NBER WORKING PAPER SERIES

THE DOG THAT DID NOT BARK:
A DEFENSE OF RETURN PREDICTABILITY

John H. Cochrane

Working Paper 12026
http://www.nber.org/papers/w12026

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2006

University of Chicago GSB and NBER. Address: Graduate School of Business, 5807 S. Woodlawn, Chicago IL 60637, 773 702 3059, john.cochrane@chicagogsb.edu. 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 and Ivo Welch for very helpful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

The Dog That Did Not Bark: A Defense of Return Predictability
John H. Cochrane
NBER Working Paper No. 12026
February 2006, Revised June 2006
JEL No. G0, G1

ABSTRACT

To question the statistical significance of return predictability, we cannot specify a null that simply turns off that predictability, leaving dividend growth predictability at its essentially zero sample value. If neither returns nor dividend growth are predictable, then the dividend-price ratio is a constant. If the null turns off return predictability, it must turn on the predictability of dividend growth, and then confront the evidence against such predictability in the data. I find that the absence of dividend growth predictability gives much stronger statistical evidence against the null, with roughly 1-2% probability values, than does the presence of return predictability, which only gives about 20% probability values. I argue that tests based on long-run return and dividend growth regressions provide the cleanest and most interpretable evidence on return predictability, again delivering about 1-2% probability values against the hypothesis that returns are unpredictable. I show that Goyal and Welch's (2005) finding of poor out-of-sample R^2 does not reject return forecastability. Out-of-sample R^2 is poor even if all dividend yield variation comes from time-varying expected returns.

John H. Cochrane
Graduate School of Business
University of Chicago
5807 S. Woodlawn
Chicago, IL 60637
and NBER
john.cochrane@gsb.uchicago.edu

1 Introduction

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 right and left hand variables and in postwar data.

Regression	b	t	$\mathrm{R}^2(\%)$	$\sigma(bx)(\%)$
$R_{t+1} = a + b(D_t/P_t) + \varepsilon_{t+1}$	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
$r_{t+1} = a_r + b_r(d_t - p_t) + \varepsilon_{t+1}$	0.097	1.92	4.0	4.0
$\Delta d_{t+1} = a_d + b_d(d_t - p_t) + \varepsilon_{t+1}$	0.008	0.18	0.00	0.003

Table 1. Forecasting regressions. 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 three-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.

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 5 percentage points, almost as large as the 7.7% level of the equity premium in this sample. Thus, the equity premium varies over time by as much as its unconditional mean. The 4% to 7% R^2 do not look that impressive, but, as emphasized by Fama and French (1988), the R^2 rises with horizon, reaching values between 30 and 60 percent, depending on time period and estimation details. The 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. Finally, prices vary if expected returns (discount rates) vary, and the return forecastability shown in Table 1 neatly accounts for *all* variation in stock prices scaled by dividends. (I present the calculation below). In place of the traditional view that price-dividend ratios rise on news of higher future dividends, or the skeptics' claim that bubbles are needed to account for price volatility, the regressions of Table 1 imply that *all* variation in market price-dividend ratios corresponds to changes in expected excess returns, i.e. risk premiums.

Statistical Significance

However, the *statistical* significance of the first row of Table 1 is marginal, with a t-statistic only a bit 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, since a change in prices moves both variables. 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 towards rejection.

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

Goetzmann and Jorion (1993) and Nelson and Kim (1993) found the distribution of the return-forecasting coefficient by simulation, and they found that the statistical significance of return-forecasting coefficients is much lower than the t-statistics of Table 1 would lead one to believe. Most clearly, Stambaugh (1999) derived the finite-sample distribution of the return-forecasting regression, showing analytically the bias in the return forecast coefficient and the finite-sample standard errors. 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, and the key is in the *dividend growth* regressions of Table 1. Dividends are clearly not forecastable at all. In fact, the small point 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 a constant, which it surely is not. Alternatively (in cointegration language), 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?" We must ask "*which* of dividend growth or returns is forecastable?" (Or really, "how much of each?")

Therefore, if we want to set up a null hypothesis in which returns are not forecastable, that null hypothesis must also specify dividend growth that *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, as Stambaugh and others find. Under the unforecastable-return null, we see return forecast coefficients as large or larger than those in the data about 20% 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) – we only see dividend growth that *fails* to be forecastable in 1-2% of the samples generated under the null. The important evidence, as in Sherlock Holmes' famous case, is the dog that does not bark.[3]

Equivalent Views and Long-Run Regressions

There are several equivalent ways of stating the same point, and I connect them. They stem from the approximate identity (derived below),

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

where b_r is the regression coefficient of log returns on the log dividend yield, b_d is the coefficient of log dividend growth on the log dividend yield, ϕ is the dividend yield autocorrelation, and $\rho \approx 0.96$ is a constant. This identity holds in each sample as well as for population moments.

[2]Additional contributions include Kothari and Shanken (1997), Paye and Timmermann (2003), Torous, Valkanov and Yan (2004) and Ang and Bekaert (2005).

[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

First, we can focus on the joint distribution of (b_r, ϕ), leaving b_d implied, rather than focus on the joint distribution (b_r, b_d), leaving ϕ implied. This is the more conventional framing of the problem, as it allows us to consider forecasting variables that do not include dividends. In the data, we see a relatively high estimate $\hat{\phi} \approx 0.94$ along with the large return-forecast estimate $\hat{b}_r \approx 0.1$ as in Table 1. There is a strong negative correlation between b_r and ϕ estimates, however. Therefore, while under the null we often see b_r higher than in the data - the *marginal* distribution of b_r does not reject – almost all of those high b_r draws come with low ϕ estimates. We almost never see events with b_r as high as we have seen in our sample *and* ϕ as high as we have seen in our sample.

Second, we can divide (1) by $(1 - \rho\phi)$ to obtain

$$\frac{b_r}{1 - \rho\phi} - \frac{b_d}{1 - \rho\phi} = 1 \tag{2}$$

The terms of this identity represent the fractions of dividend yield variance due to changing expected returns to changing expected dividend growth respectively. (I show this below.) They also represent the implied coefficients in regressions of long-run returns $\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$ and long-run dividend growth $\sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}$ on dividend yields.

We can also base a test of return forecastability on the long-run return-forecasting coefficient $b_r/(1 - \rho\phi)$. I find that tests based on these long-run coefficients reject the null with 1-2% probability values. The greater power comes from the negative correlation between b_r and ϕ. Samples with large b_r typically have small values of ϕ, and therefore do not have large values of $b_r/(1 - \rho\phi)$.

Stating null and alternative in terms of the long-run regression coefficients simplifies and clarifies the analysis considerably. It condenses the joint distribution of (b_r, ϕ) into a single number. It captures in that number the observation that we do not see high b_r without low ϕ. Since the long-run return and long-run dividend coefficients in (2) are mechanically related, we do not have to worry whether it is more interesting to test b_r, b_d or some other aspect of the joint distribution.

The question is, what set of draws do we consider "more extreme" than the observed sample, to put in the rejection region of a test statistic? If we base a test statistic on the return-forecast coefficient b_r alone, then many of the draws that produce b_r larger than the value seen in the data also have lower ϕ than in our data ((b_r, ϕ) distribution), or they have lower b_d than in our data ((b_r, b_d) distribution). These draws *do* have forecastable dividend growth, and dividend yield variation is partially due to changing dividend growth forecasts – their dogs do bark. It makes sense therefore to consider such draws "closer to the null" than our data, even though b_r is greater than in our data. 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.

Third, the identity (1) shows that we *can* in fact have both unforecastable returns $b_r = 0$ and unforecastable dividend growth $b_d = 0$ if $\phi = 1/\rho \approx 1.04$. But this specification requires an explosive root in the dividend yield. Thus, the extra information we get about *return* forecastability from *dividend* forecasts b_d or *dividend yield* forecasts ϕ comes from prior information that $\phi < 1.04$. Stronger evidence comes by imposing $\phi < 1$. This is eminently sensible extra information, as I argue at length below. It makes neither statistical nor economic sense to consider dividend yields that have explosive roots. But it *is* extra information, and it is its imposition that allows us to use information in b_d or ϕ to sharpen our knowledge about b_r. This last point is the essence of Lewellen's (2004) calculations, and he also finds strong statistical evidence against the null of unpredictable returns.

Powerful Long-Horizon Regressions?

This success 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 been subject to even greater scrutiny than the one-period regression statistics. 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 example, and they survey the literature. Their Table 5, top row, gives probability values for return forecasts from dividend-price ratios at 1-5 year horizons, based on simulations similar to mine. They report 15%, 14%, 13%, 12%, 17% values. In short, they find no advantage to long horizon regressions.

In this context, how do I find such large power advantages for long-horizon regression coefficients? The main answer is that typical long-horizon estimates do not go far out enough to see the power benefits. I confirm that regression coefficients at 5 and 10 year horizons do not reject much more often than one-year regressions, but I also show that long-horizon regressions give low (below 5%) probability values once one looks past 10 years.

The intuition is straightforward. Regression coefficients rise with horizon as a result of the one-period forecastability together with the persistence of the forecasting variable (Campbell and Shiller 1991). For example, the two-year and three-year return-forecasting coefficients are $b_r^{(2)} = b_r(1+\phi)$ and $b_r^{(3)} = b_r(1+\phi+\phi^2)$ respectively. By this fact, long-horizon return regressions are poised to exploit the negative correlation between b_r and ϕ estimates. Samples that produce a b_r larger than we see in our data typically also produce a ϕ lower than we see in our data. Thus, they typically do not produce coefficients that rise with horizon as do those in our data, and they do not produce large long-horizon regression coefficients. The trouble is that this mechanism is not *quantitatively* strong in two, five or even ten year horizon regressions. b_r and ϕ vary roughly one-for-one across samples, so the combination of b_r and ϕ that one looks at needs also to vary about one-for-one. Since $b_r \approx 0.1$, the two-year regression coefficient $b_r^{(2)} = b_r(1+\phi)$ only weights variation in ϕ by 0.1 times as much as it weights variation in b_r. Three, five, or even ten-year return regressions still do not weight ϕ heavily enough.

This finding does not mean one should construct 30 year returns and regress them directly on dividend yields or other forecasting variables. The difference between *direct* and *implied* estimates is a second reason I find strong rejections and the literature culminating in Boudoukh, Richardson and Whitelaw (2006) does not. For example, one can compute the three-year return regression by forming three-year returns and regressing them on the initial dividend yield, or one can compute the quantity $b_r^{(3)} = b_r(1+\phi+\phi^2)$ from the one-year regression coefficients b_r and ϕ. Like all "nonparametric" estimates, the direct regressions can pick up temporal dynamics not captured by low order VAR models, but at the cost of imprecise estimates. This, I think is the major message of Boudoukh, Richardson and Whitelaw (2006): in "small" samples, *direct* long-horizon estimates can lose the power advantages inherent in a better statistic by poorer measurement.

We obtain therefore a nice resolution of this statistical controversy. I reproduce results such as Boudoukh, Richardson and Whitelaw's (2006) that direct regressions at one to five year horizons have little power advantages over one year regressions, but I also agree with results such as Campbell's (2001) and Valkanov's (2003) that there are power advantages to long-horizon regressions, advantages which are maximized at very long horizons.

Out of sample R^2

Goyal and Welch (2003), (2005) find that return forecasts based on dividend yields and a number of other variables do not work out of sample. They compare forecasts in which one estimates the regression using data up to time t to forecast returns at t+1 with forecasts using the sample mean in the same period. They find 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 about 30-40% of the time. With a highly persistent right hand variable, it is hard to measure the regression coefficient accurately in "short" samples. Thus, this observation does not provide a statistical rejection of forecastable returns. Out of sample R^2 is an interesting *diagnostic*, but it 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 market-timing portfolios given currently available data.

A Common Confusion

One should not come away with the impression 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 always and as in Table 1, that returns are *very* forecastable, where the adjective "very" means by any economic metric. The point estimate (possibly with bias adjustments) remains anyone's best guess. The issue is statistical – "what is the chance that we see something as large as Table 1 by chance, if returns are truly not forecastable?" Stambaugh's (1999) answer is about 15%, and my weaker annual regressions give about 20%. 15% is still not 50% or 90%, so zero return forecastability is still not that likely. "Failing to reject the null" does not mean that we can comfortably *accept* the i.i.d. world view. The point estimate says that every time of high prices (low dividend yields) in the past 80 years has been resolved by low subsequent returns, and not by higher dividend growth. Even if there is a 15% chance that this happened by luck, that high prices "truly" correspond to high dividend growth forecasts and not to low return forecasts, the fact that there is *only* a 15% chance this is true should give one great pause before proclaiming that today's high prices really have no signal about next year's returns.

In this context, I point out that the null of unforecastable returns has other implications which 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 world with unpredictable returns. But this evidence, like the return-based evidence, also does *nothing* to change the point estimate.

2 Null hypothesis

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

$$r_{t+1} = a_r + b_r(d_t - p_t) + \varepsilon^r_{t+1} \quad (3)$$
$$\Delta d_{t+1} = a_d + b_d(d_t - p_t) + \varepsilon^d_{t+1} \quad (4)$$
$$d_{t+1} - p_{t+1} = a_{dp} + \phi(d_t - p_t) + \varepsilon^{dp}_{t+1}. \quad (5)$$

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 linearization of the definition of a return[4] $R_{t+1} = (P_{t+1} + D_{t+1})/P_t$ gives the approximate identity

$$r_{t+1} = \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} - (p_t - d_t)\,, \quad (6)$$

where $\rho = PD/(1+PD)$, PD is the price-dividend ratio about which one linearizes, and lowercase letters are logarithms of corresponding capital letters. This identity links the regression coefficients and errors of the VAR (3)-(5). First, projecting on $d_t - p_t$, identity (6) implies that the regression coefficients obey the approximate identity

$$b_r = 1 - \rho\phi + b_d. \quad (7)$$

Second, the identity (6) links the errors in (3)-(5) by

$$\varepsilon^r_{t+1} = \varepsilon^d_{t+1} - \rho\varepsilon^{dp}_{t+1}. \quad (8)$$

Thus, the three equations (3)-(5) are redundant. One can infer the coefficients and error of any one equation from those of the other two.

The identity (7) shows clearly how we cannot simply take $b_r = 0$ without changing the dividend growth forecast b_d or the dividend yield autocorrelation ϕ. If one changes b_d or ϕ, then the reduced fit of the dividend growth or dividend yield forecasts become evidence against the null as well. 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, i.e. in which both $b_r = 0$ and $b_d = 0$. To generate a coherent null with $b_r = 0$, we must assume an equally large b_d of the opposite sign, and then we must address the *failure* of this dividend growth forecastability in the data.

[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 (6).

By subtracting inflation from both sides, Equations (6)-(8) can apply to real returns and real dividend growth. Subtracting the riskfree rate from both sides, we can relate the excess log return $r_{t+1} - r_t^f$ to dividend growth less the interest rate $\Delta d_{t+1} - r_t^f$. One can either introduce an extra interest rate term and error or simply understand the need to forecast "dividend growth" to include both terms. I focus on real returns and real dividend growth, and I present some results for excess returns. Overall, the results are quite similar with all three data definitions.

To form a null hypothesis, then, I start with estimates of (3)-(5) formed from regressions of log real returns, log real dividend growth and the log dividend yield in annual 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.

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

Table 2. Forecasting regressions and null hypothesis. 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 $r_{t+1} = a_r + b_r(d_t - p_t) + \varepsilon_{t+1}^r$. 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" column describes coefficients used to simulate data under the null hypothesis that returns are not predictable.

Alas, the identity (7) is not exact. The "implied" column of Table 2 gives each coefficient implied by the other two equations and the identity linking the regression coefficients (7). 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 than $0.097/0.05 = 1.94$, and we will see as much as 2-3 percentage point differences in probability values to follow. In this and all remaining calculations I calculate ρ from the mean log dividend yield as

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

The middle three columns of Table 2 present the error standard deviations down 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

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

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

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 and dividend yield shocks (7.5%). The negatively correlated shocks result in a strong negative correlation between b_r and ϕ estimates. In turn, that correlation between estimates underlies the interesting econometrics of the return-forecast coefficient b_r and the stronger power of joint and long-run estimates.

The final columns of Table 2 present the coefficients of the null hypothesis 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 and especially larger values of ϕ below. Given $b_r = 0$ and ϕ, the necessary dividend forecast coefficient b_d follows from the identity $b_d = \rho\phi - 1 + b_r$.

We have to choose two variables to simulate and then let the third follow from the identity (6). I simulate the dividend growth and dividend yield system. However, the identity (6) holds well enough that this choice has almost no effect on the results.

In sum, the null hypotheses thus takes the form

$$\begin{bmatrix} d_{t+1} - p_{t+1} \\ \Delta d_{t+1} \\ \Delta r_{t+1} \end{bmatrix} = \begin{bmatrix} \phi \\ \rho\phi - 1 \\ 0 \end{bmatrix} (d_t - p_t) + \begin{bmatrix} \varepsilon^{dp}_{t+1} \\ \varepsilon^{d}_{t+1} \\ \varepsilon^{d}_{t+1} - \rho\varepsilon^{dp}_{t+1} \end{bmatrix} \tag{9}$$

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

2.1 A structural interpretation and the correlation of shocks

The null hypothesis can be given a deeper and more structural interpretation. This interpretation gives a bit more confidence that the null really does represent a consistent view of the world, and it is helpful to understand the strong negative correlation between return and dividend-yield shocks that is central to the whole affair.

Suppose that expected dividend growth follows an AR(1) process,

$$\Delta d_{t+1} = x_t + \delta^{d}_{t+1} \tag{10}$$
$$x_{t+1} = \phi x_t + \delta^{x}_{t+1} \tag{11}$$

and expected returns are constant. Using the Campbell-Shiller (1988) present value identity

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 alternative, but then returns, prices and dividends would no longer obey the identity $R_{t+1} = (P_{t+1} + D_{t+1})/P_t$ with end-of year prices.

that results from iterating (6) forwards,

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

we then have

$$p_t - d_t = \frac{1}{1-\rho\phi} x_t.$$

The price-dividend ratio *reveals* the expected dividend growth x_t. From the identity (6), returns follow

$$r_{t+1} = \frac{\rho}{1-\rho\phi} \delta^x_{t+1} + \delta^d_{t+1}.$$

Thus, (10) and (11) imply that dividend yields, returns, and dividend growth follow the VAR representation

$$\left[\begin{array}{c} d_{t+1} - p_{t+1} \\ \Delta d_{t+1} \\ \Delta r_{t+1} \end{array} \right] = \left[\begin{array}{c} \phi \\ \rho\phi - 1 \\ 0 \end{array} \right] (d_t - p_t) + \left[\begin{array}{c} -\frac{1}{1-\rho\phi} \delta^x_{t+1} \\ \delta^d_{t+1} \\ \frac{\rho}{1-\rho\phi} \delta^x_{t+1} + \delta^d_{t+1} \end{array} \right] \tag{12}$$

This is exactly the null hypothesis of (9).

We can relate the regression errors ε in (9) to the "structural" shocks δ. The dividend-yield shock is proportional to the shock to expected dividend growth δ^x. When there is news of higher expected future dividend growth, prices go up today. Thus, in the context of the null, we can label the dividend-yield shock as a "shock to expected dividend growth." The ex-post dividend growth shock is just that, structurally or as a regression error. The return shock is a combination of the dividend-growth shock and the expected dividend growth shock. Returns are high if there is an unexpected dividend, or if prices rise unexpectedly on news of future dividends.

Table 2 reveals that, using the shock covariance matrix from the data, expected dividend growth shocks δ^x and ε^{dp} are essentially uncorrelated with ex-post dividend growth shocks δ^d and ε^d, and both are of the same order of magnitude The strong negative correlation between return and dividend-yield shocks then *follows* naturally from the fact that the dividend-yield (or expected dividend growth) shock is part of the return shock in (9) and (12). News of good expected future dividend growth sends prices up today, raising returns and lowering dividend yields at the same time. There is no tendency for this correlation to be upset by simultaneous news about current dividend growth, which would raise returns and dividend yields together.

In sum, if our system is driven by "structural" expected dividend growth and ex-post dividend growth shocks that are essentially uncorrelated, the strong correlation between return and dividend yield shocks follows naturally and endogenously.

One can give a related "structural" interpretation to the sample estimates. They are consistent with a structure analogous to (10) and (11) in which regression errors reveal shocks to expected returns and shocks to current dividend growth, with constant expected dividend growth. The expected-return shock and the dividend-growth shock are essentially uncorrelated with each other, and the negative correlation between return and dividend yield shocks again emerges endogenously. This system is set out briefly in Section 6, and in more detail in Cochrane (2004) Ch. 20.

3 Distribution of regression coefficients and t statistics

3.1 Return and dividend growth forecasts

In each Monte Carlo draw I run regressions

$$\begin{aligned} r_{t+1} &= a_r + b_r(d_t - p_t) + v^r_{t+1} \\ \Delta d_{t+1} &= a_d + b_d(d_t - p_t) + v^d_{t+1}. \end{aligned}$$

Figure 1 plots the joint distribution of the return and dividend-forecast regression 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 produces a coefficient larger than the roughly $\hat{b}_r \approx 0.10$ sample estimate 22% of the time, and a larger t statistic than the sample $\hat{t} = 1.92$ 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).

	b_r	t_r	b_d	t_d	b_r, b_d	b_r, ϕ
Real	22.3	10.3	1.77	1.67	1.75	0.02
Excess	17.4	6.32	1.11	0.87	1.10	0.01

Table 3. Percent probability values under the $\phi = 0.941$ null. Each column gives the probability that the indicated coefficients are greater than their sample values, under the null. Columns with two variables give the probability that both are greater than their sample values, i.e. b_r, b_d gives the probability that $b_r > \hat{b}_r$ and $b_d > \hat{b}_d$. Monte Carlo simulation of the null described in Table 2 with 50,000 draws.

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 cloud of Figure 1 is vertically centered a good deal below zero and below the horizontal line of the sample estimate $\hat{b}_d$. *Dividend* growth forecasting coefficients larger than the roughly zero values observed in sample are only seen 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 only observed 1.11% of the time and the t statistic only 0.87% of the time (Table 3).

This is my central point: the *lack* of dividend forecastability in the data gives in fact far stronger statistical evidence against the null than does the *presence* of return forecastability, lowering probability values from the 20% range to the 1% range. (I discuss the $\phi = 0.99$ results in Figure 1 below.)

3.2 The ϕ view

The return forecast coefficient b_r, the dividend-yield autocorrelation ϕ, and the dividend growth forecast coefficient b_d are related by the approximate identity $b_r = 1 - \rho\phi + b_d$. Therefore, the

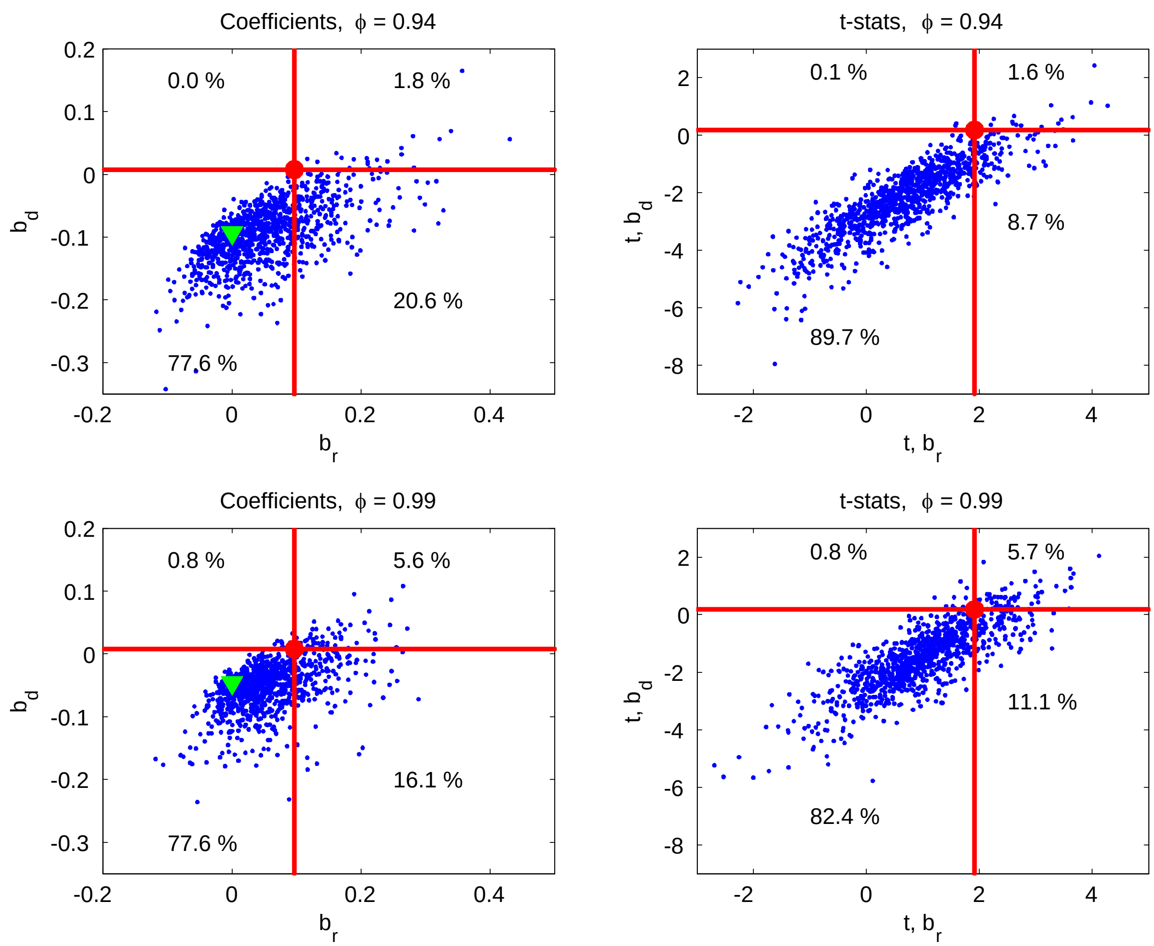

Figure 1: Joint distribution of return and dividend growth forecasting coefficients (left) and t-statistics (right). The lines and dot give the sample estimates. The triangle gives the null. 1000 simulations are plotted for clarity; each point represents 1/10% probability. Percentages are the fraction of 50,000 simulations that fall in the indicated quadrants.

same information in the (b_r, b_d) joint distribution is captured in the (b_r, ϕ) joint distribution or the (b_d, ϕ) joint distribution. Recasting the point in the (b_r, ϕ) context is especially important since most articles study return forecastability in a two-variable VAR consisting of returns and the forecasting variable, leaving the behavior of dividends implicit from identities.

The left-hand panels of Figure 2 plot the joint distribution of (b_r, ϕ). 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). However, high return coefficients b_r tend to come with low dividend yield autocorrelations ϕ. We almost never see a return forecast as high as we do in the data together with a dividend yield autocorrelation as high as the $\phi = 0.941$ we seen in the data – the Northeast quadrants are nearly empty. (The left panels of Figure 2 are the same as Lewellen's (2004) Figure 1, Panel B except Lewellen calibrates to monthly postwar data. Lewellen focuses on a different distributional calculation.)

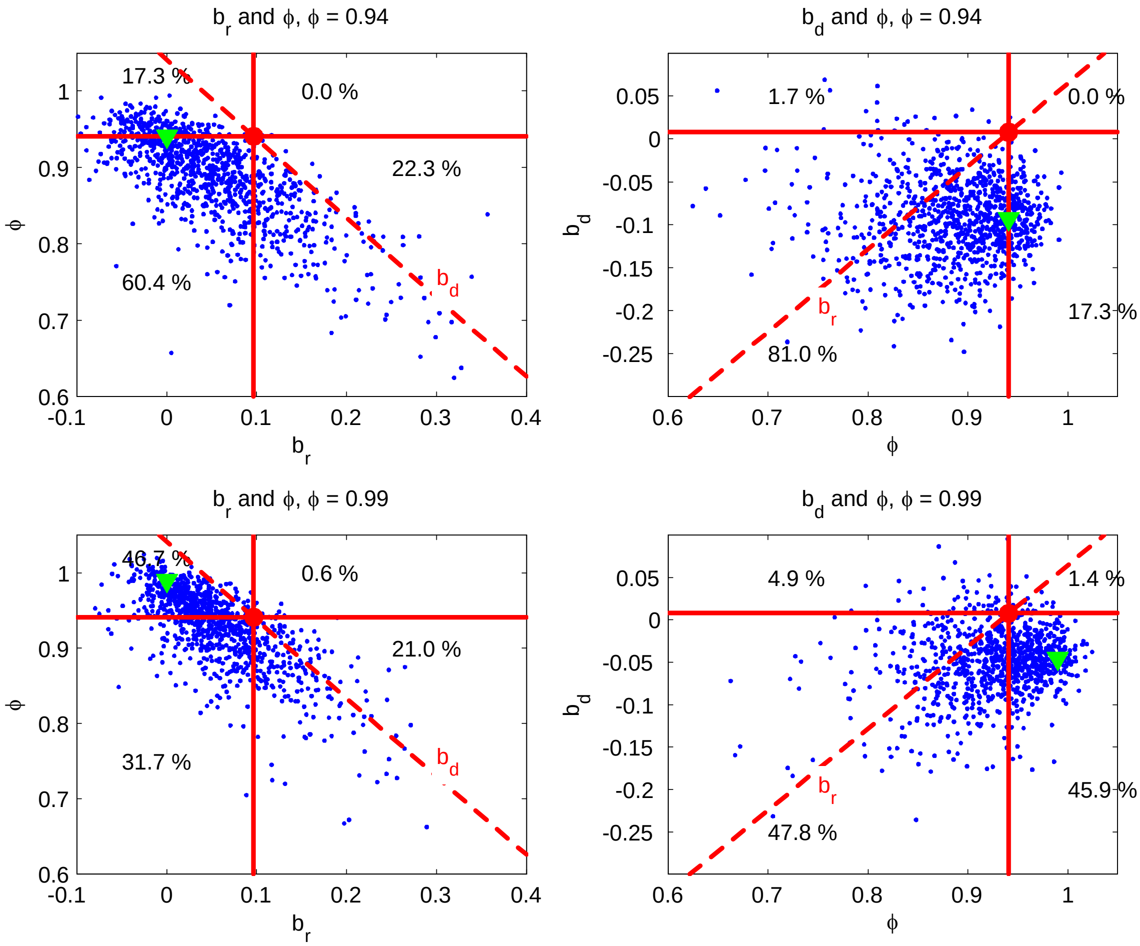

Figure 2: Joint distributions of regression coefficients. Left hand panels give the joint distribution of b_r, ϕ. Right hand panels give the joint distribution of b_d, ϕ. In each graph the triangle marks the null hypothesis used to generate the data and the circle marks the estimated coefficients $\hat{b}_r, \hat{b}_d, \hat{\phi}$. The diagonal dashed line marked "b_d" in the left hand panels marks the region $b_r = 1 - \rho\phi + \hat{b}_d$; points above and to the right are draws where b_d exceeds its sample value. The diagonal dashed line marked "b_r" in the right hand panels marks the region $b_d = \rho\phi - 1 + \hat{b}_r$; points above and to the left are draws where b_r exceeds its sample value. Numbers are the percentage of the draws that fall in the indicated quadrants.

The negative correlation between b_r and ϕ estimates, visible in the diagonal spread of points in Figure 2, is the key to this result. A lower sample value of ϕ, through the identity $b_r = 1 - \rho\phi + b_d$ *must* correspond to a larger sample value of b_r, a larger sample value of b_d, or both. If the dividend yield seems to revert quickly after a shock in a given sample, then it must be the case that one of dividend growth or prices and hence returns moves a lot after the shock, to generate the quick reversion of dividend yields. In fact, lower ϕ are primarily largely associated with higher b_r rather than lower b_d. In turn, this fact is driven by the strong negative correlation between ϕ and b_r shocks seen in Table 2. If the shocks in two regressions are negatively correlated, then a set of shocks that produces an unusually large coefficient in the first regression corresponds to set of shocks that produces an unusually small coefficient in the second regression.

To relate the (b_r, b_d) joint distribution with the (b_r, ϕ) joint distribution, the diagonal dashed line marked b_d in the top left panel of Figure 2 marks the set $b_r = 1 - \rho\phi + \hat{b}_d$ where $\hat{b}_d$ is the sample estimate. Points above and to the right of this dashed line are exactly the points above $b_d > \hat{b}_d$ in the (b_r, b_d) distribution of Figure 1. In this way, we can see both joint distributions ((b_r, b_d) and (b_r, ϕ)) on the same graph.

With both joint distributions on the same graph, we can see that the region $(b_r > \hat{b}_r, \phi > \hat{\phi})$ is in fact more restrictive than the region $(b_r > \hat{b}_r, b > \hat{b}_d)$, which is why we see lower probability values for the $(b_r > \hat{b}_r, \phi > \hat{\phi})$ region.

More importantly, we see that the region $b > \hat{b}_d$ (above and to the right of the dashed diagonal line) captures in a single number the essence of the joint (b_r, ϕ) distribution. The point of the (b_r, ϕ) distribution is that samples with large return-forecast coefficient b_r come with low dividend yield autocorrelation ϕ. By the identity $b_d = b_r + \rho\phi - 1$, such samples come with low b_d, so setting up a rejection region above and to the right of the diagonal b_d line captures the message of the negative correlation between b_r and ϕ.

Looking at the $(b_r,\ \phi)$ system has the advantage that one can make the same distributional points with an arbitrary right hand variable, one that is not connected to dividend growth via any identities. However, the strong negative correlation between b_r and ϕ estimates visible in Figure 2 is an important component of the results. In turn, the correlation of estimates derives from the strong correlation between return and dividend yield shocks seen in Table 2. As I argued above, that correlation between shocks emerges naturally in dynamic present value models, since a change in expected return or expected dividend growth gives rise to a price change that moves both dividend yield and return. An arbitrary right hand variable, especially one that does not include price, may not produce such a strong correlation.

The right hand panels of Figure 2 complete the trio of views by plotting the joint distribution of dividend growth and dividend yield forecasting coefficients (b_d, ϕ). There is no particular correlation between the two coefficients in this case, resulting from the near-zero correlation between dividend growth and dividend yield shocks, as seen in Table 2. The cloud is smeared to the left however. The distribution of b_d conditional on a given ϕ (vertical slices) becomes more spread out for lower ϕ, as one moves to the left The leftward smear of the cloud relative to the null (triangle) comes from the downward bias and large left tail of autocorrelation ϕ estimates. Thus, though the unconditional chance of seeing a dividend growth forecast as high as in the data (above the horizontal line) is already low, there are almost no observations in the Northeast corner, where we see a large dividend growth forecast *and* high sample autocorrelation.

4 Long-horizon coefficients

4.1 A long-run identity

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

$$\begin{aligned} \frac{b_r}{1-\rho\phi} - \frac{b_d}{1-\rho\phi} &= 1 \\ b_r^{lr} - b_d^{lr} &= 1. \end{aligned} \tag{13}$$

The second row defines notation. Casting the problem in terms of these coefficients simplifies and clarifies the analysis considerably.

The terms of identity (13) 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 (6) forward, giving the Campbell-Shiller (1988) present value identity

$$d_t - p_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} - E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}. \tag{14}$$

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

$$var(d_t - p_t) = cov\left(\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}, d_t - p_t\right) - cov\left(\sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}, d_t - p_t\right). \tag{15}$$

This equation states that all variation in the dividend-price 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,

$$\beta\left(\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}, d_t - p_t\right) - \beta\left(\sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}, d_t - p_t\right) = 1 \tag{16}$$

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

$$\beta\left(\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}, d_t - p_t\right) = \sum_{j=1}^{\infty} \rho^{j-1} \beta\left(r_{t+j}, d_t - p_t\right) = \sum_{j=1}^{\infty} \rho^{j-1} \phi^{j-1} b_r = \frac{b_r}{1 - \rho\phi} = b_r^{lr} \tag{17}$$

and similarly for dividend growth.

Negative b_d^{lr} is the fraction of dividend-yield volatility due to dividend growth, since if high prices and a low dividend yield signal higher future dividends, then b_d^{lr} and b_d are negative. If a high dividend yield instead means higher dividend growth, expected returns must move even further to explain the movement in dividend yield, thus explaining "more than 100%" of dividend yield variation. More than 100% and less than zero are therefore possible. This is not a decomposition into orthogonal components. This sort of calculation is the standard way to adapt the ideas of Shiller's (1981) and LeRoy and Porter's (1981) volatility tests to the fact that dividend yields rather than price levels are stationary. See Campbell and Shiller (1988) and Cochrane (1991), (1992), (2004) for more details.

Using the identity $b_r - b_d = 1 - \rho\phi$, we can also express the identity linking long-run coefficients (13) as

$$\frac{b_r}{b_r - b_d} - \frac{b_d}{b_r - b_d} = 1. \tag{18}$$

This equation expresses the same ideas in another way. $b_r - b_d$ is the total amount of predictability we see in the data. Returns or dividends must be forecastable to pull the dividend yield back after a shock, and the faster it reverts (lower ϕ), the larger b_r and $-b_d$ must be. The two terms then capture how much of the needed overall predictability $b_r - b_d$ is in returns (first term) and how much is in dividend growth (second term).

4.2 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 standard errors using the delta-method and the heteroskedasticity-corrected OLS standard errors[6] in Table 2.

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

Table 4. Long-run regression coefficients. 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.

Table 4 shows that dividend yield volatility is almost exactly accounted for entirely 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 the return forecasting coefficient is highly economically significant. This finding is a simple consequence of the familiar estimates. $\hat{b}_d \approx 0$ means $\hat{b}_d^{lr} \approx 0$ of course, and

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

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

[6] I compute standard errors for long run coefficients from standard errors for $\hat{b}_r$ and $\hat{\phi}$ as follows

$$\sigma^2(\hat{b}_r^{lr}) = \sigma^2\left[\frac{\partial b_r^{lr}}{\partial b_r}\hat{b}_r + \frac{\partial b_r^{lr}}{\partial \phi}\hat{\phi}\right] = \left(\frac{\partial b_r^{lr}}{\partial b_r}\right)^2 \sigma^2\left(\hat{b}_r\right) + \left(\frac{\partial b_r^{lr}}{\partial \phi}\right)^2 \sigma^2\left(\hat{\phi}\right) + 2\frac{\partial b_r^{lr}}{\partial b_r}\frac{\partial b_r^{lr}}{\partial \phi}\sigma(\hat{b}_r,\hat{\phi})$$

$$\frac{\partial b_r^{lr}}{\partial b_r} = \frac{1}{1-\rho\phi};\ \frac{\partial b_r^{lr}}{\partial \phi} = \frac{\rho b_r}{(1-\rho\phi)^2} = \frac{\rho}{1-\rho\phi}b_r^{lr}$$

$$\sigma^2\left(\hat{b}_r^{lr}\right) = \left(\frac{1}{1-\rho\phi}\right)^2\sigma^2(\hat{b}_r) + \left(\frac{\rho}{1-\rho\phi}\right)^2\left(b_r^{lr}\right)^2\sigma^2(\hat{\phi}) + 2\frac{\rho}{(1-\rho\phi)^2}b_r^{lr}\sigma(b_r,\phi)$$

$$\sigma^2\left(\hat{b}_r^{lr}\right) = \left(\frac{1}{1-\rho\phi}\right)^2\left[\sigma^2(\hat{b}_r) + 2\rho b_r^{lr}\sigma(\hat{b}_r,\hat{\phi}) + \left(\rho b_r^{lr}\right)^2\sigma^2(\hat{\phi})\right].$$

way. Excess returns in the last row of Table 4 show slightly stronger results. High prices-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.[7] 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 $b_d^{lr} = -1$. Using the long-horizon regression coefficients, we do not need to choose between return and dividend-growth tests.

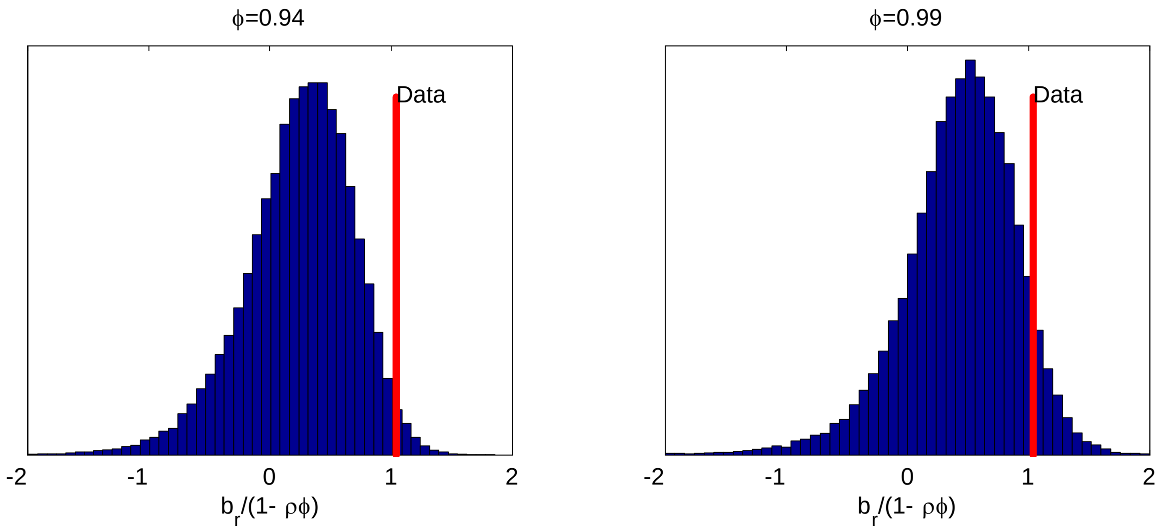

Figure 3: Distribution of $b_r/(1-\rho\phi)$. The vertical bar gives the corresponding value in the data.

Figure 3 tabulates the small-sample distribution of the long-run return-forecast estimates, and Table 4 includes the probability values, i.e. how many long-run return forecasts are greater than the sample value under the unforecastable-return null $b_{lr}^r = 0$. By the identity $b_r^{lr} - b_d^{lr} = 1$, these are the same as how many long-run dividend growth forecasts are greater than the sample value under the null $b_{lr}^d = -1$. There is about a 1.5% probability value of seeing a long-run return forecast b_r^{lr} larger than seen in the data, or a long run dividend growth forecast b_d^{lr} 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.) Comparing this 1.5% probability value to the 22% or so probability values for $b_r > \hat{b}_r$, we see that the long-run coefficient incorporates the *joint* information in returns and dividend growth, or returns and dividend-yield autocorrelation, in a *single* number.

Specifically, we saw in Figure 2 that b_r is large predominantly in samples in which ϕ is low. When ϕ is low, however, $b_r^{lr} = b^r/(1-\rho\phi)$ is not so large. Thus, the long-run coefficient captures

[7] The identities are only approximate, so to display estimates that obey the identities one must estimate two of b_r, b_d, and ϕ, 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 estimates 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 $b_d^{lr} = 0.08$ $(s.e. = 0.42)$.

in a single number the point of the joint b_r, ϕ distribution of Figure 2, that we seldom see high b_r without also seeing low ϕ.

Similarly, we saw in Figure 1 that large b_r usually come with small (large negative) b_d. In the context of the identity (18), the long-run regression coefficient is $b_r^{lr} = b_r/(b_r - b_d)$. Large b_r that also come with large negative b_d count less in b_r^{lr}, so testing the long-run coefficient captures the point of the joint (b_r, b_d) distribution in a single number.

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% for the test $b_r^{lr} = 0$ are correspondingly lower and the evidence against the null even stronger.

4.3 The advantages of long-run coefficients

Recasting the problem in terms of the long-run coefficients b_r^{lr} and b_d^{lr} provides an elegant way to characterize the null and alternative. In particular, the long-run coefficients solve the arbitrariness of the joint regions for b_r and b_d, or b_r and ϕ, by boiling them down to a single number, and they capture the null and alternative in the cleanest way.

Boiling a joint distribution down to a single test statistic is always troublesome. Should we test $b_r > \hat{b}_r$, or should we test $b_d > \hat{b}_d$? Or perhaps we should test some other subset of the (b_r, b_d) region, or the (b_r, ϕ) region? Certainly the joint probabilities $(b_r > \hat{b}_r,\ \phi > \hat{\phi})$ go too far. I present them as interesting characterizations of the joint distribution, but one would not likely set up a test region in the Northeast quadrants, since one would not likely commit to *accepting* the null outside that quadrant, i.e. with an arbitrarily large b_r but b_d or ϕ just below some cutoff.

The issue comes down to defining what is the "event" we have seen, and what other events we would consider "more extreme," and so should count as being further out in the tail. Here, the long-run coefficients neatly solve the conundrums posed by the joint distribution of short-run coefficients.

We conventionally 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 return-forecast coefficients greater than those seen in the data, $b_r > \hat{b}_r$, have dividend-growth forecast coefficients lower than seen in the data, $b_d < \hat{b}_d$, or dividend-yield autocorrelations lower than seen in the data $\phi < \hat{\phi}$. In these events, dividend growth *is* forecastable and *does* count for a portion of dividend yield variation. Are these really "more extreme" events, further from the unpredictable-return null than what we have seen in our data? Or, should we instead count such events, on full inspection, as being closer to the null than the event in our data? For example, if our data showed $\hat{b}_r = 0.2$, but volatility tests were a half success, finding half of dividend yield variance due to dividend growth forecasts, rather than the dismal failure they are in our data, would we not think of ourselves as closer to the null than we are now? That is how the long-run coefficients count things. For similar reasons, $b_d^{lr} = -1$ is a more useful statement of the null than is $b_d = -0.1$.

The long-run coefficients also solve the seeming arbitrariness of which joint distribution one chooses to look at. Since the long-run coefficients obey the identities (13) and (18), there is no difference whether we think in terms of return coefficients, dividend coefficients, or joint properties of returns b_r, dividends b_d or dividend-yields ϕ. Every statistic or pair of variables gives exactly the same answer.

The long-run coefficients seem to give the same answer as the test on dividend-growth coefficients alone, $b_d > \hat{b}_d$. In fact they are different conceptually and slightly different in this sample. The long-run coefficient test $b_d^{lr} > \hat{b}_d^{lr}$ means $b_d/(1-\rho\phi) > \hat{b}_d(1-\rho\hat{\phi})$. If we had $\hat{b}_d = 0$ exactly, these two events would be the same. With $\hat{b}_d \neq 0$, a different sample ϕ can affect the long-run dividend growth coefficient b_d^{lr}, perhaps pushing it across a boundary, for the same value of the short-run dividend growth coefficient b_d. It is the fact that $\hat{b}_d$ is so close to zero that makes the results and intuition (regions in the joint distribution regions) so similar between b_d and long-run tests in our data.

5 Autocorrelation ϕ, unit roots, bubbles, and priors

So far I have used the sample value of the dividend yield autocorrelation $\phi = 0.941$ to generate the Monte Carlo. One naturally wants to know how the results are affected by the choice of ϕ, and especially by larger values of ϕ, given the downward bias in autocorrelation estimates.

5.1 Results for different ϕ values

Table 5 collects probability values for various events as a function of ϕ. The previous figures include the case $\phi = 0.99$.

	Percent probability values										Statistics	
Null	Real Returns					Excess returns						
ϕ	b_r	b_d	b_r, ϕ	$b^{lr}_{\min}$	$b^{lr}_{\max}$	b_r	b_d	b_r, ϕ	$b^{lr}_{\min}$	$b^{lr}_{\max}$	$\sigma\,(dp)$	1/2 life
0.90	24	0.6	0.00	0.3	0.6	19	0.4	0.00	0.1	0.2	0.35	6.6
0.941	22	1.6	0.06	1.2	1.7	17	1.1	0.01	0.5	0.7	0.45	11.4
0.96	22	2.6	0.08	2.0	2.8	17	1.6	0.06	0.8	1.2	0.55	17.0
0.98	21	4.9	0.4	4.3	5.5	17	2.7	0.2	1.8	2.5	0.77	34.3
0.99	21	6.3	0.8	5.9	7.4	17	3.6	0.3	2.7	3.6	1.09	69.0
1.00	22	8.7	1.0	8.1	10	16	4.4	0.5	3.7	4.8	∞	∞
1.01	19	11	1.5	11	13	14	5.1	0.6	5.1	6.3	∞	∞
Draw ϕ	23	1.6	0.1	1.4	1.7	18	1.1	0.04	0.6	0.8		

Table 5. The effects of dividend-yield autocorrelation ϕ. 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_r, ϕ gives the chance of seeing both statistics greater than their data counterparts. b^{lr} is the long-run regression coefficient, e.g. $b_r^{lr} = b_r/(1-\rho\phi)$. $b^{lr}_{\min}$ 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 $\sigma(dp) = \sigma_{\varepsilon,dp}/\sqrt{1-\phi^2}$. Half life is the value of τ such that $\phi^\tau = 1/2$.

As ϕ rises, the identity $b_d = b_r + \rho\phi - 1$ requires larger (less negative) dividend-growth coefficients b_d in the null to go along with $b_r = 0$. At the sample $\phi = 0.941$, we needed $b_d \approx -0.1$ to go along with $b_r = 0$. As ϕ rises to $\phi = 1$, for example, we only need $b_d = \rho - 1 \approx -0.04$. As the null b_d rises, the chance of seeing values greater than in the data, $b_d > \hat{b}_d$, naturally rises. This behavior is clear comparing the top and bottom panels of Figure 1. Raising ϕ and thus raising b_d in the null raises the triangle representing the null. The cloud of simulation points rises, so the chance of seeing $b_d > \hat{b}_d$ above the horizontal line rises as well. However, the cloud doesn't rise much, and its shape is changed reflecting more severe small-sample biases. Looking down the b_d column of Table 5, the $b_d > \hat{b}_d$ probability for real returns crosses the 5% mark a bit above $\phi = 0.98$ and is still below 10% at $\phi = 1$. Excess returns are stronger as usual, with the b_d probability value still below 5% at $\phi = 1$. In all cases, the dividend-growth test b_d still has more information, with less than half the probability value of the return-forecast b_r region.

As the null rises in Figure 1, the simulation points do not move much to the left or right. Therefore, raising ϕ has little effect on the b_r statistic, which is about 22% for all values of ϕ in Table 5.

The joint distributions of Figure 2 and the corresponding Northeast-quadrant probability values b_r, ϕ in Table 5 show a similar pattern. In the (b_r, ϕ) distribution, raising ϕ raises the null triangle, raising the cloud of points somewhat. The increased downward bias in ϕ works against this rise however, as the cloud of points does not rise one for one with the triangle null.

The probability values of the long-run coefficients $b^{lr} = b/(1 - \rho\phi)$ also rise with ϕ as shown in Table 5. These probability values cross the 5% line at about $\phi = 0.98$ for real returns, and stay below 5% all the way to $\phi = 1$ for excess returns. The evidence from the long-horizon coefficients is again stronger than the one-period coefficient b_r evidence at any ϕ.

5.2 What's the right ϕ?

One can simply stop at Table 5 and catalog the probability values as a function of the assumed null ϕ. But it's natural to think a bit about how large a value of ϕ we should consider, and thus how strong the evidence really is.

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. The forward iteration used to derive the present value relation (28) from the return identity (6) is

$$p_t - d_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} + \lim_{k \to \infty} \rho^k E_t \left(p_{t+k} - d_{t+k} \right) \tag{19}$$

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 $b_r = 1 - \rho\phi + b_d$. In fact, in this case we *must* have $b_r = b_d = 0$, otherwise the terms $E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} = \sum_{j=1}^{\infty} \rho^{j-1} \phi^{j-1} b_r (d_t - p_t)$ do not converge. This is the case of "rational bubble." If $\phi = 1/\rho$ exactly, then price-dividend ratios vary on changing expectations of their future values, the last term of Equation (19), 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 doesn't 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.

Can we seriously 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 we do not reject $\phi = 1$ on that basis.

Long-run evidence 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 $\pm 1\sigma = (1.9 - 336)$, and $\pm 2\sigma = (0.14 - 4514)$. If dividend yields really follow a random walk, we should have seen observations of this sort. But market 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 fairly likely to see a market price-dividend ratio of one, or one thousand, in the next century or two? These points are mirrored in the *infinite* unconditional variance of the dividend yield tabulated in Table 5.

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 over centuries, 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 ϕ below one that is also too large to swallow. To get a handle on this question, Table 5 also includes the unconditional variance of dividend yields and the half-life of dividend yields implied by the assumed 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 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 a good deal larger than any sensible view of the world.

However, nothing dramatic happens as ϕ rises from 0.98 to 1.01, so one may take any upper limit in this range without changing the conclusions dramatically. And that conclusion remains a rejection of the null that returns are unpredictable, with the consequence that dividend growth is predictable, with probability values in the 1% to 5% range.

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

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

5.3 An overall number

It would be nice to present a single number, rather than a table of values as a function of an assumed value for the dividend-yield autocorrelation ϕ. We can do this 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 simply the likelihood function of an AR(1) process fit to the dividend yield, with the intercept and error variance parameters maximized out. Details are in the Appendix. The conditional likelihood takes the first data point as fixed. The unconditional likelihood adds the log probability of the first datapoint, 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.

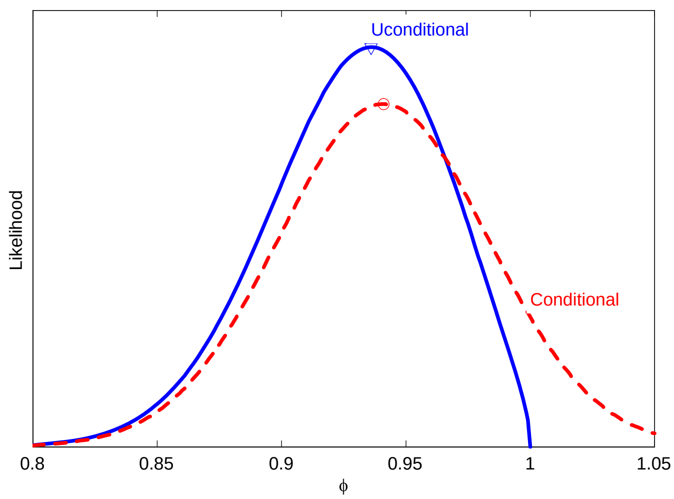

Figure 4: Likelihood function for ϕ, the autoregressive parameter for dividend yields. The likelihood is based on an autoregressive model, $d_{t+1} - p_{t+1} = a_{dp} + \phi(d_t - p_t) + \varepsilon^{dp}_{t+1}$. The intercept a_{dp} and innovation variance $\sigma^2(\varepsilon^{dp})$ are maximized out.

I repeat the simulation, but this time drawing ϕ from the unconditional likelihood plotted in Figure 4 before drawing a sample of errors ε^{dp}_t and ε^d_t. 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 $\phi > 1/\rho \approx 1.04$ where 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 a 1.6% chance of seeing a dividend-growth forecast $b_d > \hat{b}_d$ or the 1.4% - 1.7% chance of seeing the more elegant long-run regression coefficients b^{lr}_r or b^{lr}_d greater than their sample values. As usual, excess returns give even stronger rejections, with $b_d > \hat{b}_d$ occurring only 1.1% of the time, and the long-run coefficient b_r^{lr} only exceeding its sample value $\hat{b}_r^{lr}$ 0.6% - 0.8% of the time. (Lewellen 2004 presents a similar and more formally Bayesian calculation that also delivers small probability values.)

5.4 Bias in forecast estimates

Table 6 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 the consequent strong negative correlation between return and dividend-yield coefficients.

The dividend-growth coefficient b_d is not biased. As seen in Figure 2, there is no particular correlation between the b_d and ϕ estimates, again 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 value $b_r^{lr} = 0.24 - 0.43$ from the sample value $\hat{b}_r^{lr} \approx 1$.

		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

Table 6. Means of estimated parameters. Means are taken over 50,000 simulations of the Monte Carlo described in Table 2.

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 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 put 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 tiny.

6.1 Out of sample $\mathbf{R}^2$ as a test

Does this result mean that "returns are really not forecastable?" If all dividend yield variation was really due to *return* forecasts, how often would we see Goyal-Welch results? To answer this question, I set up the analogous null in which returns *are* forecastable and dividend growth *is not* forecastable, and all dividend-yield variation comes from time-varying expected returns. Let expected returns vary through time,

$$E_t(r_{t+1}) = x_{t+1} = \phi x_t - \delta^x_{t+1}.$$

(The sign of δ is arbitrary. With a negative sign, a positive δ shock raises the ex-post return, so the VAR covariance matrix becomes identical to the last case.) Now, let dividend growth be completely unforecastable,

$$\Delta d_{t+1} = \delta^d_{t+1}.$$

Imposing the Campbell-Shiller identity (28), we have

$$p_t - d_t = -\frac{1}{1-\rho\phi} x_t.$$

Returns follow

$$\begin{aligned} r_{t+1} &= \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} - (p_t - d_t) \\ &= x_t + \frac{\rho}{1-\rho\phi}\delta^x_{t+1} + \delta^d_{t+1} \\ &= (1-\rho\phi)(d_t - p_t) + \frac{\rho}{1-\rho\phi}\delta^x_{t+1} + \delta^d_{t+1}. \end{aligned}$$

Thus, we have a VAR representation

$$\begin{bmatrix} d_{t+1} - p_{t+1} \\ r_{t+1} \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 1-\rho\phi & 0 \end{bmatrix} \begin{bmatrix} d_{t+1} - p_{t+1} \\ r_{t+1} \end{bmatrix} + \begin{bmatrix} -\frac{1}{1-\rho\phi}\delta_{t+1} \\ \frac{\rho}{1-\rho\phi}\delta_{t+1} + \varepsilon_{t+1} \end{bmatrix}. \tag{20}$$

This is exactly the same VAR as before but with a $1 - \rho\phi$ in the return forecast slot rather than zero, and (not shown) zero in the dividend-growth forecast slot rather than $\rho\phi - 1$. The dividend yield regression error reveals shocks to expected returns; the dividend growth error is the second shock, and the return shock combines those two "structural" shocks. (See Cochrane 2004 Ch. 20 for more details.)

I simulate artificial data from this null as before. I start with $\phi = 0.941$, which gives the sample 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: I start in year 20, and 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, larger negative values mean the sample mean is winning.

Figure 5 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 forecast than the dividend yield, as Goyal and Welch find. However, 30-40% of the draws show even worse results than our sample. In these cases, even though *all* dividend-price variation is due to time-varying expected returns by construction, the dividend yield is an even worse "out of sample" forecaster than it is in the observed data. In fact, the *mean* of the Goyal-Welch statistic is negative, and only about 20% of the draws show a *positive* value. Under this null, it is *unusual* for dividend-yield forecasting actually to work better than the sample mean in this out-of-sample experiment.

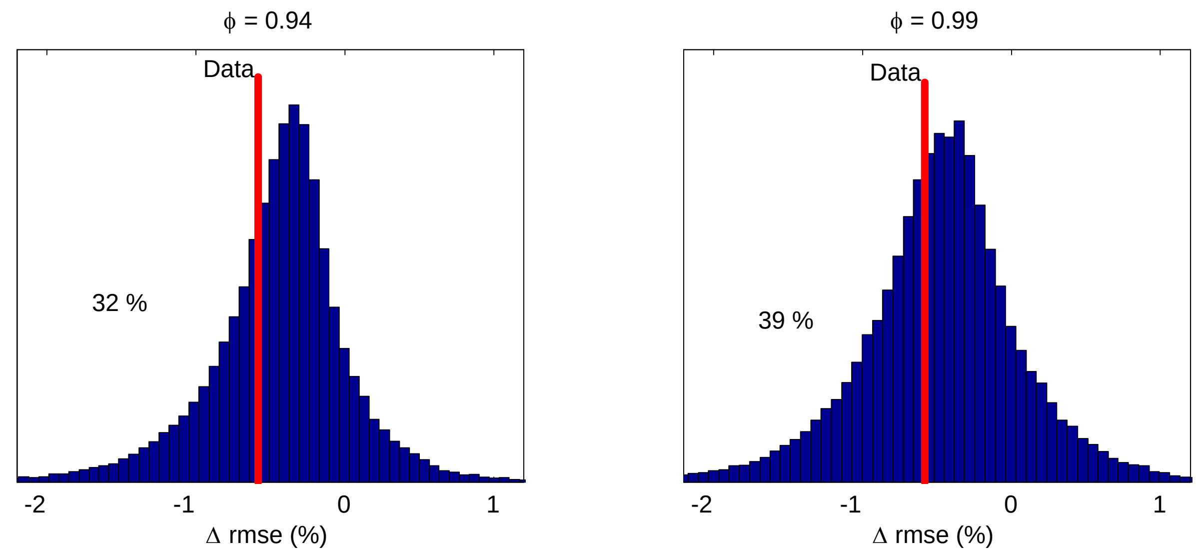

Figure 5: Distribution of the Goyal-Welch statistic under the null that returns are forecastable and dividend growth is not forecastable. The statistic is the root mean squared error from using the sample mean return from time 1 to time t to forecast returns at $t+1$, less the root mean squared error from using a dividend yield regression from time 1 to time t to forecast returns at time $t+1$.

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. Goyal and Welch's message is that regressions on dividend yields and similarly persistent variables are not likely to be useful in forming market-timing portfolios, given the difficulty of accurately estimating the return-forecasting 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 one should use a much lower parameter in a market-timing portfolio, shading the portfolio advice well back towards simple use of the sample mean. (How these more sophisticated calculations perform out of sample, extending Campbell and Thompson's 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 a *diagnostic*, not a *test*. Out-of-sample R^2 is not a new and powerful 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 portfolio advice, given uncertainties about the coefficients in our data sets.

7 Power in long-run regression coefficients?

I find much greater ability to reject the unforecastable-return null in the long-horizon coefficient $b_r^{lr} = b_r/(1-\rho\phi)$ than in the one-year coefficient b_r. How can we reconcile this result with the findings of the literature such as Boudoukh, Richardson and Whitelaw (2006), that finds no power advantage[10] in long-horizon regressions?

7.1 Long-horizon regressions compared

There are three main differences between the coefficients that I have calculated and typical long-horizon regressions. First, b_r^{lr} 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 regression coefficients, $\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$ on $d_t - p_t$. Second, $b_r^{lr} = b_r/(1-\rho\phi)$ is *implied* from the first-order VAR. Most studies examine instead *direct* regression coefficients, i.e. 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, i.e. $\sum_{j=1}^{k} r_{t+j}$ on $d_t - p_t$.

To find out which of these three differences in technique accounts for the difference in results, Table 7 evaluates long-horizon regressions. The first row of Table 7 presents the familiar one-year return forecast. We see the usual coefficient of $\hat{b}_r = 0.10$, with 22 percent probability value of observing a larger coefficient under the null.

Increasing to a 5 year horizon, we see that the 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 here findings such as Boudoukh Richardson and Whitelaw's (2006) that directly-estimated long-horizon regressions do not improve power over one-period 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 1-5 year horizons, one might well conclude that long-horizon regressions can have some, but only a little additional power.

As we increase horizon, however, 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.

[10] Boudoukh, Richardson and Whitelaw focus much of their discussion on the high correlation of short and long-term 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, i.e. power to reject the null hypothesis.

	Weighted $\sum_{j=1}^{k}\rho^{j-1}r_{t+j} = a + b_r^{(k)}(d_t - p_t) + \delta_{t+k}$						Unweighted $\sum_{j=1}^{k} r_{t+j} = a + b_r^{(k)}(d_t - p_t) + \delta_{t+k}$					
	direct			implied			direct			implied		
	coeff.	p-value, $\phi =$		coeff.	p-value, $\phi =$		coeff.	p-value, $\phi =$		coeff.	p-value, $\phi =$	
k	$b_r^{(k)}$	0.94	0.99	$b_r^{(k)}$	0.94	0.99	$b_r^{(k)}$	0.94	0.99	$b_r^{(k)}$	0.94	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

Table 7. Long-horizon forecasting regressions. In each case $b_r^{(k)}$ gives the point estimate in the data. "p-value" gives the percent probability value, i.e. the percentage of simulations in which the long-horizon regression coefficient $b_r^{(k)}$ exceeded the sample value $\hat{b}_r^{(k)}$. $\phi = 0.94, 0.99$ gives results for the two assumptions on 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, Table 7 shows that the central question is the horizon: conventional 5 and even 10 year horizons do not go far enough out to see the power advantages of long horizon regressions. To understand why long horizons help and why we need such long horizons, consider two and three-year return regressions

$$
\begin{aligned}
r_{t+1} + \rho r_{t+2} &= a_r^{(2)} + b_r^{(2)} x_t + \delta_{t+2}, \\
r_{t+1} + \rho r_{t+2} + \rho^2 r_{t+3} &= a_r^{(3)} + b_r^{(3)} x_t + \delta_{t+3}.
\end{aligned}
$$

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

$$b_r^{(3)} = b_r(1 + \rho\phi) \tag{21}$$

$$b_r^{(3)} = b_r(1 + \rho\phi + \rho^2\phi^2). \tag{22}$$

Thus, coefficients rise with horizon mechanically as a result of one-period forecastability and the autocorrelation ϕ of the forecasting variable (Campbell and Shiller 1991).

If we had a sample with a b_r as large as we see in the data, but with a ϕ smaller than we see in the data, then (21) and (22) show that we would not see as large long-horizon coefficients as we do in the data. Since large b_r tend to come with small ϕ, as seen in Figure 2, it is therefore much harder for the null to produce large long-horizon coefficients than it is for the null to produce large one-year coefficients. Long-horizon coefficients exploit the *joint* distribution of b_r and ϕ to increase power against the null.

The trouble is that this mechanism is not *quantitatively* strong for two, three, or even five year horizons, as their particular combinations of b_r and ϕ do not stress ϕ enough. To display this fact, Figure 6 plots again the joint distribution of (b_r, ϕ), together with lines that show rejection regions for long-horizon regression coefficients. The one-year horizon line is vertical line as before; the large number of points to the right of this line constitute the 22% probability value of the one-year regression. The 5 year horizon line is the set of (b_r, ϕ) points at which $b_r(1 + \rho\phi + .. + \rho^4\phi^4) = \hat{b}_r(1 + \rho\hat{\phi} + .. + \rho^4\hat{\phi}^4)$ where $\hat{b}_r$ and $\hat{\phi}$ are sample estimates. 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. The $k = \infty$ line is the set of (b_r, ϕ) points at which $b_r/(1 - \rho\phi) = \hat{b}_r/(1 - \rho\hat{\phi})$; 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.

As the figure shows, longer horizon regressions give more and more weight to ϕ. They therefore exclude more and more of the points that produce one-year coefficients b_r larger than seen in the sample, but together with dividend yield autocorrelation ϕ less than seen in the sample. The Figure shows clearly why one must consider such long horizons to exclude many points.

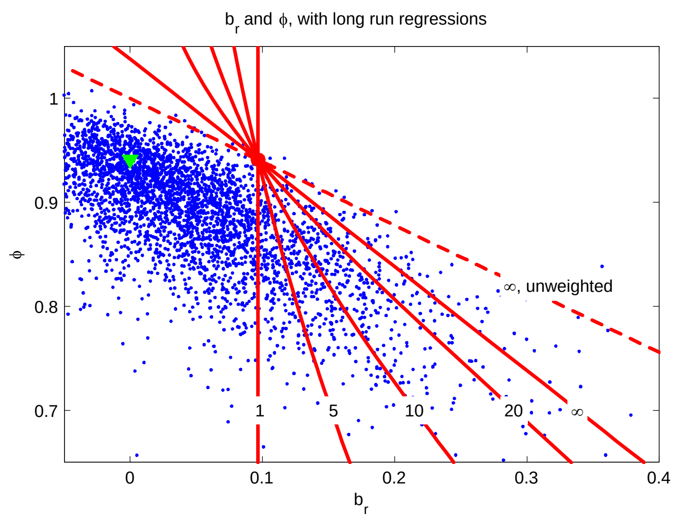

Figure 6: Joint distribution of b_r and ϕ estimates, together with regions implied by lonng-run regressions. The lines give the rejection regions implied by long-horizon return regressions at the indicated yearly horizon. For example, the points above and to the right of the line marked "2" are simulations in which coefficient $b_r^{(5)} = (1 + \rho\phi + ... + \rho^4\phi^4)b_r$ is higher than the value $\hat{b}_r^{(5)} = (1 + \rho\hat{\phi} + ... + \rho^4\hat{\phi}^4)\hat{b}_r$ in the data. The dashed line marked "∞, unweighted" plots the line where $b_r/(1 - \phi) = \hat{b}_r/(1 - \rho\hat{\phi})$ corresponding to the infinite-horizon unweighted regression.

7.2 Implications and non-implications

Table 7 shows that the distinction between weighted and unweighted long-horizon regressions makes little difference. Since $\rho = 0.96$ is close to one, that result is not surprising. Yes, the implied infinite-horizon unweighted regression coefficient does make sense. Even though the left hand variable and its variance explode, the coefficient converges to the finite value $b_r/(1 - \phi)$. Unweighted regressions value ϕ a little more than weighted regressions, producing a slightly quicker move to larger power.

Table 7 shows some interesting differences between *implied* and *direct* estimates, though the choice of horizon remains the biggest influence on power. In most cases, the direct estimates give less power against the null than the implied estimates. At a 5 year horizon, this degradation is enough that the direct estimates give less power than the 1 year horizon estimates, while the 5-year implied estimates show an improvement over the 1 year estimates. In a few cases, the direct estimates seem to better than the indirect estimates. However, this is a result of larger directly-estimated coefficients in our sample. For example the directly-estimated 15-year return coefficient is 1.38, while the implied coefficient is only 0.80. The null has to generate a much larger coefficient before rejecting, and even with the more imprecise measurement inherent in the direct regression, this is hard to do. If we set an even bar – the same sample coefficient – then the direct coefficients would show lower power than implied estimates in every case.

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

The comparison between direct and indirect estimates in Table 7 is not particularly general. Direct estimates, like kernel or nonparametric spectral density estimates, allow for unstructured temporal correlations. They generally give more sampling variation than implied estimates. Their advantage is that they measure the correct object if the long-run properties of the data are not captured well by a low-order time-series model. In simulations, the data is generated by the same first-order VAR we estimate, so Table 7 does not address the basic advantage of direct estimates. I argue in the next section that dividend yields and returns are, in fact, well represented by a low-order VAR, which argues in favor of the indirect estimates, but even that claim is special to this dataset.

How in general and in finite samples to balance the advantages and disadvantages of direct vs. indirect estimates is beyond the scope of this paper. First, I am not interested in long-horizon regression measurements per se. I am only interested in the implied long horizon estimates as useful summaries of the joint distribution of one-year coefficients b_r and ϕ. Second, the investigation would take us far afield because it must consider null hypotheses that are different from the fitted models. Finally, the point of Table 7 and this discussion is only to reconcile the power of implied long-horizon regressions with the literature that seems to find the opposite result, and to some extent to reconcile the conflicting claims of that literature.

The development of an "optimal" strategy is also beyond my scope. Which combination of b_r and ϕ or other statistics gives most evidence against the unforecastable-return null? I stop, having documented that the implied long-horizon coefficient $b_r/(1-\rho\phi)$ is enough to reject the null in our sample, without asking if there are other, even more powerful statistics. A serious investigation of that issue has also to consider more complex time-series models, rather than fine-tune a statistic for one particular parameterization of a VAR(1) and one sample size.

8 What about...

8.1 Long-horizon estimates and hidden dividend growth movements

Perhaps the VAR(1) structure is too limiting. Perhaps prices move on news of dividends several years in the future, news not seen in next year's dividend. Managers do smooth dividends, and many firms do not pay cash dividends, so the "dividend" is often a repurchase or cash acquisition that comes many years later. Imputing multi-year dividend growth forecastability from one-year forecastability and the dividend-yield autocorrelation may be severely constraining.

To address this question, I look more deeply at direct forecasts of long-horizon returns and dividend growth, regressions of the form

$$\begin{aligned}\sum_{j=1}^{k}\rho^{j-1}\Delta d_{t+j} &= a_d^{(k)}+b_d^{(k)}\left(d_t-p_t\right)+\varepsilon_{t+k}^{d}\\ \sum_{j=1}^{k}\rho^{j-1}r_{t+j} &= a_r^{(k)}+b_r^{(k)}\left(d_t-p_t\right)+\varepsilon_{t+k}^{r}.\end{aligned}$$

As with their infinite-horizon counterparts in Equations (15)-(17), these regressions amount to a variance decomposition for dividend yields. Start with the finitely-iterated version of identity (6),

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

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

$$\begin{aligned}var(d_t-p_t) &= cov\left(\sum_{j=1}^{k}\rho^{j-1}r_{t+j},d_t-p_t\right)-cov\left(\sum_{j=1}^{k}\rho^{j-1}\Delta d_{t+j},d_t-p_t\right)\\ &\quad +cov\left[\rho^{k+1}\left(d_{t+k+1}-p_{t+k+1}\right),d_t-p_t\right]\end{aligned}$$

Dividing by $var(d_t-p_t)$ we can express the variance decomposition in terms of regression coefficients,

$$1=b_r^{(k)}-b_d^{(k)}+b_{dp}^{(k+1)}. \tag{23}$$

Thus, we can read from the regression coefficients directly what fraction of the variance of dividend yields is due to k-period dividend growth forecasts, what fraction is due to k-period return forecasts, and what fraction is due to k-period forecasts of future dividend yields. As $k\rightarrow\infty$ and if the last term vanishes ($\phi<1/\rho$) we recover the identity $b_r^{lr}-b_d^{lr}=1$ studied in Section 4

Figure 7 presents direct estimates of long-horizon regression coefficients in equation (23) as a function of k. I do not calculate the last, future price-dividend ratio term as 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, e.g. $\sum_{j=1}^{k}\rho^{j-1}\Delta d_{t+j}$ on d_t-p_t. The rise in these estimates in the top panel 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, $\sum_{j=1}^{k}\rho^{j-1}\beta(\Delta d_{t+j},d_t-$

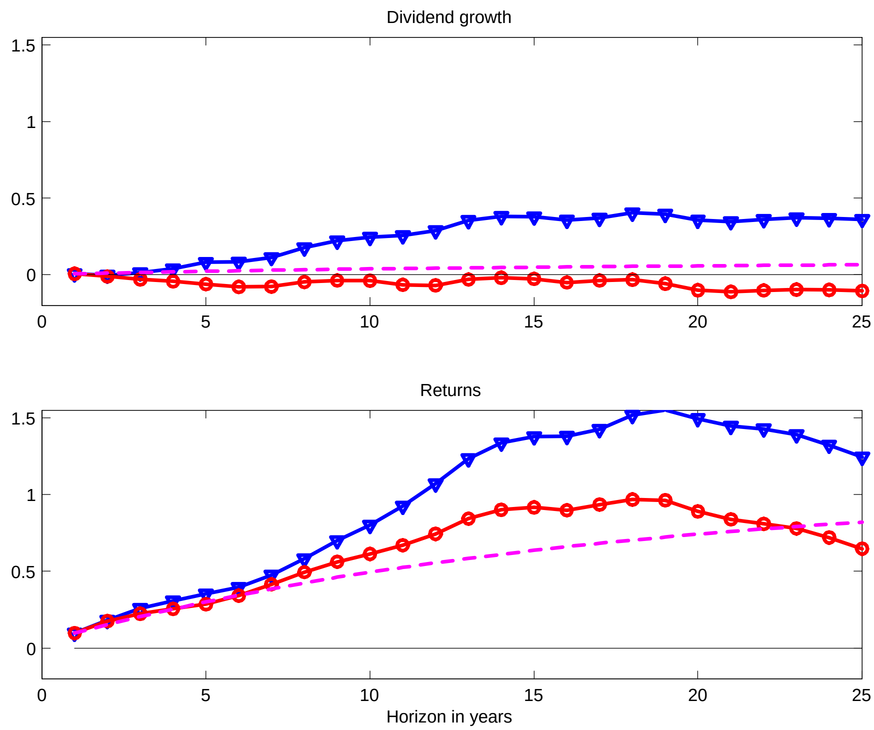

Figure 7: Regression forecasts of discounted dividend growth $\sum_{j=1}^{k}\rho^{j-1}\Delta d_{t+j}$ (top) and returns $\sum_{j=1}^{k}\rho^{j-1}r_{t+j}$ (bottom) on the log dividend yield $d_t - p_t$, as a function of the horizon k. Triangles are direct estimates, e.g. $\beta\left(\sum_{j=1}^{k}\rho^{j-1}\Delta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

p_t). This estimate only differs from the last one because it uses more data points. For example, the first year $\beta(\Delta d_{t+1}, d_t - p_t)$ in the 10-year k 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.

By contrast, the return forecasts account for essentially all dividend yield volatility once one looks out past 10 years. The regression coefficients approach and even exceed one. This (with a negative sign) is what 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.

Adding the long-run return and dividend-growth forecasts we see that the future dividend yield term in (23) is near zero past the 10 year horizon. The point estimate of return predictability is strong enough that we do not need bubbles to explain stock price variation in this direct estimates, just as we found with long-horizon regression coefficients implied by the VAR(1) in Table 4.

Despite the battering return forecasts b_r have taken 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 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. To the extent that there is any hidden multiperiod dividend growth forecastability not captured by the VAR system, it goes the "wrong" way. Though low order VAR systems do not always capture long-run dynamics well (for example, I find this to be true for GNP in Cochrane 1988), they seem to do so in this dataset.

(To keep the graph from getting too cluttered, and since our focus is on finite-sample distributions calculated from a simulation, I omit standard error bars from Figure 7. The best set of asymptotic standard errors I calculated gives the a return-forecast t-statistic of about two at all horizons. The dividend growth forecasts are completely insignificant.)

8.2 Repurchases

What about the fact that firms seem to smooth dividends, dividend payments seem to be declining in favor of repurchases, and dividend behavior may be shifting 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.

8.3 Additional variables

While there is nothing wrong with using the dividend yield, one *can* use variables that adjust for payout policies, 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 (Michaely, Richardson and Roberts 2006). For example Boudukh 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 (i.e. including repurchases). More generally, a large number of additional variables seem to forecast returns; for example see the summary in Goyal and Welch (2006). While we cannot fish across variables for t statistics anymore than we can fish across horizons, once we agree that dividend yields forecast returns, additional variables can only in the end add to the evidence for return forecastability.

Additional variables can also predict dividend growth (for example, Ribeiro 2004, Lettau and Ludvigson 2005). This fact does not imply that returns become less predictable. Our identities that dividend growth predictability and return predictability add up only apply to forecasts based on the dividend yield. Other variables can raise the predictability of dividend growth *and* of returns. To be specific, consider any set of forecasting variables that includes the dividend yield. The return identity (6) implies

$$d_t - p_t = E_t(r_{t+1}) - E_t(\Delta d_{t+1}) + \rho E_t(d_{t+1} - p_{t+1}) \tag{24}$$

and the present value identity (14) is

$$d_t - p_t = E_t\left(\sum_{j=1}^{\infty} \rho^{j-1} \Delta r_{t+j}\right) - E_t\left(\sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}\right). \tag{25}$$

Thus, by (24), if any variable helps to forecast one-period dividend growth, it *must* help to forecast returns, or help to forecast future dividend yields. By (25), if any variable helps to forecast long-run dividend growth, it must also *help* to forecast long-run returns. Again, considering more variables can only make the evidence for return predictability stronger.

9 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. Even if we convince ourselves that the return-forecasting evidence crystallized in Fama and French's (1988) regressions is statistically insignificant, we still leave unanswered the challenge crystallized by Shiller's (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 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 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.

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 cash-flow 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 variation can only be explained by rather nebulous variation in Sharpe ratios. But that is where the data have forced us, and they still do so. 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 reach throughout finance and are only beginning to be explored. The literature has focused on portfolio theory, i.e. 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, these calculations face a classic Catch-22: if there are more than measure zero of agents who should take the advice, the phenomenon will disappear. But if expected excess returns really do vary by as much as their average levels, 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, 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.

10 References

Ang, Andrew and Geert Bekaert, 2005, "Stock Return Predictability: Is it There?," Manuscript, Columbia University.

Barberis, Nicholas, 2000, "Investing for the Long Run when Returns are Predictable," *Journal of Finance* 55, 225-264.

Boudoukh, Jacob, and Matthew Richardson, 1993, "The Statistics of Long-Horizon Regressions," *Mathematical Finance* 4, 2, 103-120.

Boudoukh, Jacob, Matthew Richardson, and Robert F. Whitelaw, 2006, "The Myth of Long-Horizon Predictability," NBER Working Paper 11841, forthcoming *Review of Financial Studies.*

Boudoukh, Jacob, Roni Michaely, Matthew Richardson and Michael Roberts, 2006, "On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing" NBER Working paper #10651, forthcoming *Journal of Finance.*

Campbell, John Y., 2001 "Why Long Horizons? A Study of Power Against Persistent Alternatives," *Journal of Empirical Finance* 8, 459-491.

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

Campbell, John Y. and Samuel Thompson, 2005, "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?" NBER Working Paper No. 11468.

Campbell, John Y., and Motohiro Yogo, 2005, "Efficient Tests of Stock Return Predictability," forthcoming *Journal of Financial Economics.*

Cochrane, John H., 1988, "How Big is the Random Walk in GNP?" *Journal of Political Economy* 96, 893-920.

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

Cochrane, John H., 1992, "Explaining the Variance of Price-Dividend Ratios," *Review of Financial Studies* 5, 243-280.

Cochrane, John H., 2004, *Asset Pricing*, Revised Edition, Princeton: Princeton University Press.

Craine, Roger, 1993, "Rational Bubbles: A Test," *Journal of Economic Dynamics and Control* 17, 829-46.

Fama, Eugene F. and Kenneth R. French, 1988, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics* 22, 3-25.

Goetzmann, William N., and Philippe Jorion, 1993, "Testing the Predictive Power of Dividend Yields," *Journal of Finance* 48, 663-679.

Goyal, Amit and Ivo Welch, 2003, "Predicting the Equity Premium with Dividend Ratios," *Management Science* 49, 639-654.

Goyal, Amit and Ivo Welch, 2005, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," Manuscript, Brown University, Revision of NBER Working Paper 10483.

Hodrick, Robert J., 1992, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies* 5, 357-86.

Kandel, Shmuel and Robert F. Stambaugh, 1996, "On the Predictability of Stock Returns: an Asset-Allocation Perspective," *Journal of Finance* 51, 385-424.

Keim, Donald B. and Robert F. Stambaugh, 1986, "Predicting Returns in the Stock and Bond Markets," *Journal of Financial Economics* 17, 357-390.

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

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

Lewellen, Jonathan, 2004, "Predicting Returns with Financial Ratios," *Journal of Financial Economics* 74, 209-235.

Lettau, Martin, and Sydney Ludvigson, 2005, "Expected Returns and Expected Dividend Growth," *Journal of Financial Economics* 76, 583-626.

Mankiw, N.G., Shapiro, M., 1986. Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models," *Economic Letters* 20, 139–145.

Nelson, Charles R., and Myung J. Kim, 1993, "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance* 48, 641-661.

Paye, Bradley and Alan Timmermann, 2003, "Instability of Return Prediction Models," Manuscript, University of California at San Diego.

Pontiff, Jeffrey and Lawrence D. Schall, 1998, "Book-to-Market Ratios as Predictors of Market Returns," *Journal of Financial Economics* 49, 141-160.

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

Richardson, Matthew, and James Stock, 1989, "Drawing Inferences from Statistics Based on Multi-Year Asset Returns," *Journal of Financial Economics* 25, 323-348.

Rozeff, Michael S., 1984, "Dividend Yields are Equity Risk Premiums," *Journal of Portfolio Management* 11, 68-75.

Shiller, Robert J., 1984," Stock Prices and Social Dynamics," *Brookings Papers on Economic Activity* 2, 457-498.

Shiller, Robert J., 1981, "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71, 421-36.

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

Stambaugh, Robert F., 1999, "Predictive Regressions," *Journal of Financial Economics* 54, 375-421.

Torous, Walter, Rossen Valkanov, and Shu Yan, 2004, "On Predicting Stock Returns With Nearly Integrated Explanatory Variables," *Journal of Business* 77, 937-966.

Valkanov, R., 2003, "Long-Horizon Regressions: Theoretical Results and Applications," *Journal of Financial Economics*, 68, 2, 201-232.

11 Appendix

This Appendix documents the likelihood function plotted in Figure 4 and used in the last row of Table 5. The unconditional likelihood for an AR(1),

$$x_t = a + \phi x_{t-1} + \varepsilon_t,$$

is

$$L = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\frac{\sigma^2}{1-\phi^2}\right) - \frac{1}{2\sigma^2}\left(x_1 - \frac{a}{1-\phi}\right)^2\left(1-\phi^2\right) - \frac{T-1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\left(x_t - a - \phi x_{t-1}\right)^2.$$

The second and third terms penalize ϕ near 1. We can analytically maximize out a and σ^2 given ϕ. The first order conditions are

$$\begin{aligned} 0 &= \frac{\partial L}{\partial a} = \frac{1}{\sigma^2}\left(x_1 - \frac{a}{1-\phi}\right)\frac{(1-\phi^2)}{(1-\phi)} + \frac{1}{\sigma^2}\sum_{t=2}^{T}\left(x_t - a - \phi x_{t-1}\right) \\ a &= \frac{1}{T + 2\frac{\phi}{1-\phi}}\left[x_1(1+\phi) + \sum_{t=2}^{T}\left(x_t - \phi x_{t-1}\right)\right] \end{aligned}$$

$$\begin{aligned} 0 &= \frac{\partial L}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\left(x_1 - \frac{a}{1-\phi}\right)^2\left(1-\phi^2\right) - \frac{T-1}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=2}^{T}\left(x_t - a - \phi x_{t-1}\right)^2 \\ \sigma^2 &= \frac{1}{T}\left[\left(x_1 - \frac{a}{1-\phi}\right)^2\left(1-\phi^2\right) + \sum_{t=2}^{T}\left(x_t - a - \phi x_{t-1}\right)^2\right] \end{aligned}$$

Figure 4 uses these values of σ^2 and a for any given ϕ to plot the likelihood as a function of ϕ only.

The conditional likelihood function in Figure 4 is

$$L = -\frac{T-1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\left(x_t - a - \phi x_{t-1}\right)^2$$

For each ϕ I use the usual estimates of the other parameters,

$$\begin{aligned} 0 &= \frac{\partial L}{\partial \sigma^2} = -\frac{(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=2}^{T}\left(x_t - a - \phi x_{t-1}\right)^2 \\ \sigma^2 &= \frac{1}{T-1}\left[\sum_{t=2}^{T}\left(x_t - a - \phi x_{t-1}\right)^2\right] \end{aligned}$$

$$\begin{aligned} 0 &= \frac{\partial L}{\partial a} = \frac{1}{\sigma^2}\sum_{t=2}^{T}\left(x_t - a - \phi x_{t-1}\right) \\ a &= \frac{1}{T-1}\sum_{t=2}^{T}\left(x_t - \phi x_{t-1}\right). \end{aligned}$$