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
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

Table 1 presents regressions of the real and excess value-weighted stock return on its dividend-price ratio, in annual data. The results are quite similar using nominal returns, earnings yields or book to market ratios, and in postwar data.

	b	t	R^2(%)	$\sigma(\beta x)$(%)
R_{t+1}	3.39	2.28	5.8	4.9
$R - R^f$	4.00	2.62	7.6	5.8
D_{t+1}/D_t	0.07	0.06	0.00	0.001
r_{t+1}	0.097	1.92	4.0	4.0
$r - r^f$	0.110	2.18	5.1	4.5
Δd_{t+1}	0.008	0.18	0.00	0.003

Table 1. Regression of real CRSP value-weighted return and dividend growth on dividend/price ratio, 1926-2004. The top two rows regress gross returns and dividend growth on the D/P ratio; the bottom two rows regress log returns and log dividend growth on the log D/P ratio. $\sigma(\beta x)$ gives the standard deviation of the fitted value of the regression.

Returns seem predictable, and excess returns even more so. The point estimates have very large *economic* significance. The units in the top row are percent return on percent dividend yield. A coefficient of zero means that if the dividend yield goes up one percentage point, prices are expected to grow one percentage point less; the one percentage point lower expected capital gain matches the one percentage point higher dividend yield to give no change in expected return. A coefficient of one results if price expectations do not change at all. The one percentage point rise in dividend yield translates directly to a one percentage point higher return. The coefficient of *three to four* means that if dividend yields go up one percentage point, prices are expected to go *up* another two to three percentage points, strongly *reinforcing* rather than offsetting the change in dividend yield.

As another measure of economic significance, the last column of Table 1 presents the standard deviation of the fitted value of the regressions. The return-forecasting regression gives a standard deviation of expected returns of five to six percentage points. The *variation* in expected returns is almost as large as the *level* of the sample equity premium. Furthermore, as emphasized by Fama and French (1988), the coefficients and R^2 rise with horizon reaching values between 30 and 60 percent, depending on time period and estimation details. The large R^2 at long horizon is another measure of the large economic significance of return forecastability. Finally, if one calculates the fraction of the variance of the price-dividend ratio due to time-varying expected returns (discount rates), the point estimate of the return forecast shown in Table 1 neatly accounts for *all* variation in stock prices scaled by dividends, leaving no need or room for changing expected dividend growth or bubbles to affect price-dividend ratios. I present this calculation below.

However, the *statistical* significance of the first row is marginal. And the ink was hardly dry on the first studies[1] to run regressions like those of Table 1 before a large literature sprang

[1]Rozeff (1984), Shiller (1984), Keim and Stambaugh (1986), Campbell and Shiller (1987), and Fama and French (1988).

up examining their econometric properties and questioning their statistical significance. The right hand variable (dividend yield) is very persistent, and innovations in returns are highly correlated with innovations in dividend yields, since a change in prices moves both variables. As a result, the return-forecasting coefficient inherits near-unit-root properties of the dividend yield. It is biased upward, and its t-statistic is biased towards rejection. Goetzmann and Jorion (1993) and Nelson and Kim (1993) find the distribution of the return-forecasting coefficient by simulation, and find greatly reduced evidence for return forecastability. Stambaugh (1999) derives the finite-sample properties of the return-forecasting regression, showing the bias in the return forecast coefficient and the standard errors. In monthly regressions, Stambaugh finds that in place of OLS p-values of 6% (1927-1996) and 2% (1952-1996), the correct p-values are 17% and 15% – far from statistically significant.

More recently, Goyal and Welch (2003), (2005) show that return forecasts based on dividend yields and a menagerie 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.[2]

Does this evidence mean return forecastability is dead? No, and the key is in the *second* regression of Table 1. Dividends are clearly not forecastable at all. In fact, the small point estimate has 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 dividend/price ratio is a constant, which it surely is not. Alternatively, if 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 have to ask "*which* of dividend growth or returns is forecastable?" (Or really, "how much of each?") The null hypothesis for unforecastable returns must also raise the forecastability of dividend growth, and then it must also confront the *lack* of such forecastability in the data.

I set up such a null, and I evaluate the joint distribution of return and dividend-growth forecasting coefficients. I confirm that the return forecasting coefficient, taken alone, is not significant. Under the $b_r = 0$ null, we see return forecasts 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. The answer depends on specification, but the best overall number is about a 1-2% probability value (last row of Table 5). The important evidence, as in Sherlock Holmes' famous case, is the dog that does not bark.[3]

[2] Additional contributions include Kothari and Shanken (1997), Paye and Timmermann (2003), Torous, Valkanov and Yan (2004), Ang and Bekaert (2005), Richardson and Whitelaw (2006), and papers cited below.

A related literature finds studies whether long-horizon regressions capture any information not present in one-period regressions. Given the large persistence of the dividend yield and related forecasting variables, the answer is that, by and large, they do not. This is good news for my purpose, as I can focus entirely on one-period regression statistics. Important contributions include Campbell and Shiller (1988), Richardson and Stock (1989), Hodrick (1992), Boudoukh and Richardson (1993), Valkanov (2003), Boudokh Richardson and Whitelaw (2006).

[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," remarked Sherlock Holmes.
From "The Adventure of Silver Blaze" by Arthur Conan Doyle

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 reject the view that returns are forecastable. 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.

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

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

where b_r is the 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.

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, and it allows us to examine forecasting variables that do not include dividends. Larger b_r estimates typically come with lower ϕ estimates. This fact is driven by a strong negative correlation between return and dividend yield shocks. Thus, while under the $b_r = 0$ null we do often see b_r as high as in the data - the *marginal* distribution of b_r does not reject – almost all of those high b_r draws come with low ϕ draws. 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, one 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 return forecasts and to changing dividend growth forecasts respectively. They are also the 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. Tests based on long-run coefficients also reject the null with 1-2% probability values. Again, samples with high b_r typically have low ϕ. Therefore, they do not have particularly 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. They condense the joint distribution of b_r, ϕ into a single number, and capture in that number the observation that we do not see high b_r without high 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 part of the joint distribution. The question is, what set of events 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 b_r greater than that observed in the data, then many of the events in the rejection region have lower ϕ than in our data ((b_r, ϕ) distribution), or they have b_d much lower (large negative numbers) than in our data ((b_r, b_d) distribution). These events *do* have forecastable dividend growth, and dividend yields partially forecast by variation in dividend growth – their dogs do bark. The long-run coefficients reject decisively, because they put these events in the set that are "closer to the null" than the event we have seen. And rightly so.

Third, the identity (1) shows that we can in fact have both $b_r = 0$ and $b_d = 0$ if $\phi = 1/\rho \approx 1.04$. But this specification requires an explosive root in the dividend yield. Thus, the extra information about return forecastability from b_d or ϕ comes from prior information that $\phi < 1.04$, and 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.

2 Null hypothesis

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

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

Lagged returns and dividend yields 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

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

First, projecting on $d_t - p_t$, this identity implies that the regression coefficients obey the approximate identity

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

Second, it means that the errors in (3)-(5) obey

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

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

[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 $\rho = \frac{P/D}{1+P/D}$,

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

Iterating forward results in the present value identity (11).

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 those forecasts become evidence against the null as well. In particular, for a nonexplosive $\phi < 1/\rho \approx 1.04$, we cannot choose both $b_r = 0$ and $b_d = 0$. Fixing ϕ, 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.

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 term b_{rf} and ε_{rf} or simply understand the need to forecast "dividend growth" to include both terms.

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 return-forecasting coefficient is about $b_r \approx 0.10$, the dividend growth forecasting coefficient $b_d \approx 0$, and the OLS estimate of the dividend yield autocorrelation is about $\phi \approx 0.94$. The standard errors are about the same, 0.05 in each case.

Alas, the identity (7) is not exact. The "implied b" column of Table 2 gives each coefficient implied by the other two equations and the identity (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 and correlations. Returns have almost 20% standard deviation. Dividend growth has a large 14% standard deviation. In part this number comes from large variability in dividends in the prewar data. In part, the standard method for recovering dividends from the CRSP returns[5] means that dividends paid early in the year are reinvested at the market return to the end of the year. Return and dividend yield shocks are negatively correlated; price changes affect both variables. Dividend yield and dividend growth shocks are almost uncorrelated however, which will drive several differences between return - dp and dividend growth - dp systems.

The final columns of Table 2 present 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 in detail below. Given $b_r = 0$ and ϕ, the necessary dividend forecast

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

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

I then can find dividend growth by

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

Cochrane (1991) shows that this procedure implies that dividends paid early in the year are reinvested at the return R to the end of the year. Accumulating dividends at a different rate is an attractive 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.

coefficient b_d follows from the identity $b_d = \rho\phi - 1 + b_r$.

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

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. This is a pleasant choice since the errors are nearly uncorrelated, and with $b_d = 0$ the shocks are nicely interpretable as pure "expected return" and "cashflow" shocks (See Cochrane 2004 Ch. 20). However, the identity (6) holds well enough that this choice has almost no effect on the results. The null hypotheses thus takes the form

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

I use the sample estimate of the covariance matrix of ε^{dp} and ε^d. I simulate 5000 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

The null hypothesis can be given a deeper structural interpretation. Suppose that expected dividend growth follows an AR(1) process,

$$\Delta d_{t+1} = x_t + v_{t+1} \tag{9}$$
$$x_{t+1} = \phi x_t + \delta_{t+1}, \tag{10}$$

and that expected returns are constant. Using the Campbell-Shiller (1988) present value identity that results from iterating (6) forwards,

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

we then have

$$p_t - d_t = \frac{1}{1-\rho\phi} x_t. \tag{12}$$

From the identity (6), returns follow

$$r_{t+1} = \frac{\rho}{1-\rho\phi}\delta_{t+1} + v_{t+1} \tag{13}$$

Thus, (9)-(10) implies that dividend yields, returns, and dividend growth follow the representation

$$\begin{bmatrix} d_{t+1} - p_{t+1} \\ r_{t+1} \\ \Delta d_{t+1} \end{bmatrix} = \begin{bmatrix} \phi \\ 0 \\ \rho\phi - 1 \end{bmatrix} (d_t - p_t) + \begin{bmatrix} -\frac{1}{1-\rho\phi}\delta_{t+1} \\ \frac{\rho}{1-\rho\phi}\delta_{t+1} + v_{t+1} \\ v_{t+1} \end{bmatrix}. \tag{14}$$

We can recover the covariance of the "structural" shocks v, δ from the covariance matrix of the regression errors – σ_v, σ_δ and $\sigma_{v\delta}$ are exactly identified– and thus we can interpret the restricted regression with $b_r = 0$ and $b_d = \rho\phi - 1$ model as an instance of this structural model. The implied values are a very small innovation variance for expected dividend growth, $\sigma(\delta) = 0.008$, a quite large innovation variance for unexpected dividend growth $\sigma(v) = 0.144$ and a nearly zero correlation between the two $corr(v, \delta) = -0.073$.

Thus, we understand the strong positive correlation between return and dividend growth shocks in Table 2 as a *consequence* of these underlying uncorrelated expected dividend growth and ex-post dividend growth shocks, and the fact that unexpected dividend growth ε shocks quite naturally enter both the return and dividend growth (second two) equations in (14). We understand the strong negative correlation between return and dividend yield shocks in Table 2 as a *consequence* of the fact that the "present value" $1/(1-\rho\phi)$ of expected dividend growth shocks δ enters both dividend yield and return (first two) equations in (14). We understand the near-zero correlation of dividend-yield and dividend growth shock in Table 2 as a *consequence* of the fact that δ and ε are (sensibly) nearly uncorrelated, and appear separately in the dividend yield and dividend growth (first and last) equations of (14). The same error structure emerges if we specify that expected returns vary through time and expected dividend growth is constant.

This little calculation verifies that the null really can comes from a consistent view of the world in which expected returns are constant and changing expected dividend growth generates observed movements in dividend yields. It also verifies that it makes sense to change the forecasting coefficients b_r, b_d, ϕ and keep intact the error covariance structure.

3 Distribution of regression coefficients and t statistics

In each Monte Carlo draw I run regressions

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

Figure 1 plots the joint distribution of the return and dividend-forecast regression coefficients and 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

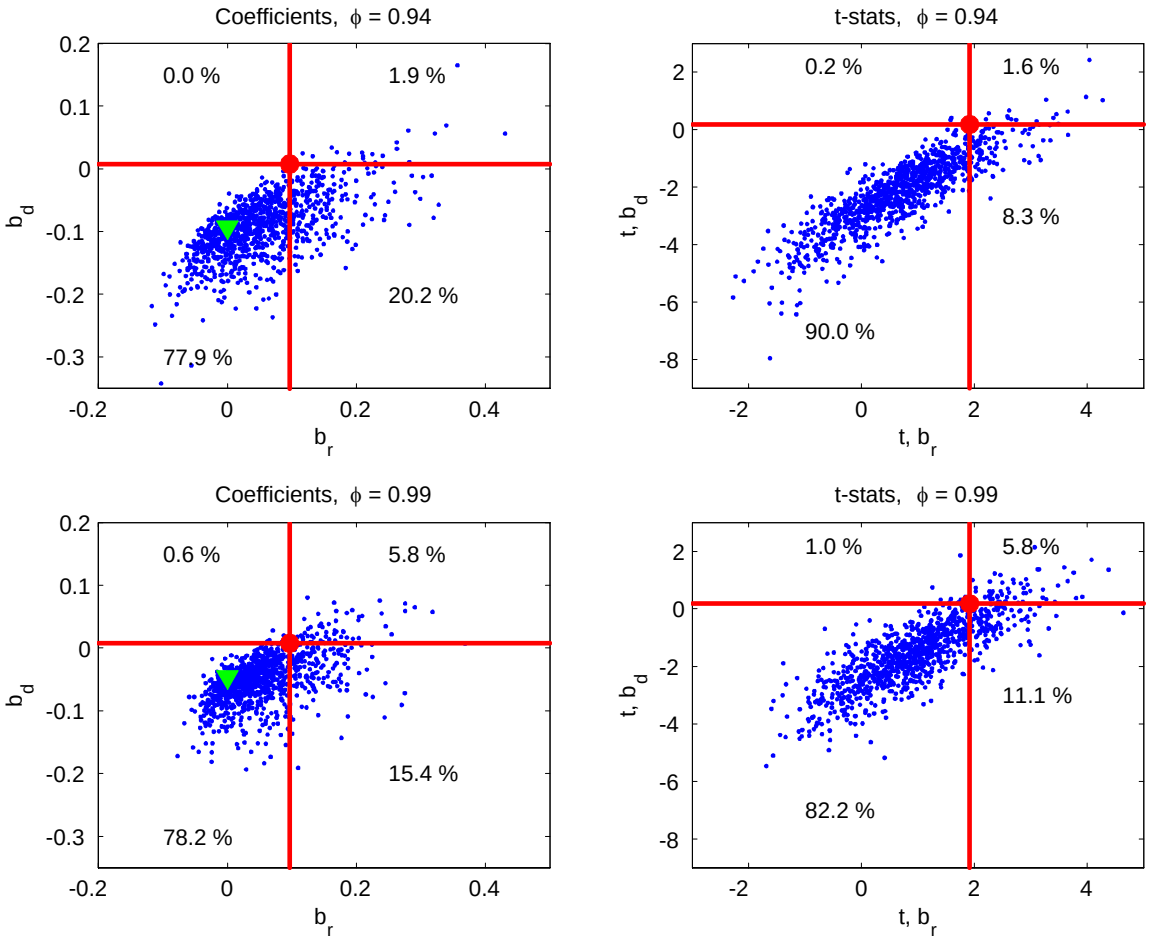

Figure 1: Joint distribution of return and dividend growth forecasting coefficients (left) and t-statistics (right). Red lines and dot give the sample estimates. The green triangle gives the null. 1000 simulations are plotted for clarity; each point represents 1/10 % probability. Percentages are the fraction of 5000 simulations that fall in the indicated quadrants.

roughly $\hat{b}_r \approx 0.10$ sample estimate 22% of the time, and a larger t statistic than the sample t = 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) This finding confirms the results of Goetzmann and Jorion (1993), Nelson and Kim (1993), and Stambaugh (1999).

However, the null must assume that dividends *are* forecastable. As a result, almost all simulations give a strong, negative dividend growth forecast coefficient b_d; the cloud of Figure 1 is vertically centered a good deal below zero and below the sample estimate $\hat{b}_d$. *Dividend* growth forecasting coefficients and t statistics larger than the roughly zero values observed in sample are only seen 1.90% of the time, and the t statistic is only greater than the sample value 1.76% of the time (points above the horizontal lines in Figure 1, b_d column of Table 3). Results are even stronger for excess returns, for which $b_d > \hat{b}_d$ is only observed 1.16% of the time and the t statistic only 0.82% of the time.

	b_r	t_r	b_d	t_d	b_r, b_d	b_r, ϕ	b_r, ϕ_{impl}
Real	22.1	9.88	1.90	1.76	1.88	0.04	0.00
Excess	17.1	5.60	1.16	0.82	1.16	0.02	0.00

Table 3. Percent probability values under the $\phi = 0.941$ null. Each column gives the probability that the indicated coefficients are greater than the sample values. Monte Carlo simulation of the null described in Table 2 with 5000 draws. $\phi_{impl} = 0.945$ uses the value of ϕ implied by the b_d and b_r estimates and the identity $b_r = 1 - \rho\phi + b_d$

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 for returns to the 1% range. (I discuss the $\phi = 0.99$ results in Figure 1 below.)

To emphasize this point, Figure 2 plots the *conditional* distribution of each forecast coefficient given that the other one comes out to its sample value. These are horizontal and vertical slices of the distributions in Figure 1 along the horizontal and vertical lines. The left hand panel of Figure 2 shows that *given* the dividend growth coefficient $b_d = \hat{b}_d$, the observed return coefficient $\hat{b}_r$ is not that surprising. However, the right hand panel of Figure 2 show that given the *return* coefficient, the (lack of) dividend forecast really is surprising. Given the estimate $\hat{b}_r \approx 0.1$, we should see most of the time a dividend growth forecast coefficient of approximately $b_d \approx -0.05$, and we only see coefficients above the approximately zero sample value 2% of the time.

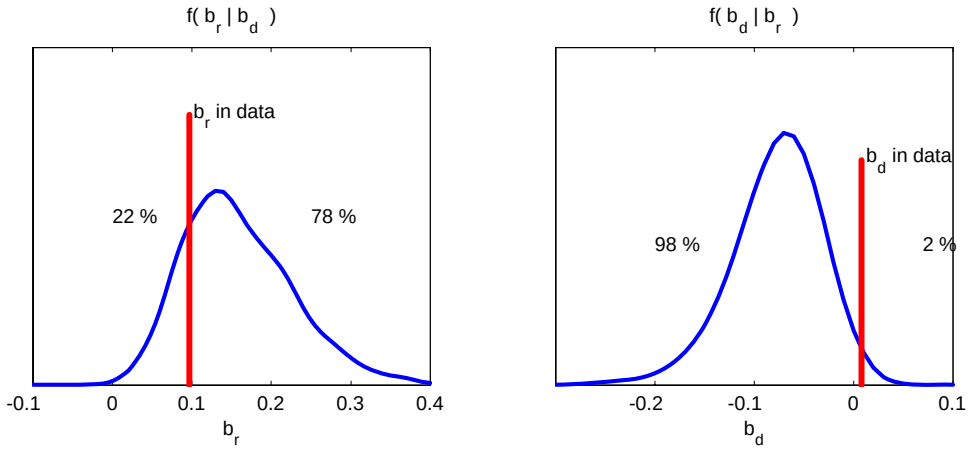

Figure 2: Conditional distribution of return forecast coefficients b_r given the dividend growth forecast coefficient b_d and vice versa.

3.1 The ϕ view

The return forecast coefficient, the dividend-yield autocorrelation and the dividend growth forecast coefficient are related by the approximate identity $b_r = 1 - \rho\phi + b_d$. Therefore, the exact same information in the joint distribution of return and dividend growth forecast coefficients

(b_r, b_d) is captured in the joint distribution of return and dividend yield forecast coefficients (b_r, ϕ), or dividend growth and dividend yield coefficients (b_d, ϕ), and it is useful to relate the three ways of looking at the data.

Recasting the point in the (b_r, ϕ) is context is especially important since the 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.

To address this question, the left-hand panels of Figure 3 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 20% of the time (area to the right of the vertical line). However, high return coefficients b_r almost always 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 North East quadrants. The exact numbers depend on whether one starts the tail region at the directly estimated sample value $\hat{\phi} = 0.941$ or the sample value of ϕ implied from $\hat{b}_d$ and $\hat{b}_r$, $\hat{\phi} = 0.945$. The probability of the joint region based on $\hat{\phi} = 0.941$ is 0.04%, i.e. two times in 5000 draws, while the probability based on $\phi > 0.945$ is 0.00%, i.e. never in 5000 draws. (The bottom left panel of Figure 3 is the same as Lewellen's (2004) Figure 1, Panel B except Lewellen calibrates to monthly postwar data. Lewellen also focuses on a different distributional calculation.)

The negative correlation between b_r and ϕ estimates is the key to this result. A lower sample value of ϕ, through $b_r = 1 - \rho\phi + b_d$ must correspond to a larger b_r, a larger, b_d or both. If the dividend yield reverts quickly after a shock, then it must be the case that one of dividend growth or prices and hence returns is unusually large after the shock, to bring the dividend yield back in line. In fact, lower ϕ are primarily largely associated with higher b_r, driven by the strong negative correlation between ϕ and b_r shocks seen in Table 2.

The diagonal dashed line marked b_d in the left panel of Figure 3 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 Figure 1. The comparison between the diagonal b_d line and the vertical b_r line shows the difference between looking at the return forecast b_r and the dividend forecast b_d in (b_r, ϕ) space. Given the strong negative correlation between b_r and ϕ, the b_d region above the diagonal line of Figure 3 excludes many points allowed by the vertical $b_r > \hat{b}_r$ region. In this way, the $b_d > \hat{b}_d$ test captures the intuition that the high b_r estimates null typically happen *with* low ϕ estimates, estimates not observed in our sample. Again, I discuss the $\phi = 0.99$ results below.

Looking at the $(b_r,\ \phi)$ system has the added 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 3 is an important component of the result. In turn, the correlation of estimates derives from the strong correlation between return and dividend yield shocks seen in Table 2, and that correlation between shocks emerges naturally in dynamic present value models such as the one sketched at the end of section 2, since a change in price moves both dividend yield and return. An arbitrary right hand variable, especially one that does not include price, is likely not to feature such a strong correlation.

The right hand panels of Figure 3 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 fact that the correlation

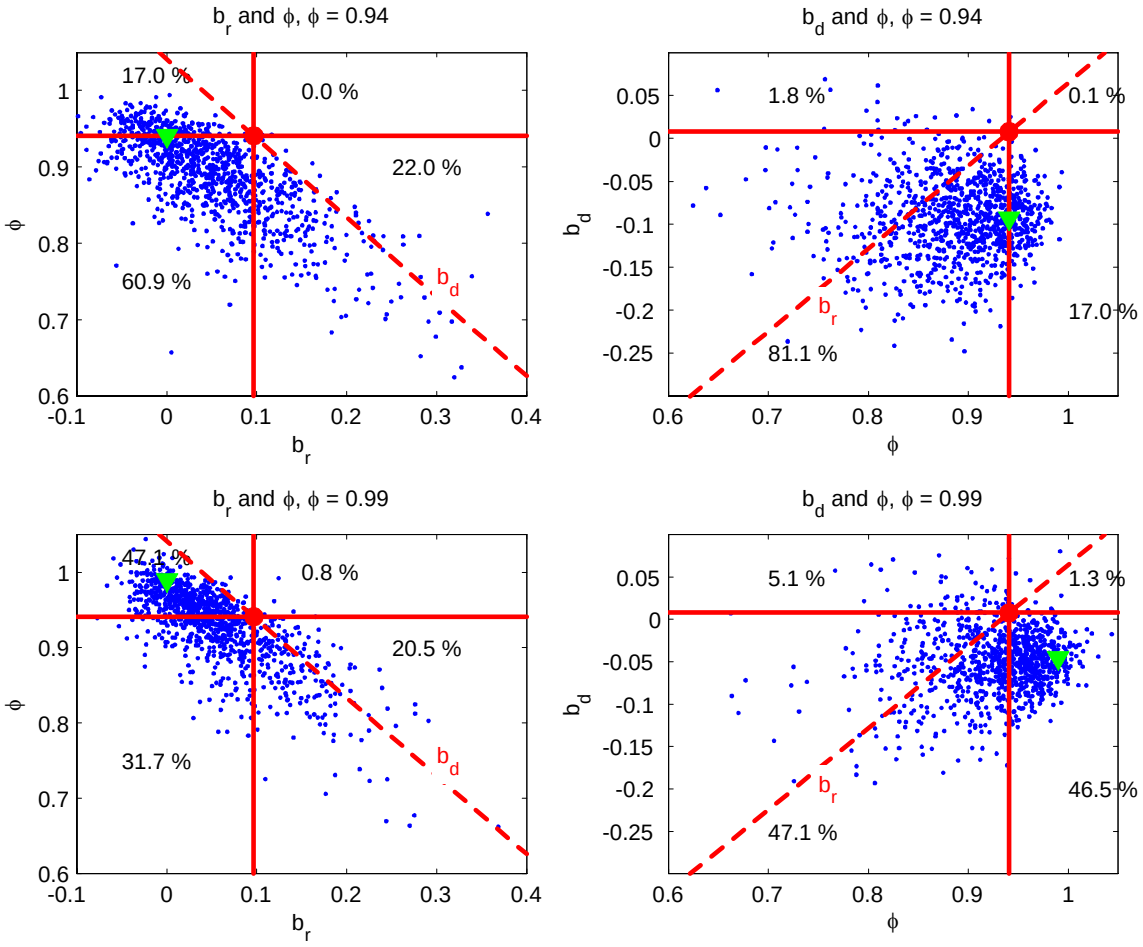

Figure 3: Joint distributions of regression coefficients. Left hand panels give the joint distribution of b_r, ϕ. Right hand panels give the joint distribution of b_d, ϕ. In each graph the triangle marks the null hypothesis used to generate the data and the circle marks the estimated coefficients $\hat{b}_r, \hat{b}_d, \hat{\phi}$. 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.

between dividend growth and dividend yield shocks is nearly zero, 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. The leftward smear of the cloud relative to the null (triangle) comes from the downward bias and large left tail of autocorrelation ϕ estimates, and the fact that lower ϕ estimates are via $b_r = 1 - \rho\phi + b_d$ allow the appearance of greater dividend growth forecastability than is really there. 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 North East corner, where we see a large dividend growth forecast *and* high sample autocorrelation.

4 Long-horizon coefficients

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

$$\begin{aligned} \frac{b_r}{1-\rho\phi} - \frac{b_d}{1-\rho\phi} &= 1 \\ b_r^{lr} - b_d^{lr} &= 1. \end{aligned} \tag{15}$$

The terms of this lovely identity have important interpretations. First, b_r^{lr} is the regression coefficients 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 points, iterate the return identity (6) forward, giving the Campbell-Shiller (1988) present value identity

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

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

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

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,

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

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

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

and similarly for dividend growth.

Negative b_d^{lr} is the fraction of dividend-yield volatility due to dividend growth, since b_d^{lr} and b_d should be negative. High dividend yields – low prices – should correspond to lower dividend growth. If a high dividend yield instead means higher dividend growth, expected returns must move even further to explain the 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; in fact with a single state variable (d-p) the components are perfectly correlated. 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 (1992), (2004) for more details on this variance decomposition.

Using the identity $b_r - b_d = 1 - \rho\phi$, we can also express the identity linking long-run coefficients (15) as

$$\frac{b_r}{b_r - b_d} - \frac{b_d}{b_r - b_d} = 1. \tag{16}$$

This equation expresses the same idea 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 and how much is in dividend growth.

Variable	b^{lr}	s.e.	t	% p
r	1.086	0.437	2.48	1.52
Δd	0.086	0.437	2.48	1.52

	b_r^{lr}		t		% p	
	min	max	min	max	min	max
Real	1.04	1.09	2.48	2.50	1.48	1.98
Excess	1.18	1.23	2.62	2.72	0.38	0.64

Table 4. Long-run regressions. b_r^{lr} is computed as $b_r^{lr} = b_r/(1-\rho\phi)$ where b_r is the regression coefficient of one year returns r_{t+1} on $d_t - p_t$, ϕ is the autocorrelation of $d_t - p_t$, $\rho = 0.961$, and similarly for b_d^{lr}. The standard error is calculated from standard errors for b_r and ϕ by the delta method. b_r, ϕ standard errors correct for heteroskedasticity. The t statistic for Δd is the statistic for the hypothesis $b_d^{lr} = -1$. Top panel entries are based on direct estimates of b_d, ϕ using b_r implied by the identity $b_r = 1 - \rho\phi + b_d$. The bottom panel gives the maximum and minimum values over the three choices of which two variables are estimated leaving the third implied. Probability values are generated by Monte Carlo under the $\phi = 0.941$ null.

Table 4 presents estimates of the long-horizon regression coefficients. I calculate standard errors[6] and a t statistic based on the OLS standard errors for the underlying coefficients b_r, b_d, ϕ. The top panel of 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

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

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

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

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

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 way.

The top panel of Table 4 drives 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. 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.

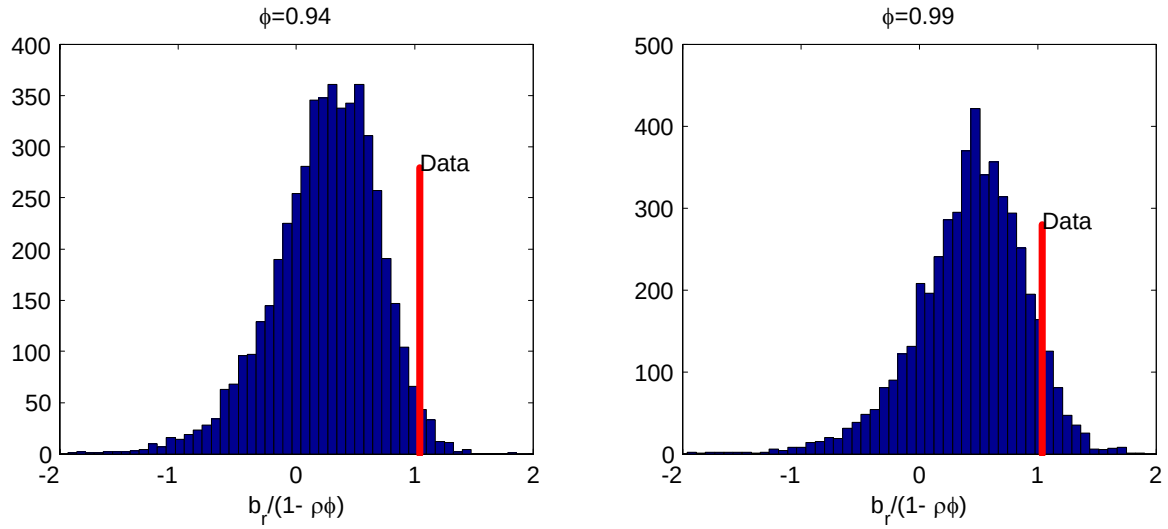

Figure 4: Distribution of $b_r/(1-\rho\phi)$, the fraction of dividend-yield variance explained by return forecasts, and the implied coefficient of long run returns $\sum \rho^{j-1} r_{t+j}$ on dividend yields. The vertical bar gives the corresponding value in the data.

Figure 4 tabulates the small-sample distribution of the long-run return and dividend growth estimates, and the bottom panel of Table 4 includes the probability values, i.e. how many long-run return forecasts are greater than the sample value under the null $b_{lr}^r = 0$. By the identity $b_r^{lr} - b_d^{lr} = 1$, these are exactly the same as how many long-run dividend growth forecasts are greater than the sample value under the null $b_{lr}^d = -1$. The finite sample distribution gives a 1.52% probability value of seeing larger b_r^{lr} (or $b_d^{lr} - 1$). Comparing this value to the 20% or so probability values for $b_r > \hat{b}_r$, and 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 3 that b_r is large predominantly in samples in which ϕ is low. When ϕ is low, however, $1 - \rho\phi$ is large, so $b_r^{lr} = b^r/(1-\rho\phi)$ is not so large. Thus, the long-run coefficient captures the point of the joint b_r, ϕ distribution of Figure 3.

Similarly, we saw in Figure 1 that large b_r usually come with small (large negative) b_d. In the context of (16), 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. Of course, since $b_r - b_d = 1 - \rho\phi$ this is the same observation, as the only way both returns and dividend growth can be more forecastable in the right direction is for dividend yields to have a lower autocorrelation.

For the identity $b_r^{lr} - b_d^{lr} = 1$ to hold exactly, one must use estimates for which the return identity holds exactly, implying one of b_r, b_d or ϕ from the other two. The first two rows of Table 4 present results calculated from b_d and ϕ, implying b_r. The first row of the bottom panel shows the maximum and minimum values over the three ways of making the calculation, i.e. implying each variable in turn from estimates of the other two. The coefficients change slightly, from 1.04 to 1.09, resulting in just enough of a difference in probability values, 1.48% to 1.98%, to merit showing the range of variation.

The last row of Table 4 shows the results for excess returns. Again, excess returns paint a stronger picture. Returns are more forecastable, and correspondingly dividend growth less interest rates go further in the wrong direction, accounting now for -18% to -23% of dividend yield variation. The probability values of 0.38% - 0.64% for the test $b_r^{lr} = 0$ are correspondingly lower.

4.1 The advantages of long-run coefficients

Recasting the problem in terms of the long-run coefficients b_r^{lr} and b_d^{lr} provides the most 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 linear combination or 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 that is an upper right quadrant, since one would not likely commit to *accepting* the null with an arbitrarily large b_r but b_d or ϕ just below some preannounced value.

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 $b_r = \hat{b}_r \approx 0.1$, and "more extreme" events as $b_r > \hat{b}_r$. But, as the joint distributions point out, most of the events with $b_r > \hat{b}_r$ have $b_d < \hat{b}_d \approx 0$ or low values of $\phi < \hat{\phi}$. In these events, dividend growth *is* forecastable and *does* count for an often substantial portion of dividend yield variation. For example, we might see $\phi = 0.8$, $b_d = -0.11$ and $b_r = 1 - 0.96 \times 0.80 - 0.11 = 0.12$. This $b_r = 0.12$ is greater than $\hat{b}_r \approx 0.10$ seen in our data, so conventionally counts as a "more extreme" event. But in this draw, a rise in the dividend yield corresponds about half and half to future dividend growth and future returns; volatility tests are a half-success rather than the total failure they are in our data. Is this really a "more extreme" event, further from the unpredictable-return null than

what we have seen in our data? Or, should we instead count this event as being much closer to the null than the event in our data? The latter seems much more plausible, and that is how the long-run coefficient counts things. Characterizing the null as $b_d = -0.1$ leads to similar problems, since a different ϕ leads to $b_r \neq 0$. $b_d^{lr} = -1$ is the same as $b_r^{lr} = 0$ for any value of ϕ.

Furthermore, since the long-run coefficients obey the identities (15) and (16), 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 $b_d > \hat{b}_d$, but in fact they are different conceptually and slightly different in this sample. $b_d^{lr} > \hat{b}_d^{lr}$ 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 b_d^{lr}, perhaps pushing it across a boundary, for the same value of b_d. Events with $b_d > \hat{b}_d$ can have a variance decompositions closer to the null than is our sample. It is just 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 ϕ. One naturally wants to know how the results are affected by the choice of ϕ, especially larger values given the downward bias in autocorrelation estimates.

	Percent probability values										Statistics	
Null	Real					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	23.6	0.64	0.00	0.34	0.58	19.3	0.34	0.00	0.08	0.18	0.35	6.6
0.941	22.2	1.60	0.06	1.20	1.68	17.5	0.96	0.00	0.36	0.58	0.45	11.4
0.96	21.7	2.58	0.08	2.02	2.80	17.0	1.52	0.02	0.76	1.04	0.55	17.0
0.98	21.2	4.92	0.42	4.30	5.50	15.9	2.92	0.20	1.80	2.54	0.77	34.3
0.99	21.3	6.28	0.76	5.86	7.40	16.0	3.44	0.34	2.86	3.56	1.09	69.0
1.00	22.2	8.66	1.00	8.06	10.10	17.1	4.82	0.56	3.86	4.94	∞	∞
1.01	19.6	11.00	1.46	10.72	12.94	15.0	5.40	0.70	5.14	6.60	∞	∞
Draw ϕ	23.1	1.64	0.10	1.40	1.70	18.2	0.96	0.04	0.70	0.84		

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 5. "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 $b_r/(1-\rho\phi)$. $b^{lr}_{\min}$ 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. $\sigma(dp)$ gives the implied standard deviation of the dividend yield $\sigma_{\varepsilon,dp}/\sqrt{1-\phi^2}$. Half life is the value of t such that $\phi^t = 1/2$.

Table 5 collects the probability values for various events as a function of ϕ. The previous figures include the case $\phi = 0.99$, to illustrate the effects of changing ϕ on the distribution of statistics.

As ϕ rises, the identity $b_r = 1 - \rho\phi + b_d$ requires larger (less negative) b_d in the null $b_r = 0$. At the sample $\phi = 0.94$, we needed $b_d \approx -0.1$. 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 $b_d > \hat{b}_d$ 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, and the cloud of points rises with it, so the chance of seeing $b_d > \hat{b}_d$ rises as well. This rise has little effect on the b_r statistic, which is about 20% for all values of ϕ. However, the cloud doesn't rise much, and its shape is changed reflecting more severe small-sample biases. Looking down the b_r and b_d columns of Table 5, the 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, b_d still has more information, with less than half the probability value of the b_r region.

The joint distributions of Figure 3 and the corresponding probability values b_r, ϕ in Table 5 show a similar pattern. In the left hand (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. Again, raising ϕ has little effect on the number of points to the right of the vertical $b_r = \hat{b}_r$ line, which is why these probability values stay put at about 20%. Raising ϕ does put more points above the diagonal b_d line, but again not that much, and still almost no points in the joint b_r, ϕ region.

The probability values of the more attractive long-run coefficients $b^{lr} = b/(1 - \rho\phi)$ also rise with ϕ. 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 stronger than the b_r evidence at any ϕ.

5.1 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 (11) from the return identity (6) is

$$p_t - d_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} + \lim_{k \to \infty} \rho^k E_t \left(p_{t+k} - d_{t+k} \right) \tag{17}$$

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

If we have $\phi = 1/\rho = 1/0.96 \approx 1.04$, 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 (17). 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 $\phi = 1$, dividend yields are a random walk. $\phi = 1$ still requires 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 translates directly to a higher return. Alternatively, if returns are unchanged, lower dividend growth must offset the higher dividend yield. $\phi = 1$ does not cause much trouble for the present value model; that blows up at $\phi = 1/\rho \approx 1.04$. $\phi = 1$ 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 $\hat{\phi} = 0.94$ that statistical evidence will naturally be marginal. 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 a that basis.

A random walk does not fit long-run evidence. 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[7] 5.6 and 112, and the ± 2 standard deviation band is between[8] 1.24 and 502. In 300 years, the bands are similarly $\pm 1\sigma = (1.9 - 336)$, $\pm 2\sigma = (0.14 - 4514)$. If dividend yields are a random walk we should have seen observations of this sort, but market price-dividend ratios of two or several hundred have never been approached.

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 argument rests 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 just 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 one or 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

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

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

is predictable, with probability values in the 1% to 5% range.

5.2 An overall number

Results as a function of ϕ and then thoughts about upper limits for ϕ are not that satisfying in the end. To produce a single number, one wants to integrate over possible values of ϕ with a prior on ϕ. The last row of Table 5 presents this calculation, using the unconditional maximum likelihood of ϕ as the density.

Figure 5 presents the unconditional likelihood function for the autoregressive parameter ϕ of the dividend yield. I maximize out the other parameters, the intercept a_{dp} and the innovation variance $\sigma^2_{\varepsilon,dp}$. (Details are in the appendix.) I use the unconditional likelihood (the usual likelihood function plus the likelihood of the first data point) in order to impose the view that dividend yields are stationary with a finite variance, $\phi < 1$, since the unconditional likelihood function goes to 0 at $\phi = 1$.

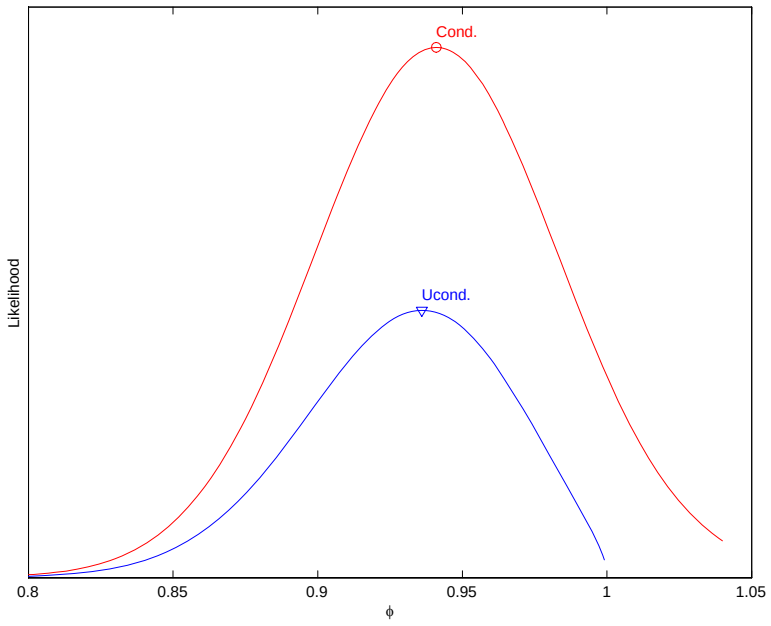

Figure 5: Likelihood function for ϕ, the autoregressive parameter for dividend yields. The intercept a_{dp} and innovation variance $\sigma^2_{\varepsilon dp}$ are maximized out

Next, I repeat the simulation but this time drawing ϕ from the unconditional likelihood plotted in Figure 5 before drawing a sample of errors ε_t^{dp} and ε_t^d. The last row of Table 5 summarizes the results. (The graphs do not look all that much different than the ones shown so far.) As one might expect from a visual integration of Table 5, the results are quite similar to the $\phi = 0.94$ case. Most importantly, rather than a 23.1% chance of seeing $b_r > \hat{b}_r$, we can reject the null based on a 1.64% chance of seeing $b_d > \hat{b}_d$ or the 1.40-1.70% chance of seeing the more elegant long-run regression coefficients b_r^{lr} or b_d^{lr} greater than their sample values. As usual, excess returns give even stronger rejections, with $b_d > \hat{b}_d$ occurring 0.96% of the time, and the long-run b_r^{lr} test only 0.70-0.84% of the time. (Lewellen 2004 presents a similar calculation, also

delivering small probability values.)

Figure 5 and the fact that numbers behave smoothly across the $\phi = 1$ boundary in Table 5 actually suggest that the results will not be all that different if one chooses ϕ from the conditional likelihood function, allowing the possibility that $\phi \geq 1$, but ignoring the information in the first data point if $\phi < 1$. Most of the weight of the density is in fact below $\phi = 1$. The only hitch is what to do about the small probabilities that $\phi \geq 1/\rho \approx 1.04$, where the price-dividend ratio becomes infinite.

5.3 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 ϕ estimate is downward biased. The return forecast coefficient b_r is upward biased, with a bias of approximately 0.05 accounting 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. The dividend growth coefficient however is not biased. As seen in of Figure 3, 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. 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 b_d^{lr} is biased up as well. However, the strong rejections of $b_r^{lr} = 0$ or equivalently $b_d^{lr} = -1$ are a sign that the coefficients are well enough measured that we can distinguish the biased value of $b_r^{lr} = 0.24 - 0.42$ from the sample value of $b_r^{lr} \approx 1$. The main source of bias here is the downward bias in ϕ, which induces a downward bias in the measured variance of dividend yields. Therefore $b_d/(1 - \rho\phi)$ is biased even though b_d is not.

		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	-0.096	0.886	0.24	-0.76
$\phi = 0.99$	Null	0	-0.046	0.990	0	-1
	Mean	0.056	-0.050	0.927	0.42	-0.58

Table 6. Means of estimated parameters. Means are taken over 5000 simulations of the Monte Carlo described in Table 2.

6 Out-of-sample R^2

Goyal and Welch (2005) show in a careful and comprehensive study that dividend yield and just about every other regressor thought to forecast returns does 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.

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 with $b_d = 0$. Let expected *returns* vary through time,

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

(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,

$$\Delta d_{t+1} = \varepsilon_{t+1}.$$

Imposing the Campbell-Shiller identity (11), we have

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

Returns follow

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

Thus, we have a VAR representation

$$\begin{bmatrix} d_{t+1} - p_{t+1} \\ r_{t+1} \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 1-\rho\phi & 0 \end{bmatrix} \begin{bmatrix} d_{t+1} - p_{t+1} \\ r_{t+1} \end{bmatrix} + \begin{bmatrix} -\frac{1}{1-\rho\phi}\delta_{t+1} \\ \frac{\rho}{1-\rho\phi}\delta_{t+1} + \varepsilon_{t+1} \end{bmatrix}. \tag{18}$$

This is exactly the same VAR as before but with a $1-\rho\phi$ in the return forecast slot rather than zero.

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, using $b_r = 1 - \rho\phi$. 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 6 shows the distribution of this statistic across simulations. In the data, marked by the vertical "Data" line, the statistic is negative; the sample mean is a better 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, the dividend yield is an even worse "out of sample" forecaster than it is in the observed data. In fact, the *mean* of the 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.

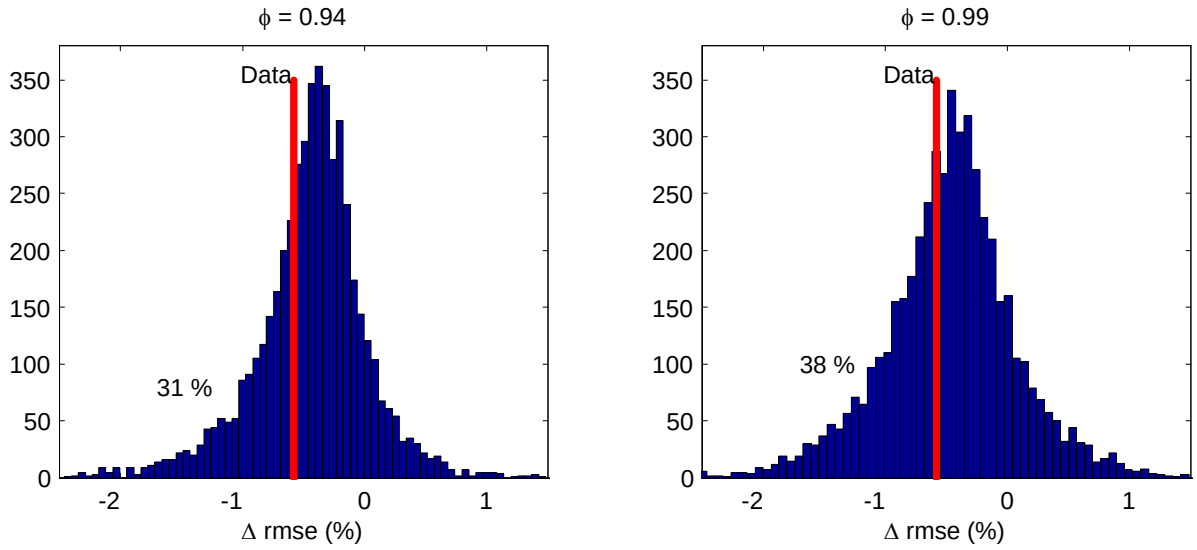

Figure 6: Distribution of the Goyal-Welch statistic under the null that returns are forecastable and dividend growth is not forecastable. The statistic is the root mean squared error from using the sample mean return to forecast returns, less the root mean squared error from using a dividend yield regression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

I conclude that 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. Also, one might think that the null the sample mean would do poorly, reasoning that with predictability high past returns would signal low future returns. However, though under this null though returns are predictable from dividend yields, returns are essentially unpredictable from past returns, so the sample mean does not lead one astray in this way.

6.1 Reconciliation

Both views are right. Goyal and Welch's message is that regressions using dividend yields and other 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 with highly persistent right hand variables. 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 not an unusually powerful statistic that gives stronger evidence about return forecastability than the regression coefficients. 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 What about...

7.1 Long-horizon estimates

So far, I have imposed a VAR(1) structure on the null with unforecastable returns. Perhaps this restriction is too limiting. Perhaps prices move on news of dividends several years in the future, news not seen in next year's dividend. After all, we know managers smooth dividends, so imputing the multi-year dividend growth forecastability from the forecastability one year ahead may be severely constraining.

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

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

Individual regressions $\Delta d_{t+j} = a + b (d_t - p_t) + \varepsilon_{t+k}$ paint a similar picture, and the long-run regressions are of course partial sums of such individual regressions. Again, these regressions amount to a variance decomposition for dividend yields of the type studied by Cochrane (1992). Start with the finitely-iterated version of identity (6),

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

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

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

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

$$1 = b_r^{(k)} - b_{\Delta d}^{(k)} + b_{dp}^{(k+1)}. \tag{19}$$

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 $k \to \infty$ and if the last term vanishes ($\phi < 1/\rho$) we recover the identity $b_r^{lr} - b_d^{lr} = 1$ studied above.

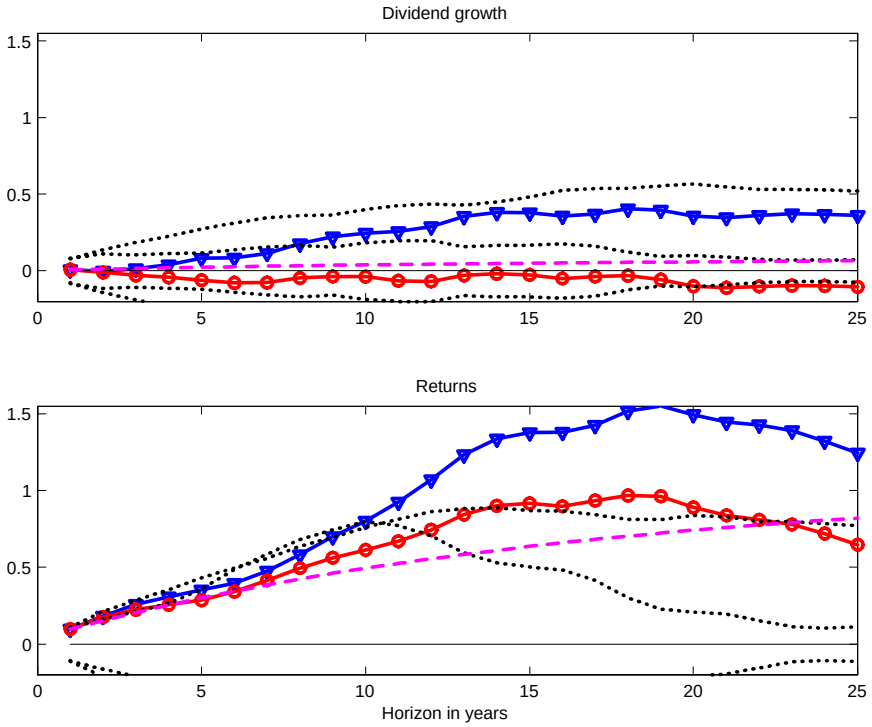

Figure 7: Regression forecasts of discounted dividend growth $\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}$ (top) and returns $\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$ (bottom) on the log dividend yield $d_t - p_t$, as a function of the horizon k. Triangles are direct estimates. Circles sum individual estimates, e.g. $\sum_{j=1}^{k} \rho^{j-1} \beta \left(\Delta d_{t+j}, d_t - p_t \right)$. The dashed line is the value implied by the VAR, e.g. $\sum_{j=1}^{k} \rho^{j-1} \phi^{j-1} b_d$. Dots are +/- two standard errors from zero. The tighter set use a Newey-West correction with lags = twice the horizon. The larger set impose the null and homoskedasticity to avoid "nonaparametric" standard error estimation.

Figure 7 presents direct estimates of long-horizon regression coefficients in (19) as a function of k. I do not calculate the last, future price-dividend ratio term as it is implied by the other two.

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, $\sum_{j=1}^{k} \rho^{j-1} \beta(\Delta d_{t+j}, d_t - p_t)$. This estimate only differs 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. This (with a negative sign) is what dividend forecasts *should* look like if we are to hope that they explain price variation, and they do not come close, even in these direct estimates.

The return forecast coefficient and thus the sum of coefficients is close to one past the 10 year horizon. There is no identity forcing return and dividend coefficients to add up here, as the future dividend yield term in (19) could enter as well. In fact, we do not need bubbles.

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 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 present the long-run coefficients implied by the VAR, $\sum_{j=1}^{k} \rho^{j-1} b_r \phi^{j-1} = b_r \frac{1-(\rho\phi)^k}{1-\rho\phi}$ 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. Time-series empirical work is full of examples in which direct long-horizon estimates give quite different answers from those implied by models fit to short-run properties of the data, for example Cochrane (1988). This case does not appear to be one of them.

The dotted standard errors in Figure 7 use the Newey-West scheme with lags equal to twice the horizon in order to control for serial correlation due to overlap. I use the Newey-West scheme because the standard Hansen-Hodrick correction with lags equal to the horizon yields negative variances in some instances. These standard errors struck me as suspiciously optimistic, especially the apparent increase in precision with horizon in the lower panel of Figure 7. "Nonparametric" estimates can perform poorly in small samples, especially when using up to 50 lags in a 77 year sample. In an attempt to provide a bit more trustworthy standard errors, the dashed standard errors impose the null that one period returns or dividend growth are i.i.d. (independent of current and past dividend yields and independent of past returns or dividend growth) in order to estimate the spectral density matrix. This assumption produces a much simplified spectral density matrix, which should result in better small sample performance. The calculation is in the Appendix. The dashed standard errors show the return forecasts to have about the same significance at all horizons, which is the message of the econometric literature that investigates long-horizon forecasts. They show that the long-horizon dividend forecasts are completely insignificant.

7.2 Hidden long-run movements

We cannot rule out a null hypothesis that prices are driven by news of extremely far-distant dividend growth, that the real decompositions change places after the 25 years shown in Figure 7. For example, we might suppose that dividend growth exhibits rare "structural breaks," and prices vary on varying assessments of the probability of such a break. Though we do not have any evidence *for* such long-run dividend growth forecastability, we don't have much evidence *against* it either, so this null cannot be rejected.

By itself, this is an unsatisfactory solution. Explaining price variation by far-off dividend forecasts with no independent measurement of those forecasts is really no different from explaining price variation by fads and fashions. The only way to make either idea respectable is to find some *independent* confirmation of the event. Noone has yet suggested a way to independently confirm that expectations of long-term dividend forecasts are moving.

7.3 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. If the firm pays no dividends, ever, as measured by CRSP, then the stock is worthless. 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.

The danger posed by repurchases, then, is another possibility of long-delayed dividend growth. Prices may move on news of future cashflows, and those cashflows do eventually work their way into measured dividends, but it takes so long that we do not measure the correlation between prices and eventual dividends even in 25 years. Again, we need some independent measurement for this view to rescue the idea that dividend growth is forecastable and returns are not.

8 Conclusion

If returns really are *not* forecastable, then dividend growth must *be* forecastable in order to generate 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-to-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 varying expected dividend growth does explain the variation of dividend yields, I can check both dividend and return forecastability. I find that the *absence* of dividend growth forecastability in our data provides much stronger evidence against the null than does the presence of return forecastability, with probability values in the 1-2% range rather than in the 20% range.

The long-run coefficient $b_r^{lr} = b_r/(1-\rho\phi) = b_r/(b_r - b_d)$ captures these observations in a single number, and ties 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 coefficients also give 1-2% rejections.

The stronger rejection comes from a different view of what events are "more extreme" than the one seen in our data. Many samples with higher return forecasting coefficients b_r than we have seen also come with much greater dividend forecastabilty than we have seen (large negative b_d or small ϕ). In these samples, some or even a lot of dividend yield variation is accounted for by dividend growth forecasts. The conventional $b_r > \hat{b}_r$ test counts these samples as "more extreme," in the rejection region. Tests based on the dividend growth coefficient or the long-run coefficients count these events as "closer to the null" thus delivering the smaller larger probability values for events that really are "more extreme" than our data.

I have concentrated on dividend yields for simplicity and to give the tightest interpretation of the alternative – if returns are not predictable, then something else must be. Other variables do predict dividend growth (Ribeiro 2004, Lettau and Ludvigson 2005), but they also predict returns. Adding more variables can only make the evidence stronger.

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, at least high prices could 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 in all of this 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 value.

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 who are not affected by the change in risk or risk aversion that drives excess return forecastability can benefit by market-timing portfolio rules. However, the signals are slow-moving, really affecting the static portfolio choices of different generations rather than dynamic portfolio choices of short-run investors, parameter uncertainty greatly reduces the potential benefit, and these calculations face the classic Catch-22: if there are more than measure zero of agents who take the advice (and you don't find a corresponding measure who want to move in the opposite direction), 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 that premium, and more deeply recognizing likely changes in that premium over the lifespan of a project and the multiple pricing factors that predictability implies.

9 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.

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.

10 Appendix

10.1 Likelihood for an AR(1)

The unconditional likelihood for an AR(1),

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

is

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

The second and third terms penalize ϕ near 1. There is no full analytic solution, but we can analytically maximize out a and σ^2 given ϕ. The derivatives are

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

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

Figure 5 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 5 is

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

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

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

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

10.2 Simple standard errors for long-horizon forecasts

The general GMM formula for standard errors of OLS regressions

$$y_t = \beta x_t + \varepsilon_t$$

is (see Cochrane 2004)

$$var(\hat{\beta}) = \frac{1}{T} E(x_t x_t')^{-1} \sum_{j=-\infty}^{\infty} E\left(\varepsilon_t x_t x_{t-j}' \varepsilon_{t-j}\right) E(x_t x_t')^{-1}.$$

Applied to a forecasting regression

$$y_{t+k} = \beta x_t + v_{t+k}$$

we have

$$var(\hat{\beta}) = \frac{1}{T} E(x_t x_t')^{-1} \sum_{j=-\infty}^{\infty} E\left(v_{t+k} x_t x_{t-j}' v_{t-j+k}\right) E(x_t x_t')^{-1}$$

When the horizon k is long, we need many terms of the sum. These terms can be poorly estimated in "small" samples. By imposing structure on the null we can obtain simpler formulas that can perform more reliably in small samples.

The long-horizon forecast regression is

$$y_{t+k} = \sum_{j=1}^{k} \rho^{j-1} r_{t+j} = \alpha + \beta x_t + v_{t+k}$$

I impose the null that the returns are unforecastable. Then the regression is

$$y_{t+k} = x_t \beta + \varepsilon_{t+1} + \rho \varepsilon_{t+2} + ..\rho^{k-1} \varepsilon_{t+k}$$

and we can recover σ_ε^2 from the regression residual by

$$\sigma_v^2 = \sigma^2\left(\varepsilon_{t+1} + \rho \varepsilon_{t+2} + ..\rho^{k-1} \varepsilon_{t+k}\right) = \sigma_\varepsilon^2\left(1 + \rho^2 + .. + \rho^{2(k-1)}\right) = \frac{1-\rho^{2k}}{1-\rho^2}\sigma_\varepsilon^2$$

Plugging in to the standard error formula we have

$$\begin{aligned} var(\hat{\beta}) &= \frac{1}{T} E(x_t x_t')^{-1} \times \\ &\quad \sum_{j=-\infty}^{\infty} E\left(\left(\varepsilon_{t+1} + \rho \varepsilon_{t+2} + .. + \rho^{k-1} \varepsilon_{t+k}\right) x_t x_{t-j}' \left(\varepsilon_{t-j+1} + \rho \varepsilon_{t-j+2} + .. + \rho^{k-1} \varepsilon_{t-j+k}\right)\right) E(x_t x_t')^{-1} \end{aligned}$$

I assume that the ε_t are *iid*, and independent (as well as orthogonal to) past x. I do not assume that ε_t are independent of contemporaneous and future x – return innovations today do affect the dividend yield tomorrow. Thus

$$\begin{aligned} & E\left[\left(\varepsilon_{t+1} + \rho \varepsilon_{t+2} + .. + \rho^{k-1} \varepsilon_{t+k}\right) x_t x_t' \left(\varepsilon_{t+1} + \rho \varepsilon_{t+2} + .. + \rho^{k-1} \varepsilon_{t+k}\right)\right] \\ = \; & E\left[\left(\varepsilon_{t+1} + \rho \varepsilon_{t+2} + .. + \rho^{k-1} \varepsilon_{t+k}\right)^2\right] E\left[x_t x_t'\right] \\ = \; & \left(1 + \rho^2 + .. + \rho^{2(k-1)}\right) \sigma_\varepsilon^2 E\left[x_t x_t'\right] \\ = \; & \frac{1-\rho^{2k}}{1-\rho^2} \sigma_\varepsilon^2 E\left[x_t x_t'\right] \end{aligned}$$

The first lag term is

$$E\left[\left(\varepsilon_{t+1}+\rho\varepsilon_{t+2}+..+\rho^{k-1}\varepsilon_{t+k}\right)x_t x_{t-1}'\left(\varepsilon_t+\rho\varepsilon_{t+1}+..+\rho^{k-1}\varepsilon_{t-1+k}\right)\right]$$

Note by the independence assumption

$$E\left[\varepsilon_{t+1}x_t x_{t-1}'\varepsilon_t\right]=E\left(\varepsilon_{t+1}\right)E\left(x_t x_{t-1}'\varepsilon_t\right)$$

thus the fact that ε_t is not independent of x_t does not stop us from eliminating terms with different dates on the ε. Thus, the first lag term simplifies to

$$\begin{aligned}
& E\left[\left(\varepsilon_{t+1}+\rho\varepsilon_{t+2}+..+\rho^{k-1}\varepsilon_{t+k}\right)x_t x_{t-1}'\left(\varepsilon_t+\rho\varepsilon_{t+1}+..+\rho^{k-1}\varepsilon_{t-1+k}\right)\right] \\
= \; & E\left[\left(\varepsilon_{t+1}+\rho\varepsilon_{t+2}+..+\rho^{k-1}\varepsilon_{t+k}\right)\left(\rho\varepsilon_{t+1}+..+\rho^{k-1}\varepsilon_{t-1+k}\right)\right]E\left(x_t x_{t-1}'\right) \\
= \; & \left(\rho+\rho^3+..+\rho^{2(k-2)+1}\right)\sigma_\varepsilon^2 E\left(x_t x_{t-1}'\right) \\
= \; & \rho\frac{1-\rho^{2(k-1)}}{1-\rho^2}\sigma_\varepsilon^2 E\left(x_t x_{t-1}'\right)
\end{aligned}$$

Continuing,

$$\begin{aligned}
var(\hat{\beta}) &= \frac{1}{T}E(x_t x_t')^{-1}\left\{E(x_t x_t')\frac{1-\rho^{2k}}{1-\rho^2}\sigma_\varepsilon^2\right. \\
&\quad +\left[E\left(x_t x_{t-1}'\right)+E\left(x_{t-1}x_t'\right)\right]\rho\frac{1-\rho^{2(k-1)}}{1-\rho^2}\sigma_\varepsilon^2 \\
&\quad \left.+\left[E\left(x_{t-2}x_t'\right)+E\left(x_t x_{t-2}'\right)\right]\rho^2\frac{1-\rho^{2(k-3)}}{1-\rho^2}\sigma_\varepsilon^2+...\right\}E(x_t x_t')^{-1} \\
&= \frac{1}{T}E(x_t x_t')^{-1}\sigma_\varepsilon^2\sum_{j=-k}^{k}E\left(x_t x_{t-j}'\right)\frac{1-\rho^{2(k-j)}}{1-\rho^2}\rho^{|j|}E(x_t x_t')^{-1} \\
&= \frac{1}{T}E(x_t x_t')^{-1}\sigma_v^2\sum_{j=-k}^{k}\left(\frac{1-\rho^{2(k-j)}}{1-\rho^{2k}}\rho^{|j|}\right)E\left(x_t x_{t-j}'\right)E(x_t x_t')^{-1}
\end{aligned}$$

Unweighted long-horizon regressions are also often used,

$$y_{t+k}=\sum_{j=1}^{k}r_{t+j}=\alpha+\beta x_t+v_{t+k}$$

We can obtain the result in this case by taking the limit $\rho\to 1$,

$$\lim_{\rho\to 1}\left(\frac{1-\rho^{2(k-j)}}{1-\rho^{2k}}\rho^{|j|}\right)=\frac{\frac{d}{d\rho}\left(1-\rho^{2(k-j)}\right)|_{\rho=1}}{\frac{d}{d\rho}\left(1-\rho^{2k}\right)|_{\rho=1}}=\frac{-2\left(k-j\right)\frac{1}{\rho}\rho^{2(k-j)}|_{\rho=1}}{-2k\frac{1}{\rho}\rho^{2k}|_{\rho=1}}=\frac{k-j}{k}$$

Thus, the answer in this case is

$$var(\hat{\beta})=\frac{1}{T}E(x_t x_t')^{-1}\sigma_v^2\sum_{j=-k}^{k}\frac{k-|j|}{k}E\left(x_t x_{t-j}'\right)E(x_t x_t')^{-1}.$$