (1 - \rho^5\phi^5)/(1 - \rho\phi) b_r$.

Thus, coefficients rise with horizon mechanically as a result of one-period forecastability and the autocorrelation ϕ of the forecasting variable (Campbell and Shiller, (1988).

These regressions exploit the negative correlation of b_r and ϕ to increase power, just as do the infinite-horizon regressions studied above. Because large b_r tend to come with small ϕ it is harder for the null to produce large long-horizon coefficients than it is for the null to produce large one-year coefficients.

The trouble is that this mechanism is not *quantitatively* strong for 2-year, 3-year, or even 5-year horizons, as their particular combinations of b_r and ϕ do not stress ϕ enough. To display this fact, Figure 5 plots again the joint distribution of $\{b_r, \phi\}$, together with lines that show rejection regions for long-horizon regression coefficients. The 1-year horizon line is vertical as before. The 5-year horizon line is the set of $\{b_r, \phi\}$ points at which $b_r(1 + \rho\phi + \cdots + \rho^4\phi^4)$ equals its value in the data. Points above and to the right of this line are simulations in which the five-year (unweighted, implied) regression coefficient is larger than the sample value of this coefficient. As the figure shows, this line excludes a few additional points, but not many, which is why the associated probability value declines from 22% to only 17%. The $k = \infty$ line is the set of (b_r, ϕ) points at which $b_r/(1 - \rho\phi)$ equals its value in the data; points above and to the right of this line are simulations in which the infinite-horizon long-run regression coefficients studied above are greater than their sample values.

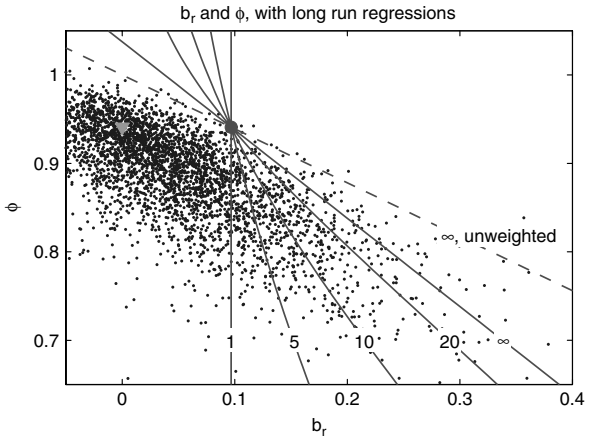

Figure 5
Joint distribution of b_r and ϕ estimates, together with regions implied by long-run regressions. The lines give the rejection regions implied by long-horizon return regressions at the indicated horizon. For example, the points above and to the right of the line marked "5" are simulations in which coefficient $b_r^{(5)} = (1 + \rho\phi + \cdots + \rho^4\phi^4)b_r$ are larger than its value in the data. The dashed line marked "∞, unweighted" plots the line where $b_r/(1 - \phi)$ equals its value in the data, corresponding to the infinite-horizon unweighted regression.

As Figure 5 shows, longer horizon regressions give more and more weight to ϕ, and therefore generate smaller probability values. The figure shows why one must consider such long horizons to exclude many points and obtain a powerful test.

5.2 Implications and nonimplications

Table 6 shows that the distinction between weighted and unweighted long-horizon regressions makes almost no difference. Since $\rho = 0.96$ is close to one, that result is not surprising. (Yes, the implied infinite-horizon unweighted regression coefficient makes sense. Even though the left-hand variable and its variance explode, the coefficient converges to the finite value $b_r/(1-\phi)$.)

Table 6 shows some interesting differences between *implied* and *direct* estimates. In most cases, the direct estimates give less power against the null than the implied estimates. In a few cases, the direct estimates seem better than the indirect estimates, but this is a result of larger directly-estimated coefficients in our sample. If we set an even bar, then the direct coefficients show lower power than implied estimates in every case. It is interesting, however, that the direct estimates are not much worse, even at very long horizons. (Boudoukh and Richardson (1994) also find little difference between methods given the horizon.)

Table 6 does not say much *in general* about whether it is better to compute long-horizon statistics directly, "nonparametrically," or whether one should compute them by calculating implied coefficients from a low-order model. The latter strategy works better here, but that fact is hardly surprising since the data are generated by the same AR(1) I use to calculate implied long-horizon statistics. In some other circumstances, direct estimates of long-horizon statistics can pick up low-frequency behavior that even well-fit short-horizon models fail to capture. Cochrane (1988) is a good example. In other circumstances, the short-order model is a good approximation, small sample biases are not too severe, and especially when cointegrating vectors are present and one variable (price) summarizes conditional expectations, implied long-horizon statistics can perform better. I present some evidence below that the VAR(1) is a good fit for the data on returns and dividend yields we are studying here, suggesting the latter conclusion, but of course that conclusion is also limited to this data set.

The statistical power of "long-horizon regressions" in this analysis really has nothing to do with the horizon *per se*. The test is powerful in these simulations because it forms a joint region of short-horizon coefficients b_r and ϕ that has good power. The *one*-year horizon b_d test gives almost exactly the same results. The economic interpretation, and the economic motivation for defining "distance from the null" in terms of long-horizon statistics involves horizon, of course, but says nothing about whether

one should estimate long-horizon statistics directly or indirectly as a consequence of a fitted VAR(1).

In sum, Table 6 shows that long-horizon regression coefficients have the potential for substantially greater power to reject the null of unforecastable returns. However, one must look a good deal past the 5-year horizon to see much of that power. Furthermore, direct estimates of long-horizon coefficients introduce additional uncertainty, and that uncertainty can be large enough to obscure the greater power for some horizons, null hypotheses, and sample sizes. This summary view reconciles the analytical and simulation results on both sides, including Boudoukh, Richardson, and Whitelaw (2006) simulations showing low power, and Campbell (2001) and Valkanov (2003) analyses showing good power in large samples and at very long horizons.

6. Out-of-Sample R^2

Goyal and Welch (2005) show in a comprehensive study that the dividend yield and many other regressors thought to forecast returns do not do so out of sample. They compare two return-forecasting strategies. First, run a regression $r_{t+1} = a + bx_t + \varepsilon_{t+1}$ from time 1 to time τ, and use $\hat{a} + \hat{b}x_\tau$ to forecast the return at time $\tau + 1$. Second, compute the sample mean return from time 1 to time τ, and use that sample mean to forecast the return at time $\tau + 1$. Goyal and Welch compare the mean squared error of the two strategies, and find that the "out-of-sample" mean squared error is often larger for the return forecast than for the sample mean.

Campbell and Thompson (2005) give a partial rejoinder. The heart of the Goyal–Welch low R^2 is that the coefficients a and b are poorly estimated in "short" samples. In particular, sample estimates often give conditional expected excess returns less than zero, and recommend a short position. Campbell and Thompson rule out such "implausible" estimates, and find out-of-sample R^2 that are a bit better than the unconditional mean. Goyal and Welch respond that the out-of-sample R^2 are still small.

6.1 Out of sample R^2 as a test

Does this result mean that "returns are really not forecastable?" If all dividend-yield variation were really due to *return* forecasts, how often would we see Goyal–Welch results?

To answer this question, I set up the null analogous to (7) in which returns *are* forecastable, dividend growth *is not* forecastable, and all dividend-yield variation comes from time-varying expected returns,

$$\begin{bmatrix} d_{t+1} - p_{t+1} \\ \Delta d_{t+1} \\ r_{t+1} \end{bmatrix} = \begin{bmatrix} \phi \\ 0 \\ \rho\phi - 1 \end{bmatrix} (d_t - p_t) + \begin{bmatrix} \varepsilon^{dp}_{t+1} \\ \varepsilon^{d}_{t+1} \\ \varepsilon^{d}_{t+1} - \rho\varepsilon^{dp}_{t+1} \end{bmatrix}$$

Mechanically, this is the same as the previous VAR (7) except $b_r = \rho\phi - 1$ and $b_d = 0$ rather than the other way around. It can be derived from an analogous "structural" model that expected returns follow an AR(1), $E_t(r_{t+1}) = x_t = \phi x_{t-1} - \delta_t^x$, dividend growth is unforecastable $\Delta d_{t+1} = \varepsilon_{t+1}^d$, and dividend yields are generated from the present value identity (9) (Cochrane (2004) ch. 20). This null is very close to the sample estimates of Table 2, but turns off the slight dividend predictability in the "wrong" direction.

I simulate artificial data from this null as before. I start with $\phi = 0.941$, which implies a return-forecasting coefficient $b_r = 1 - \rho\phi \approx 0.1$. I also consider $\phi = 0.99$ to address small-sample bias worries, which implies a lower value of $b_r = 1 - \rho\phi \approx 0.05$. In each sample, I calculate the Goyal-Welch statistic: starting in year 20, I compute the difference between root mean squared error from the sample-mean forecast and from the fitted dividend-yield forecast. A larger positive value for this statistic is good for return forecastability; a larger negative value implies that the sample mean is winning.

Figure 6 shows the distribution of this statistic across simulations. In the data, marked by the vertical "Data" line, the statistic is negative; the sample mean is a better out-of-sample forecaster than the dividend yield, as Goyal and Welch (2005) find. However, 30–40% of the draws show even worse results than our sample. In fact, the *mean* of the Goyal-Welch statistic is negative, and only about 20% of the draws show a *positive* value. Even though under this null *all* dividend-price variation is due to time-varying expected returns by construction, it is *unusual* for dividend-yield forecasting to actually work better than the sample mean in this

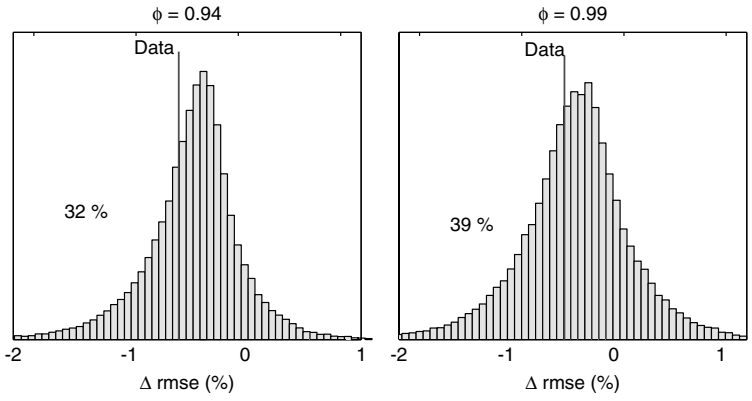

Figure 6
Distribution of the Goyal–Welch statistic under the null that returns are forecastable and dividend growth is not forecastable. The statistic is the root mean squared error from using the sample mean return from time 1 to time t to forecast returns at $t + 1$, less the root mean squared error from using a dividend yield regression from time 1 to time t to forecast returns at time $t + 1$.

out-of-sample experiment. $\phi = 0.99$ makes it even more likely for sample means to win the Goyal–Welch race.

Thus, the Goyal–Welch statistic does not reject the time-varying expected return null. Poor out-of-sample R^2 is exactly what we expect given the persistence of the dividend yield, and the relatively "short" samples we have for estimating the relation between dividend yields and returns.

6.2 Reconciliation

Both views are right, if correctly interpreted. Goyal and Welch (2005) message is that regressions on dividend yields and similarly persistent variables are not likely to be useful in forming real-time forecasts or market-timing portfolios, given the difficulty of accurately estimating the coefficients in our "short" data sample. This conclusion echoes Kandel and Stambaugh (1996) and Barberis (2000), who show in a Bayesian setting that uncertainty about the parameter b_r means that market-timing portfolios should use a much lower parameter, shading the portfolio advice well back toward the use of the sample mean. How these more sophisticated calculations perform out of sample, extending Campbell and Thompson (2005) idea, is an interesting open question.

However, poor out-of-sample R^2 does not reject the null hypothesis that returns are predictable. Out-of-sample R^2 is not a new and powerful test statistic that gives stronger evidence about return forecastability than the regression coefficients or other standard hypothesis tests. One can simultaneously hold the view that returns are predictable, or more accurately that the bulk of price-dividend ratio movements reflect return forecasts rather than dividend-growth forecasts, *and* believe that such forecasts are not very useful for out-of-sample forecasting and portfolio advice, given uncertainties about the coefficients in our data sets.

7. What about. . .

7.1 Repurchases, specification, and additional variables

What about the fact that firms seem to smooth dividends, many firms do not pay dividends, dividend payments are declining in favor of repurchases, and dividend behavior may shift over time?

Dividends as measured by CRSP capture all payments to investors, including cash mergers, liquidations, and so forth, as well as actual dividends. If a firm repurchases *all* of its shares, CRSP records this event as a dividend payment. If a firm repurchases *some* of its shares, an investor may choose to hold his shares, and the CRSP dividend series captures the eventual payments he receives. Thus, there is nothing wrong in an accounting sense with using the CRSP dividends series. The price really is the present value of these dividends. These worries are therefore really

statistical rather than conceptual, whether the VAR(1) model adequately captures the time-series data.

Dividends that are frequently zero, and consist of infrequent large lumps would clearly not fit a linear VAR(1). For this reason, I limit attention to the *market* price-dividend ratio, so that this lumpiness is averaged out across firms. Shifts in dividend behavior also do not mean there is anything conceptually wrong with a dividend-yield regression. We can understand this regression as a characterization of the unconditional moments, averaging over such shifts. Dividend-smoothing can only do limited damage, since earnings must eventually be paid out as dividends.

Although there is nothing *wrong* with using the dividend yield to forecast returns, one *can* use variables that adjust for payout policies or stochastic shifts in dividend behavior, as we can use any other variable in the time-t information set, to forecast returns. Such forecasts can give even stronger evidence of return predictability, since the payout yield is "more stationary" than the dividend yield (Boudoukh Michaely, Richardson, and Roberts (2007)). For example, Boudoukh, Richardson, and Whitelaw (2006) report a 5.16% R^2 using the dividend yield, but 8.7, 7.7, and 23.4% R^2 using various measures of the payout yield[12] (i.e., including repurchases). Lettau and Van Nieuwerburgh (2006) show that allowing for a shift in dividend payout behavior also raises the forecast R^2 substantially. Price/earnings, book/market and similar variables forecast returns, and one can understand these as sensible variations that account for the behavior of measured dividends.

More generally, a large number of additional variables seem to forecast returns; for example, see the summary in Goyal and Welch (2005). Although we cannot fish across variables for t-statistics any more than we can fish across horizons, once we agree that dividend yields forecast returns, additional variables can only add to the evidence for return forecastability.

Again, the point of the dividend-yield specification in this paper is not to find the best return-forecasting specification; the point is to show how return-forecast statistics work in the simplest possible specification. Everything I do here can only give even stronger evidence with these additional variables or more complex specifications.

Additional variables can also help to predict dividend growth. Ribeiro (2004), and Lettau and Ludvigson (2005) give examples. This fact does *not* imply that returns must become less predictable. The identities that dividend-growth predictability and return predictability add up apply only to forecasts based on the dividend yield. Other variables can raise the predictability of both dividend growth *and* of returns. To be specific, consider any set of forecasting variables Ω_t that includes the dividend

[12] The stunning 23.4 R^2 value comes from one large outlier in the early 1930s.

yield. The return identity (4) implies

$$d_t - p_t = E(r_{t+1}|\Omega_t) - E(\Delta d_{t+1}|\Omega_t) + \rho E(d_{t+1} - p_{t+1}|\Omega_t) \tag{19}$$

and the present value identity (9) is

$$d_t - p_t = E_t\left(\sum_{j=1}^{\infty} \rho^{j-1}\Delta r_{t+j}\middle|\Omega_t\right) - E_t\left(\sum_{j=1}^{\infty} \rho^{j-1}\Delta d_{t+j}\middle|\Omega_t\right) \tag{20}$$

By Equation (19), if any variable helps to forecast one-period dividend growth, it *must* help to forecast returns, or help to forecast future dividend yields. By Equation (20), if any variable helps to forecast long-run dividend growth, it must also *help* to forecast long-run returns (Lettau and Ludvigson (2005)). Again, considering more variables can only make the evidence for return predictability stronger, even if those variables also help to forecast dividends.

7.2 Direct long-horizon estimates and hidden dividend growth

Concerns about dividend smoothing, repurchases, and so on mean that, despite aggregation, prices might move today on news of dividends several years in the future, news not seen in next year's dividend. The 1-year VAR would miss this pattern. We can address this worry by examining direct forecasts of long-horizon returns and dividend growth, regressions of the form

$$\sum_{j=1}^{k} \rho^{j-1}\Delta d_{t+j} = a_d^{(k)} + b_d^{(k)}(d_t - p_t) + \varepsilon_{t+k}^d$$

$$\sum_{j=1}^{k} \rho^{j-1} r_{t+j} = a_r^{(k)} + b_r^{(k)}(d_t - p_t) + \varepsilon_{t+k}^r$$

As with their infinite-horizon counterparts in Equations (10)–(12), these regression coefficients amount to a variance decomposition for dividend yields. They obey the identity

$$1 = b_r^{(k)} - b_d^{(k)} + b_{dp}^{(k+1)} \tag{21}$$

where $b_{dp}^{(k+1)}$ is the regression coefficient of $\rho^{k+1}(d_{t+k+1} - p_{t+k+1})$ on $d_t - p_t$. We can interpret these regression coefficients as estimates of what fraction of dividend-yield variance is due to k-period dividend-growth forecasts, what fraction is due to k-period return forecasts, and what fraction is due to k+1-period forecasts of future dividend yields. As $k \to \infty$ and for $\phi < 1/\rho$ the last term vanishes and we recover the identity $b_r^{lr} - b_d^{lr} = 1$ studied in Section 2.2.

Figure 7 presents direct estimates of long-horizon regression coefficients in Equation (21) as a function of k. I do not calculate the last, future price-dividend ratio term because its value is implied by the other two terms.

In the top panel of Figure 7, we see that dividend-growth forecasts explain small fractions of dividend yield variance at all horizons. The triangles in Figure 7 are direct regressions, $\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}$ on $d_t - p_t$. The rise in these estimates means that long-run dividend growth moves in the wrong direction, explaining negative fractions of dividend-yield variation. The circles in Figure 7 sum individual regression coefficients, $\sum_{j=1}^{k} \rho^{j-1} \beta(\Delta d_{t+j}, d_t - p_t)$. This estimate differs from the last one only because it uses more data points. For example, the first year $\beta(\Delta d_{t+1}, d_t - p_t)$ in the 10-year horizon return is estimated using $T - 1$ data points, not $T - 10$ data points of the direct (triangle) estimate. Here we at least see the "right," negative, sign, though the magnitudes are still trivial.

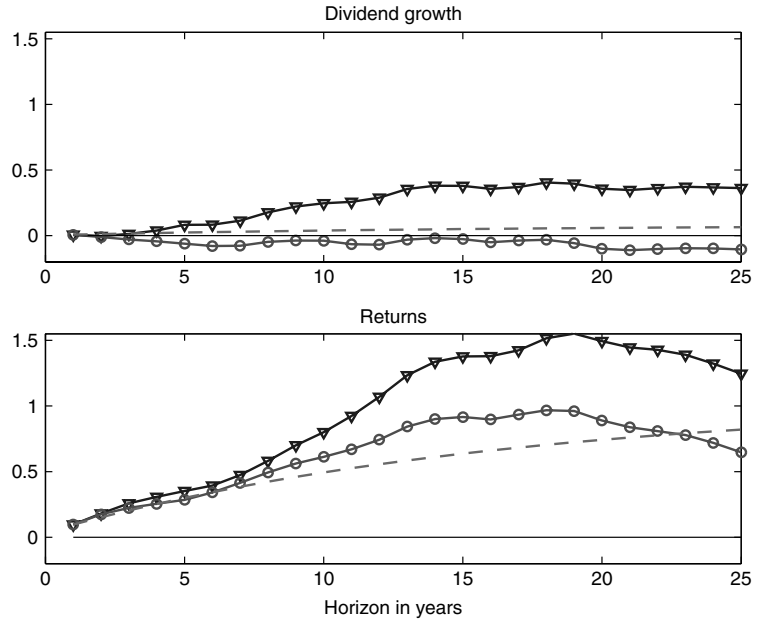

Figure 7
Regression forecasts of discounted dividend growth $\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}$ (top) and returns $\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$ (bottom) on the log dividend yield $d_t - p_t$, as a function of the horizon k. Triangles are direct estimates: I form the weighted long-horizon returns and run them on dividend yields—for example, $\beta\left(\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}, d_t - p_t\right)$. Circles sum individual estimates: I run dividend growth and return at year $t + j$ on the dividend yield at t and then sum up the coefficients—for example, $\sum_{j=1}^{k} \rho^{j-1} \beta\left(\Delta d_{t+j}, d_t - p_t\right)$. The dashed lines are the long-run coefficients implied by the VAR—for example, $\sum_{j=1}^{k} \rho^{j-1} \phi^{j-1} b_d$.

By contrast, the return forecasts account for essentially all dividend-yield volatility once one looks out past 10-years. The long-horizon return regression coefficients approach and even exceed one. This, with a negative sign, is what long-horizon dividend forecasts *should* look like if we are to hope that changing expectations of dividend growth explain price variation. They do not come close, even in these direct estimates that allow for unstructured temporal correlations and long-delayed dividend payments.

Despite the battering return forecasts measured by b_r took in the 1990s, cutting return coefficients b_r almost in half, both these direct and the above indirect $b_r^{lr} = b_r/(1-\rho\phi)$ long-horizon estimates of Table 4 are very little changed since Cochrane (1992). The longer sample has a lower b_r, but a larger ϕ, so $b_r/(1-\rho\phi)$ is still just about exactly one.

The dashed lines in Figure 7 present the long-run coefficients implied by the VAR, $\sum_{j=1}^{k} \rho^{j-1}\phi^{j-1}b_r = b_r\left(1-\rho^k\phi^k\right)/(1-\rho\phi)$ and similarly for dividend growth, to give a visual sense of how well the VAR fits the direct estimates. The point estimates of the long-run regressions show slightly *stronger* return forecastability than the values implied by the VAR, and dividend growth that goes even more in the "wrong" positive direction, though the differences are far from statistically significant. Though low-order VAR systems do not always capture long-run dynamics well (for example, Cochrane (1988)), they seem to do so in this data set.

To keep the graph from getting too cluttered, I omit standard error bars from Figure 7. The best set of asymptotic standard errors I calculated gives the return-forecast t-statistic of about two at all horizons. The dividend-growth forecasts are completely insignificant.

7.3 Bias in forecast estimates

Table 7 presents the means of the estimated coefficients under the null hypothesis. As we expect for a near-unit-root process, the dividend-yield autocorrelation estimate ϕ is biased downward. The return forecast coefficient b_r is biased upward. The bias of approximately 0.05 accounts for roughly half of the sample estimate $\hat{b}_r \approx 0.10$. This bias results from the strong negative correlation between return and dividend-yield errors and

Table 7
Means of estimated parameters

		b_r	b_d	ϕ	b_r^{lr}	b_d^{lr}
$\phi = 0.941$	Null	0	−0.093	0.941	0	−1
	Mean	0.049	−0.097	0.886	0.24	−0.77
$\phi = 0.99$	Null	0	−0.046	0.990	0	−1
	Mean	0.057	−0.050	0.926	0.43	−0.57

Means are taken over 50,000 simulations of the Monte Carlo described in Table 2

the consequent strong negative correlation between return and dividend-yield coefficients.

The dividend-growth coefficient b_d is not biased. There is no particular correlation between the b_d and ϕ estimates, deriving from the nearly zero correlation between dividend-growth and dividend-yield shocks. Thus, the dividend-growth forecast does not inherit any near-unit-root issues from the strong autocorrelation of the right-hand variable. This observation should give a little more comfort to the result that $b_d \approx 0$ is a good characterization of the data. The long-horizon return coefficient b_r^{lr} is biased up, and more so for higher values of ϕ. Correspondingly, the long-horizon dividend-growth coefficient b_d^{lr} is biased up as well. However, the strong rejections of $b_r^{lr} = 0$ or equivalently $b_d^{lr} = -1$ mean that we can still distinguish the biased null value $b_r^{lr} = 0.24 - 0.43$ from the sample value $\hat{b}_r^{lr} \approx 1$.

The probability values documented above do not ignore the fact that coefficients are biased in small samples. They show that we can reject the null hypothesis despite the biases, which are fully accounted for in the small-sample test statistics documented above.

Table 7 documents biases under the unpredictable-return null, so it does not really answer the question, "If we want to adjust for small-sample biases, what point estimate should be our best guess of the world?" (The question presumes we want an unbiased estimate, not, for example, the maximum likelihood estimate, which we already have.) We can, however, read a rough answer to this question from Table 7. The size of the biases is about the same under the null presented in Table 7, $b_r \approx 0, b_d \approx -0.05, \phi \approx 0.99$, as it is under the alternative $b_r \approx 0.05$, $b_d \approx 0$, $\phi \approx 0.99$. Therefore, this latter set of parameters will produce, on average, estimates close to those observed in our sample, $\hat{b}_r = -0.10$, $\hat{b}_d = 0, \hat{\phi} \approx 0.94$.

Most importantly, this set of parameters implies $b_r^{lr} \approx 0.05/(1 - 0.99 \times 0.96) \approx 1.0$ and $b_d^{lr} = 0$. Thus, the "bias corrected" point estimates keep the view that all dividend-yield volatility comes from return forecasts $b_r^{lr} \approx 1$ and $b_d^{lr} \approx 0$ intact. They imply that more of the long-run forecastability comes from dividend-yield autocorrelation ϕ (the build-up of coefficients with horizon) and less from one-period return forecastability b_r, but the combination is unchanged.

8. Conclusion

If returns really are *not* forecastable, then dividend growth must *be* forecastable in order to generate the observed variation in dividend-price ratios. We should see that forecastability. Yet, even looking 25 years out, there is not a shred of evidence that high market price-dividend ratios are associated with higher subsequent dividend growth (Figure 7). Even

if we convince ourselves that the return-forecasting evidence crystallized in Fama and French (1988) regressions is statistically insignificant, we still leave unanswered the challenge crystallized by Shiller (1981) volatility tests: If not dividend growth or expected returns, what *does* move prices?

Setting up a null in which returns are not forecastable, and changes in expected dividend growth explain the variation of dividend yields, I can check both dividend-growth and return forecastability. I find that the *absence* of dividend-growth forecastability in our data provides much stronger evidence against this null than does the *presence* of one-year return forecastability, with probability values in the 1–2% range rather than in the 20% range.

The long-run coefficients best capture these observations in a single number, and tie them to modern volatility tests. The point estimates are squarely in the bull's eye that all variation in market price-dividend ratios is accounted for by time-varying expected returns, and none by time-varying dividend-growth forecasts. Tests based on these long-run coefficients also give 1–2% rejections.

Both long-run regressions and dividend-growth regressions exploit the negative correlation between return-forecast and dividend-yield autocorrelation coefficients, and prior knowledge that dividend-yield autocorrelation cannot be too high, to produce more powerful tests. Large long-run return forecasts result from large short-run return forecasts together with large autocorrelations, but a null with a limited autocorrelation and negative correlation between return-forecast and autocorrelation estimates produces a large long-run return forecast only infrequently.

Excess return forecastability is not a comforting result. Our lives would be so much easier if we could trace price movements back to visible news about dividends or cashflows. Failing that, it would be nice if high prices forecast dividend growth, so we could think agents see cashflow information that we do not see. Failing that, it would be lovely if high prices were associated with low interest rates or other observable movements in discount factors. Failing that, perhaps time-varying expected excess returns that generate price variation could be associated with more easily measurable time-varying standard deviations, so the market moves up and down a mean-variance frontier with constant Sharpe ratio. Alas, the evidence so far seems to be that most aggregate price/dividend variation can be explained only by rather nebulous variation in Sharpe ratios. But that is where the data have forced us, and they still do. The only good piece of news is that observed return forecastability *does* seem to be just enough to account for the volatility of price dividend ratios. If both return and dividend-growth forecast coefficients were small, we would be forced to conclude that prices follow a "bubble" process, moving only on news (or, frankly, opinion) of their own future values.

The implications of excess return forecastability have a reach throughout finance and are only beginning to be explored. The literature has focused on portfolio theory—the possibility that a few investors can benefit by market-timing portfolio rules. Even here, the signals are slow moving, really affecting the static portfolio choices of different generations rather than dynamic portfolio choices of short-run investors, and parameter uncertainty greatly reduces the potential benefits. Most seriously, all portfolio theory calculations face a classic catch-22: if there is a substantial number of agents who, on net, should take the advice, the phenomenon will disappear.

Portfolio calculations are just the tip of the iceberg, however. If expected excess returns really do vary by as much as their average levels, and if all market price-dividend ratio variation comes from varying expected returns and none from varying expected growth in dividends or earnings, much of the rest of finance still needs to be rewritten. For example, Mertonian state variables, long a theoretical curiosity but relegated to the back shelf by an empirical view that investment opportunities are roughly constant, should in fact be at center stage of cross-sectional asset pricing. For example, much of the beta of a stock or portfolio reflects covariation between firm and factor (e.g., market) discount rates rather than reflecting the covariation between firm and market cash flows. For example, a change in prices driven by discount rate changes does not change how close the firm is to bankruptcy, justifying strikingly inertial capital structures as documented by Welch (2004). For example, standard cost-of-capital calculations featuring the CAPM and a steady 6% market premium need to be rewritten, at least recognizing the dramatic variation of the initial premium, and more deeply recognizing likely changes in that premium over the lifespan of a project and the multiple pricing factors that predictability implies.

References

Ang, A., and G. Bekaert. 2007. Stock Return Predictability: Is It There? *Review of Financial Studies* 20:651–707.

Barberis, N. 2000. Investing for the Long Run when Returns Are Predictable. *Journal of Finance* 55:225–64.

Boudoukh, J., R. Michaely, M. Richardson, and M. Roberts. 2007. On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing. *Journal of Finance* 62:877–916.

Boudoukh J., and M. Richardson. 1994. The Statistics of Long-Horizon Regressions. *Mathematical Finance* 4:103–20.

Boudoukh J., M. Richardson, and R. F. Whitelaw. 2006. The Myth of Long-Horizon Predictability. *Review of Financial Studies*, 10.1093/rfs/hhl042.

Campbell, J. Y. 2001. Why Long Horizons? A Study of Power Against Persistent Alternatives. *Journal of Empirical Finance* 8:459–91.

Campbell, J. Y., and R. J. Shiller. 1988. The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies* 1:195–228.

Campbell, J. Y., and S. Thompson. 2005. Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average? NBER Working Paper No. 11468, *Review of Financial Studies* forthcoming.

Campbell, J. Y., and M. Yogo. 2006. Efficient Tests of Stock Return Predictability. *Journal of Financial Economics* 81:27–60.

Cochrane, J. H. 1988. How Big Is the Random Walk in GNP? *Journal of Political Economy* 96:893–920.

Cochrane, J. H. 1991. Volatility Tests and Efficient Markets: Review Essay. *Journal of Monetary Economics* 27:463–85.

Cochrane, J. H. 1992. Explaining the Variance of Price-Dividend Ratios. *Review of Financial Studies* 5:243–80.

Cochrane, J. H. 2004. *Asset Pricing*, (rev. ed.), Princeton, NJ: Princeton University Press.

Craine, R. 1993. Rational Bubbles: A Test. *Journal of Economic Dynamics and Control* 17:829–46.

Fama, E. F., and K. R. French. 1988. Dividend Yields and Expected Stock Returns. *Journal of Financial Economics* 22:3–25.

Goetzmann, W. N., and P. Jorion. 1993. Testing the Predictive Power of Dividend Yields. *Journal of Finance* 48:663–79.

Goyal, A., and I. Welch. 2003. Predicting the Equity Premium with Dividend Ratios. *Management Science* 49:639–54.

Goyal, A., and I. Welch. 2007. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, 10.1093/rfs/hhm014.

Kandel, S., and R. F. Stambaugh. 1996. On the Predictability of Stock Returns: An Asset-Allocation Perspective. *Journal of Finance* 51:385–424.

Keim, D. B., and R. F. Stambaugh. 1986. Predicting Returns in the Stock and Bond Markets. *Journal of Financial Economics* 17:357–90.

Kothari, S. P., and J. Shanken. 1997. Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis. *Journal of Financial Economics* 44:169–203.

LeRoy, S. F., and R. D. Porter. 1981. The Present-Value Relation: Tests Based on Implied Variance Bounds. *Econometrica* 49:555–74.

Lettau, M., and S. Ludvigson. 2005. Expected Returns and Expected Dividend Growth. *Journal of Financial Economics* 76:583–626.

Lettau, M., and S. Van Nieuwerburgh. 2006. Reconciling the Return Predictability Evidence. *Review of Financial Studies* forthcoming.

Lewellen, J. 2004. Predicting Returns with Financial Ratios. *Journal of Financial Economics* 74:209–35.

Mankiw, N. G., and M. Shapiro. 1986. Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models. *Economic Letters* 20:139–45.

Nelson, C. R., and M. J. Kim. 1993. Predictable Stock Returns: The Role of Small Sample Bias. *Journal of Finance* 48:641–61.

Paye, B., and A. Timmermann. 2003. Instability of Return Prediction Models. Manuscript, University of California at San Diego.

Ribeiro, R. 2004. Predictable Dividends and Returns: Identifying the Effect of Future Dividends on Stock Prices. Manuscript, University of Pennsylvania.

Rozeff, M. S. 1984. Dividend Yields Are Equity Risk Premiums. *Journal of Portfolio Management* 11:68–75.

Shiller, R. J. 1981. Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends? *American Economic Review* 71:421–36.

Shiller, R. J. 1984. Stock Prices and Social Dynamics. *Brookings Papers on Economic Activity* 2:457–98.

Stambaugh, R. F. 1986. Bias in Regressions with Lagged Stochastic Regressors. Manuscript, University of Chicago.

Stambaugh, R. F. 1999. Predictive Regressions. *Journal of Financial Economics* 54:375–421.

Torous, W., R. Valkanov, and S. Yan. 2004. On Predicting Stock Returns With Nearly Integrated Explanatory Variables. *Journal of Business* 77:937–66.

Valkanov, R. 2003. Long-Horizon Regressions: Theoretical Results and Applications. *Journal of Financial Economics* 68(2):201–32.

Welch, I. 2004. Capital Structure and Stock Returns. *Journal of Political Economy* 112:106–31.