

Predicting Dividends in Log-Linear Present Value Models*

Andrew Ang
Columbia University and NBER

This Version: 8 August, 2011

JEL Classification: C12, C15, C32, G12
Keywords: predictability, dividend yield, cashflows,
dividend growth, time-varying risk premium

*I thank Yuhang Xing for excellent research assistance, and Geert Bekaert, Joe Chen, Jun Liu, Geert Rouwenhorst, and Tuomo Vuolteenaho for helpful discussions. I especially thank Michael Brandt, Bob Hodrick, and Zhenyu Wang for detailed comments. I am also grateful to the editor, Ghon Rhee, for very useful comments. Email address: aa610@columbia.edu

Abstract

In a present value model, high dividend yields imply that either future dividend growth must be low, or future discount rates must be high, or both. While previous studies have largely focused on the predictability of future returns from dividend yields, dividend yields also strongly predict future dividends, and the predictability of dividend growth is much stronger than the predictability of returns at a one-year horizon. Inference from annual regressions over the 1927-2000 sample imputes over 85% of the variation of log dividend yields to variations in dividend growth. Point estimates of the predictability of both dividend growth and discount rates is stronger when the 1990-2000 decade is omitted.

1 Introduction

In a present value model, the market price-dividend ratio is the present value of future expected dividend growth, discounted at the required rate of return of the market. If the dividend yield, the inverse of the price-dividend ratio, is high, then future expected dividend growth must be low, or future discount rates must be high, or both. While there is a very large body of research focusing on the predictability of future returns by the dividend yield, the forecasting power of dividend yields for future dividend growth has been largely ignored. In fact, Cochrane's (2011) presidential address to the American Finance Association overlooks totally the predictive ability of the dividend yield to forecast future cashflows and concentrates entirely on the dividend yield's ability to forecast future returns.¹ In this paper, I highlight the evidence of predictability of dividend growth by the dividend yield, and estimate the relative importance of future dividends for explaining the variation of the dividend yield.

I begin by standard simple regressions of long-horizon dividend growth and long-horizon total returns (which include both capital gain and dividend income). To characterize the predictability of dividend growth and expected returns, I work with the log-linear dividend yield model of Campbell and Shiller (1988b). Although this setup only approximates the true non-linear dividend yield process, this approach maps the one-period regression coefficients directly to the variance decompositions.² However, since long-horizon regression coefficients can be very different from one-period regression coefficients, I also run weighted long-horizon regressions following Cochrane (1992) to compute variance decompositions. Here, future dividend growth or returns are geometrically downweighted by a constant, which is determined from the log-linear approximation.

In my analysis, I am careful to use robust t-statistics and account for small sample biases (see Nelson and Kim, 1993). Using a log-linear Vector Autoregression (VAR) as a data generating process, I show that Newey-West (1987) and robust Hansen-Hodrick (1980) t-statistics have large size distortions (see also Hodrick, 1992; and Ang and Bekaert, 2007). On the other hand, Hodrick (1992) t-statistics are well-behaved and have negligible size distortions. Simulating under the alternative hypothesis of dividend growth or return predictability by log dividend

¹ For a partial list of the literature using dividend yields to predict returns, see Fama and French (1988), Campbell and Shiller (1988a), Cochrane (1992, 2011), Lettau and Ludvigson (2001), Lewellen (2004), Campbell and Yogo (2006), Ang and Bekaert (2007), Goyal and Welch (2008), Campbell and Thompson (2008), Lettau and Van Nieuwerburgh (2008), Chen (2009), Chen and Zhao (2009), and van Binsbergen and Koijen (2010).

² Campbell (1991), Campbell and Ammer (1993), Ammer and Mei (1996), Vuolteenaho (2002), and Chen and Zhao (2009), among others, use the Campbell and Shiller (1988b) log-linear model.

yields, I find that Hodrick (1992) t-statistics are also the most powerful among these three t-statistics. Whereas using Wald tests to determine the significance of variance decompositions produces severe small sample distortions, testing the variance decompositions from regression coefficients has much better small sample behavior. Further, if log dividend yields are used as predictive instruments rather than dividend yields in levels, the Stambaugh (1999) bias resulting from a correlated regressor variable is negligible.

The first striking result is that using data from 1927-2000 on the CRSP value-weighted market index, dividend growth is strongly predictable by log dividend yields. A 1% increase in the log dividend yield, lowers next year's forecast of future dividend growth by 0.13%. Dividend growth predictability is much stronger at short horizons (one-year) than at long horizons. In contrast, returns are not forecastable by log dividend yields at any horizon, unless the returns during the 1990s are excluded.

Second, if the 1990s are omitted, the evidence of both dividend growth predictability and return predictability becomes stronger.³ From 1927-1990, the magnitude and significance of the predictability coefficient of dividend growth still dominates, by a factor of two, the predictability coefficient of returns at an annual horizon. Without the 1990s, dividend growth predictability is significant at longer horizons (up to four years) with data at a monthly frequency.

Third, using one-period regressions (restricted VARs) to infer the variance decomposition of dividend yields assigns over 85% of the variance of the log dividend yield to dividend growth over the full sample. This is because, at one-year horizons, the magnitude of the predictability coefficient of dividend growth is much larger than the predictability coefficient of returns. While it is hard to make any statistically significant statements about the variance decompositions using the asymptotic critical values from Wald tests, I can attribute a major portion of the variance of the log dividend yield to dividend growth, and this attribution is highly significant once I account for the size distortions of the small sample distributions.

Finally, inference from weighted long-horizon regressions to compute the variance decomposition is treacherous because of the serious size distortions induced by the use of overlapping data. Use of Newey-West (1987) or robust Hansen-Hodrick (1980) standard errors leads to incorrect inference that attributes most of the variation in log dividend yields to expected returns. With robust t-statistics, no statistically significant statement can be made about the variance

³ Both Goyal and Welch (2001) and Ang and Bekaert (2007) document that when the 1990s are included in the sample period, dividend yields do not predict excess returns at any horizon. Authors who employ standard errors implied from nearly-integrated variables usually find weak or no evidence of predictability by dividend yields. See, for example, Richardson and Stock (1993) and Valkanov (2003).

decompositions. However, the point estimates show that the predictability of expected returns, although small at short horizons, increases at long horizons, as found by Shiller (1981) and others. In contrast, while dividend yields strongly predict dividend growth at short horizons, the point estimates of the long-horizon predictability of dividend growth are insignificant and smaller.

Why has the predictability of dividend growth been over-looked in the literature relative to the predictability of returns?⁴ Previous studies concentrate on the predictive regressions with expected total or excess returns and do not consider the predictability of dividend growth. For example, while Fama and French (1988) and Hodrick (1992) consider putting long-horizon expected (excess) returns on the LHS of a regression, they do not forecast long-horizon dividend growth with dividend yields. In Campbell and Shiller's (1989) VAR tests of the dividend discount model, dividend growth does not have its own separate forecasting equation by log dividend yields. In Campbell and Ammer (1993), no cashflows appear directly in the VARs even though past cashflows are observed variables. Instead, Campbell and Ammer specify the process for returns and only indirectly infer news about dividend growth from the VAR as a remainder term. In contrast to these studies, I explicitly run regressions with dividend growth on the LHS, and include dividend growth as a separate variable with its own law of motion in the overall data-generating process. Chen and Zhao (2009) also show that not including direct measures of cashflows and discount rates leads to incorrect inference about dividend growth predictability; all the VAR data-generating processes I consider include both returns and dividend growth.

The rest of the article proceeds as follows. Section 2 describes the construction of dividend yields, growth rates and returns from the CRSP market index. Section 3 motivates the empirical work using Campbell and Shiller's (1988b) log-linear relation. Section 4 outlines the regression framework and compares the size and power of various robust t-statistics. I decompose the variance of the log-dividend yield in Section 5, imputed by one-period regressions and Cochrane (1992) long-horizon weighted-regressions. Section 6 concludes.

⁴ Since the first draft of this paper in 2002, there has been a growing literature that finds that cashflow risk plays an important role in explaining the variation of returns, including Bansal and Yaron (2004), Bansal, Dittmar and Lundblad (2005), Lettau and Wachter (2005), Hansen, Heaton and Li (2008), Chen (2009), Chen and Zhao (2009), and van Binsbergen and Koijen (2010).

2 Data

All the data are from the CRSP value-weighted portfolio from Jan 1927 to Dec 2000, both at a monthly and at an annual frequency. All the time subscripts t are in years, so t to $t+1$ represents one year, and t to $t + 1/12$ represents one month. To compute monthly dividend yields, I use the difference between CRSP value-weighted returns with dividends $VWRET D$ and CRSP value-weighted returns excluding dividends $VWRET X$. The monthly income return from t to $t + 1/12$ is computed from:

$$\frac{\bar{D}_{t+1/12}}{P_t} = VWRET D_{t+1/12} - VWRET X_{t+1}, \quad (1)$$

where I denote the monthly dividend in month $t + 1/12$ as $\bar{D}_{t+1/12}$. The bar superscript in $\bar{D}_{t+1/12}$ indicates that this is a monthly, as opposed to annual, dividend. Dividends are summed over the past twelve months, as is standard practice, to remove seasonality in the dividend series and to form an annual dividend, D_t :⁵

$$D_t = \sum_{i=0}^{11} \bar{D}_{t-i/12}.$$

The log dividend yield is given by:

$$dy_t = \log \left(\frac{D_t}{P_t} \right). \quad (2)$$

To compute continuously compounded dividend growth rates, g , I use:

$$g_t = \log \left(\frac{D_t}{D_{t-1}} \right),$$

which gives a time-series of annual log dividend growth. This series is available at a monthly frequency but refers to dividend growth over an annual horizon.

I express monthly equity returns $\bar{r}_{t+1/12}$ as continuously compounded returns:

$$\bar{r}_{t+1/12} = \log \left(\frac{P_{t+1/12} + \bar{D}_{t+1/12}}{P_t} \right),$$

I work with annual horizons, so the annual equity return and the annual excess equity return are obtained by summing up equity returns over the past 12 months:

$$r_t = \sum_{i=0}^{11} \bar{r}_{t-i/12}. \quad (3)$$

⁵ Chen (2009) shows that re-investing dividends in the market portfolio tends to understate the predictability of dividend growth because it contaminates dividend growth with stock returns.

Note that although these are annual horizons, total equity returns are available at a monthly frequency.

In the empirical analysis, I use both monthly and annual frequencies, but focus most of my work at the annual horizon. Monthly data for annual returns and dividend growth has the problem of each observation sharing data over 11 overlapping months, so the moving average errors induced by the monthly frequency are much larger than for an annual frequency. However, in all cases, using monthly data has almost the same results as using annual data. Table 1 lists summary statistics of the market dividend yields, dividend growth, and total equity returns (including capital gains and dividend income). I report annual frequencies; the summary statistics for monthly frequencies are similar. The data are split into two subsamples, from January 1927 to December 1990, and the full sample January 1927 to December 2000. The 1990s bull market saw very high returns with decreasing dividend yields, so I am careful to run the predictability regressions with and without the 1990s. Most of the summary statistics of Table 1 are well known. Total equity returns have almost zero autocorrelation and log dividend yields are highly autocorrelated (0.76 over the full sample). While dividend growth is weakly autocorrelated (0.30), this is not significant at the 5% level.

3 Motivating Framework

The market price-dividend ratio P_t/D_t is the present value of future expected dividends, discounted back by the market's total expected return:

$$\frac{P_t}{D_t} = E_t \left[\sum_{i=1}^{\infty} \exp \left(\sum_{j=1}^i (-r_{t+j} + g_{t+j}) \right) \right]. \quad (4)$$

Assuming there are no bubbles, high price-dividend ratios indicate that either expected future cashflow growth must be high, or expected future discount rates must be low, or both. Equation (4) is a highly non-linear specification, and while closed-form expressions of (4) are available in affine economies, I follow Campbell and Shiller (1988b) and linearize the price-dividend expression in (4) to obtain an approximate linear expression. This allows time-series tools to be directly applied, but the linear identities do not fully capture, by construction, the full dynamics of the price-dividend ratio.⁶

⁶ For non-linear present-value models see, among others, Ang and Liu (2001, 2007), Mamaysky (2002), Bakshi and Chen (2005), Ang and Bekaert (2007), and Bekaert, Engstrom and Grenadier (2010).

Campbell and Shiller (1988b) derive an approximate one-period identity for the total return:

$$\exp(r_{t+1}) = \frac{P_{t+1} + D_{t+1}}{P_t}.$$

Letting lower case letters denote logs of upper case letters and re-arranging, Campbell and Shiller derive:

$$p_t - d_t \approx k - r_{t+1} + g_{t+1} + \rho(p_{t+1} - d_{t+1}), \quad (5)$$

where $p_t - d_t$ is the log price-dividend ratio, $g_{t+1} = \Delta d_{t+1}$ is one-period dividend growth, $\rho = 1/(1 + \exp(\overline{p - d}))$, where $\overline{p - d}$ denotes the average log price-dividend ratio, and k is a linearization constant given by $k = -\log(\rho) - (1 - \rho)\log(1/\rho - 1)$.

Iterating this approximation forward, it is easy to derive a log-linear equivalent specification to equation (4):

$$p_t - d_t = \frac{k}{1 - \rho} + E_t \left[\sum_{j=1}^{\infty} \rho^{j-1} (-r_{t+j} + g_{t+j}) \right]. \quad (6)$$

Multiplying each side by -1 gives an approximate log-linear identity for the dividend yield:

$$dy_t = d_t - p_t = -\frac{k}{1 - \rho} + E_t \left[\sum_{j=1}^{\infty} \rho^{j-1} (-g_{t+j} + r_{t+j}) \right], \quad (7)$$

where dy_t is the log dividend yield. According to equation (7), a high dy_t today implies that either future dividend growth rates are low, or future discount rates are high, or both. Hence, if we regress future growth rates onto dy_t we would expect to see negative coefficients, or if we regress future returns onto dy_t we would expect to see positive coefficients. I examine these predictive regressions directly.

Equation (7) further allows the variance of the log dividend yield to be decomposed as:

$$\text{var}(dy_t) = -\text{cov} \left(dy_t, E_t \left[\sum_{j=1}^{\infty} \rho^{j-1} g_{t+j} \right] \right) + \text{cov} \left(dy_t, E_t \left[\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} \right] \right). \quad (8)$$

In a simple VAR, the variance decomposition (8) can be easily evaluated. In particular, letting $X_t = (dy_t \ g_t \ r_t)'$ follow a VAR:

$$X_t = \mu + AX_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, \Sigma)$, the variance of the log dividend yield due to cashflows is given by:

$$-\text{cov} \left(dy_t, E_t \left[\sum_{j=1}^{\infty} \rho^{j-1} g_{t+j} \right] \right) = -e'_2 A (I - \rho A)^{-1} \Sigma_X e_1, \quad (9)$$

where e_i is a vector of zero's with a 1 in the i th position, and Σ_X is the unconditional covariance matrix of X_t :

$$\Sigma_X = \text{devec}[(I - A \otimes A)^{-1} \text{vec}(\Sigma)].$$

The variance of the log dividend yield due to total expected returns can be computed by:

$$\text{cov} \left(dy_t, E_t \left[\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} \right] \right) = e_3' A (I - \rho A)^{-1} \Sigma_X e_1. \quad (10)$$

The expressions (9) and (10) can be easily tested if they are equal to zero by conducting Wald tests. Given that there are well-known problems with Wald tests (see, among others, Burnside and Eichenbaum, 1996; Hansen, Heaton and Yaron, 1996; and Bekaert and Hodrick, 2001), especially substantial size distortions in small samples, I conduct statistical inference using simulated small sample distributions for the Wald statistics.

While the VAR allows easy computation of the variance decompositions, a straight-forward application of the VAR yields many parameters likely to be insignificant in the companion form A of the VAR. Including these insignificant parameters in the VAR may make the variance decompositions unreliable by increasing the standard error of the variance decompositions since these parameters are estimated very imprecisely. The expressions (9) and (10) also involve an inverse, and it is not obvious how simple predictive regression coefficients, from regressions of future dividend growth or returns onto log dividend yields, are related to the variance decompositions of these variables.

To map the one-period predictive regression coefficients of dividend growth and returns into variance decompositions of the log dividend yield attributable to these variables, I work with a specialized VAR:

Proposition 3.1 *Suppose the log dividend yield follows dy_t , dividend growth g_t and total returns r_t follow the following process:*

$$\begin{aligned} dy_{t+1} &= \alpha_{dy} + \rho_1 dy_t + \rho_2 g_t + \rho_3 r_t + \epsilon_{1t} \\ g_{t+1} &= \alpha_g + \beta_g dy_t + \epsilon_{2t} \\ r_{t+1} &= \alpha_r + \beta_r dy_t + \epsilon_{3t}, \end{aligned} \quad (11)$$

where $\epsilon_t = (\epsilon_{1t} \epsilon_{2t} \epsilon_{3t})' \sim N(0, \Sigma)$. Then the proportion of the variance of log dividend yields attributable to variation in dividend growth is given by:

$$-\frac{1}{\text{var}(dy_t)} \text{cov} \left(dy_t, \sum_{j=1}^{\infty} \rho^{j-1} g_{t+j} \right) = \frac{-\beta_g}{-\beta_g + \beta_r}, \quad (12)$$

where $\text{var}(dy_t) = -\text{cov}\left(dy_t, \sum_{j=1}^{\infty} \rho^{j-1} g_{t+j}\right) + \text{cov}\left(dy_t, \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}\right)$. The proportion of the variance of log dividend yields attributable to variation in total returns is given by:

$$\frac{1}{\text{var}(dy_t)} \text{cov}\left(dy_t, \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}\right) = \frac{\beta_r}{-\beta_g + \beta_r}. \quad (13)$$

There are several appealing features about Proposition 3.1. First, the setup is very simple and clean, since the one-period regression coefficients β_g and β_r are directly related to the variance decompositions. The larger the one-period predictability coefficients $-\beta_g$ and β_r of future dividend growth and total returns, respectively, the larger the proportion of the variance of the log dividend yield we can attribute directly to dividend growth or total returns. To estimate a model under the null where the variance decomposition to dividends (returns) is zero, but dividend yields may predict returns (dividends), then I can impose the restriction that $\beta_g = 0$ ($\beta_r = 0$). The natural alternative models are then $\beta_g \neq 0$ and $\beta_r \neq 0$, respectively.

Second, in Proposition 3.1 only the predictability coefficients matter for the percentage variance decomposition, and the covariances Σ and the process of the dividend yield (ρ_1 , ρ_2 and ρ_3) do not enter (12) or (13). The larger the magnitude of β_g , relative to β_r , the greater the attribution of the variance of dividend yields to dividend growth. The reason why the proportion of the variance decomposition only depends on the regression coefficients is that the other coefficients (ρ_i and Σ) affect the level of the covariance terms in the same manner, and cancel out when computing the proportions. Thus, intuitively the greater the magnitude of the predictive coefficients, the larger the dividend yield variance attribution to dividend growth or returns.

Finally, I test (11) against an unrestricted VAR alternative. The four coefficients appearing under the alternative (the autocorrelation of dividend growth, the autocorrelation of returns, the coefficient predicting dividend growth with lagged returns and the coefficient predicting returns with lagged dividend growth) are all insignificant in univariate regressions. I also cannot reject the null that each of the individual coefficients are equal to zero in a restricted VAR. Finally, a joint test that these coefficients equal zero also fails to reject.

One problem with the formulation in Proposition 3.1 is that it only uses one-period regressions to impute the unconditional variance decomposition, which is a critique that can be leveled at any VAR-based technique.⁷ Since one-period regression estimates are often different from long-horizon regression estimates, a direct way to compute the variance decomposition is

⁷ However, the usual criterion for lag selection length, such as the Schwartz or Aikake criteria picks one lag as the optimal choice.

to sum (7) only up to horizon k and leave a terminal value:

$$dy_t = \text{constant} + E_t \left[\sum_{j=1}^k \rho^{j-1} (-g_{t+j} + r_{t+j}) \right] + dy_{t+k}. \quad (14)$$

Then, taking variances of both sides yields:

$$\text{var}(dy_t) = -\text{cov} \left(dy_t, \sum_{j=1}^k \rho^{j-1} g_{t+j} \right) + \text{cov} \left(dy_t, \sum_{j=1}^k \rho^{j-1} r_{t+j} \right) + \text{cov}(dy_t, dy_{t+k}). \quad (15)$$

The last term approaches zero as k becomes large.

As Cochrane (1992) notes, the coefficient

$$\frac{-\text{cov} \left(dy_t, \sum_{j=1}^k \rho^{j-1} g_{t+j} \right)}{\text{var}(dy_t)} \quad (16)$$

is the proportion of the dividend yield variance attributable to the variation in dividend yields. This can be estimated by regressing $-\sum_{j=1}^k \rho^{j-1} g_{t+j}$ on dy_t . Similarly, the percentage variance decomposition for r is given by:

$$\frac{\text{cov} \left(dy_t, \sum_{j=1}^k \rho^{j-1} r_{t+j} \right)}{\text{var}(dy_t)}, \quad (17)$$

which can be estimated by regressing $\sum_{j=1}^k \rho^{j-1} r_{t+j}$ on dy_t . The difference between these regressions and standard long-horizon regressions, is that each horizon is down-weighted geometrically by a factor of ρ . As $k \rightarrow \infty$, it can be shown that if X_t follows a VAR, then (16) converges to (9).

Computing the variance decomposition using long-horizon weighted regressions such as (16) necessitates the use of overlapping data. This induces severe size distortions, as noted by several studies including Hodrick (1992), Richardson (1993), Campbell (2001), and Ang and Bekaert (2007). Hence, before estimating the variance decompositions, I run several Monte Carlo experiments to determine the best choice of t-statistics to conduct statistical inference.

4 Predictive Regressions

4.1 Methodology

The direct implication of (7) is that high log dividend yields today predict future low dividend growth, or future high returns, or both. To test this directly, I regress long-horizon dividend

growth or long-horizon total returns onto the current log dividend yield. Following Fama and French (1988), the main predictive regressions I consider are:

$$z_{t+k} = \alpha + \beta_k dy_t + \epsilon_{t+k,k} \quad (18)$$

where dy_t is the log dividend yield, and z_{t+k} is the k -period ahead variable which is being forecast. I predict dividend growth rates k years ahead, in which case

$$z_{t+k} = (1/k)(g_{t+1} + \dots + g_{t+k}),$$

and for the case of predicting total returns k years ahead:

$$z_{t+k} = (1/k)(r_{t+1} + \dots + r_{t+k}).$$

Although the regression (18) can be consistently estimated by OLS, it is subject to some serious statistical problems which affect inference about the coefficients β_k . First, the regression (18) is run both at annual and monthly frequencies. In the case of $k > 1$, both frequencies use overlapping data, which induce moving average errors. Second, the LHS variables are heteroskedastic. Third, the dividend yield on the RHS is an endogenous regressor, and this causes the coefficients to be biased (see Stambaugh, 1999). Without correcting for these effects, one may erroneously conclude that the β_k coefficients are significantly different from zero.

Using GMM, the parameters $\theta = (\alpha \beta_k)'$ in equation (18) have an asymptotic distribution $\sqrt{T}(\hat{\theta} - \theta) \overset{d}{\sim} N(0, \Omega)$ where $\Omega = Z_0^{-1} S_0 Z_0^{-1}$, $Z_0 = E(x_t x_t')$, $x_t = (1 dy_t)'$. I compare three estimates of S_0 . The first estimate is Newey-West (1987) with k lags:

$$\hat{S}_0 = k \sum_{j=-k}^k \frac{k - |j|}{k} C(j), \quad (19)$$

where

$$C(j) = \frac{1}{T} \sum_{t=j+1}^T (w_{t+k} w'_{t+k-j}).$$

and $w_{t+k} = \epsilon_{t+k,k} x_t$.

While the Newey-West estimate guarantees invertibility, it down-weights higher order autocorrelations. Hodrick (1992) develops a heteroskedastic extension of Hansen and Hodrick (1980), termed the 1A estimate, which avoids this problem:

$$\hat{S}_0 = C(0) + \sum_{j=1}^{k-1} [C(j) + C(j)']. \quad (20)$$

Unfortunately, this estimator of S_0 is not guaranteed to be positive semi-definite and so its behavior may be erratic in small samples.

The final estimate of S_0 I consider is the Hodrick (1992) 1B estimate, which exploits covariance stationarity to remove the overlapping nature of the error terms in the standard error computation. Instead of summing $\epsilon_{t+k,k}$ into the future to obtain an estimate of S_0 , $x_t x'_{t-j}$ is summed into the past:

$$\hat{S}_0 = \frac{1}{T} \sum_{t=k}^T \bar{w}_t \bar{w}'_t \quad (21)$$

where

$$\bar{w}_t = \epsilon_{t+1,1} \left(\sum_{i=0}^{k-1} x_{t-i} \right).$$

Note that for a horizon $k = 1$ with annual data, the Hodrick 1A and 1B estimates are exactly the same. Richardson and Smith (1991), Hodrick (1992), and Ang and Bekaert (2007) show that the small sample properties of t-statistics improve dramatically by summing the data to remove the moving average structure in the error terms, rather than summing the error terms themselves.

4.2 Empirical Results

I start by examining univariate regressions, where I regress future dividend growth or total returns on the market portfolio onto lagged log dividend yields. In Table 2 I denote the Newey-West t-statistics as t-NW, the Hodrick (1992) 1A t-statistics as t-1A, and the Hodrick (1992) 1B t-statistics as t-1B. Panel A runs the regressions at an annual frequency while Panel B reports the results of the regressions at a monthly frequency. For the annual frequency, the one year horizon corresponds from January to December over a particular calendar year. For the monthly frequency, a one-year horizon also takes the other annual observations within the calendar year, for example, February to February. In both cases, I report results over a sample excluding the 1990s (1927-1990) and over the full sample (1927-2000).

Table 2 contains a number of remarkable stylized facts. First, in Table 2, the point estimates of the coefficients are approximately the same using either an annual (Panel A) or a monthly frequency (Panel B). However, using monthly frequency data does not imply that the t-statistics are necessarily larger because of greater power owing to more observations. For example, when I regress returns onto log dividend yields at an annual frequency over the five-year horizon, the Newey-West t-statistic using annual (monthly) data is 1.90 (1.85). Nevertheless, the same patterns are evident, particularly in the point statistics, in using both monthly and annual data.

Second, the t-statistics differ considerably depending on which covariance estimate is used. For example, focusing on the log dividend yield predicting future dividend growth over the full sample at an annual frequency (Panel A), one would conclude that log dividend yields significantly predict future dividend growth using t-1A or t-1B statistics at the one-year horizon, but one would find no evidence of predictability of future dividends using t-NW statistics. One would also conclude that long-horizon total returns are weakly predictable by log dividend yields using t-NW statistics (t-NW=1.90 at the five-year forecasting horizon), but find no evidence of any predictive power for returns by dividend yields using t-1B statistics (t-1B=1.27). I carefully address the small sample properties of each covariance estimator in the next section to ensure that I use the best-behaved t-statistic for inference.

Third, there is a wide discrepancy between both the point estimates of the coefficients and their t-statistics with and without the 1990s. Looking first at the different point estimates from 1927-1990 and from 1927-2000 in the annual frequency regressions, the coefficient on log dividend yields predicting dividend growth at the one-year horizon over the full sample is about half the size of the coefficient when the 1990s are omitted (-0.129 and -0.232 respectively). For longer horizons, the *dy* coefficients are only slightly smaller in magnitude. The reduction in the magnitude of the coefficients from 1927-1990 to the full sample is much more pronounced for predicting total returns. The one-year horizon coefficient over 1927-2000 for dividend yields forecasting total returns is only 0.018, versus 0.125 over the sample excluding the 1990s. At a five-year horizon, the *dy* coefficients forecasting discount over the full sample is 0.080 compared to 0.135 over the sample omitting the 1990s.

Turning now to examine the magnitude of the t-statistics with and without the 1990s, we see that over the full sample there is evidence that dividend growth is predictable using t-1B statistics, at short horizons (one year). However, there is no evidence across all t-statistics that discount rates are predictable by dividend yields at any horizon. This result is robust to using data at both monthly and annual frequencies. Omitting the 1990s increases the predictability of both dividend growth and returns by dividend yields. At the one-year horizon, the t-statistics on dividend growth predictability are much larger in magnitude than their counterparts for discount rate predictability.

Table 2, however, shows that using the sample ending in 1990, the long horizon predictability of returns is stronger than the long horizon predictability of dividend growth. The standard interpretation in the literature is that expected returns may reflect more slowly moving permanent components (see Shiller, 1981; and Shiller and Beltratti, 1992, among

many others). With the 1990s, there is no evidence of any long horizon predictability of total returns. Nevertheless, the point estimates of dividend growth predictability decrease in magnitude with horizon over both the 1927-1990 and 1927-2000 samples, while the point estimates of discount rate predictability generally increase with horizon. This may mean that although future dividends are predictable, at least in the short term, in the long-run, dividends may account for little of the total variation in the dividend yield. Using robust t-statistics, I directly address attributing the variance of the log dividend yield into dividend growth and expected total return sources in Section 5.

4.3 Size and Power of T-Statistics

The wide discrepancy between the t-statistics in Table 2 means that I must determine which t-statistics have the best small sample properties in order to perform correct inference about the predictability of dividend growth or returns by the dividend yield. To preview the results, I show that the Hodrick (1992) 1B standard errors have the best size properties and are the most powerful among the Newey-West and Hodrick t-statistics. To abstract from monthly seasonality issues and since the true de-seasonalized monthly dividend growth is unobservable, I focus on the annual frequency. Since the behavior of the point estimates and t-statistics in Table 2 is similar across the monthly and annual frequencies, I expect that the main results will carry over to the monthly frequency. I focus on a system with total returns and demonstrate the size and power properties using the data-generating process (11) of Proposition 3.1.

The estimates of (11) over the full sample are listed in Table 3. The equations where dy predicts g and r are the same as the one-year horizon estimates at an annual frequency over the full sample in Table 2. All the coefficients ρ_i affecting the log dividend yield are highly significant, and shocks to the log dividend yield are highly negatively correlated with shocks to total returns (-0.88). To conduct the Monte Carlo analysis of the various t-statistics, I use these estimates to simulate small samples of 73 years, exactly the same length as the data. The size results for the behavior of the (1992) Hodrick t-statistics for testing predictability of discount rates are similar to those found by Ang and Bekaert (2007).

I check the performance of the t-statistics for inference regarding both dividend growth and total return predictability by log dividend yields. To examine predictability of dividend growth by dividend yields, I simulate data from a constrained estimation of (11), setting $\beta_g = 0$, so that dividend growth is IID. In each simulated sample, I run the regression (18) for cumulated dividend growth on the LHS onto log dividend yields and record the number of rejections

corresponding to nominal size levels of 5% and 10% for Newey-West (1987) and Hodrick (1992) 1A and 1B t-statistics. To check power of the t-statistics I simulate under the alternative where $\beta_g = -0.129$. I run a similar exercise to check the small sample behavior of the t-statistics for predictability of total returns under the null of $\beta_r = 0$ versus the alternative of predictability of returns ($\beta_r = 0.018$).

Table 4 lists empirical size and power of Newey-West and Hodrick t-statistics. Turn first to the predictability of dividend growth in Panel A. At the one-year horizon, there are very few size distortions for all the t-statistics. At a nominal size of 5% (10%) at the one-year horizon, the empirical sizes using t-NW and t-1B statistics are 6.7% (5.7%) and 10.2% (10.2%), respectively. However, at long horizons the size distortions for the t-NW and t-1A t-statistics are large, especially compared to the behavior of the t-1B statistic. For example, at the five-year horizon corresponding to a nominal size level of 5%, the t-NW statistic rejects 19.2% of the time and the t-1A statistic rejects 15.9% of the time. In contrast, the t-1B statistics have negligible size distortions and reject only 4.4% of the time when the nominal size is 5%. Hence, the Hodrick 1-B t-statistics are slightly conservative for dividend growth predictability at long horizons. The size-adjusted power of all the t-statistics is high, around 95% at the one-year horizon, but it decreases to the 70-80% level at the five-year horizon. While at all horizons the t-1B statistics are more powerful than the t-NW or t-1A statistics, the superior performance of the Hodrick 1B t-statistic is most clear at long horizons. At a five-year horizon, the size-adjusted power of the t-1B statistic is 83.9% compared to 71.1% for the t-1A and 70.2% for the t-NW.

In Panel B of Table 4, I list empirical size and size-adjusted power of the total return regressions. At a 5% nominal size, all three t-statistics are more likely to reject the null of no discount rate predictability than the null of no dividend growth predictability. At the one-year horizon, the empirical sizes corresponding to a 5% nominal level are 8.0% for t-NW and 7.6% for t-1A and t-1B, respectively. At the five-year horizon, the size distortions are very large for t-NW and t-1A (21.4% and 18.6% empirical size, respectively), while using Hodrick 1B standard errors produces small size distortions (6.5%). Turning to size-adjusted power, I see that the t-1B statistic is again most powerful. However, the power for all three t-statistics is low, around 30% at the one-year horizon, decreasing to below 25% at the five-year horizon. The reason is that over the 1927-2000 sample, the coefficient β_r for dividend yields predicting discount rates is very small, only 0.018, as opposed to 0.125 over the 1927-1990 sample.

In Figure 1, I graph the size-adjusted power function of the t-statistics. In the top two graphs, I show the power at the one-year horizon as a function of the coefficient β_g under the

alternative and β_r under the alternative, respectively. The power corresponding to the $\beta_g = -0.129$ and $\beta_r = 0.018$ in the data is shown as an asterisk for the Hodrick 1-B t-statistics. Over the sample omitting the 1990s, the coefficients of β_g and β_r are -0.232 and 0.123, respectively. At these coefficients, the power of all the t-statistics is one. In the bottom two graphs, I show the size-adjusted power at the five-year horizon. Here, the superior power of the t-1B statistics is pictorially depicted by the solid line lying above the power curves for the t-NW and t-1A statistics. Power is lower at the five-year horizon than for the one-year horizon, particularly for the dividend growth regressions.

Table 4 and Figure 1 show that power of all the t-statistics decreases with horizon. Mark and Sul (2002) show that long horizon tests asymptotically may have greater power over some regions of the parameter space using local-to-zero analysis. For my Monte Carlo experiments, this is not the case. Asymptotic local-to-zero analysis uses a family of alternative hypotheses, indexed by (unobservable) nuisance parameters, while all the alternative hypotheses are fixed. Campbell (2001) also finds that there are many alternatives where there are no small sample power advantages for long horizon regressions, as do Berkowitz and Giorgianni (2001).

4.4 Bias Under the Null of No Predictability

The bottom of Panels A and B of Table 4 report the bias of the predictive regressions. At the bottom of Panel A, I report the bias of the predictive dividend growth regression under the null of $\beta_g = 0$, which is extremely small (approximately less than 0.001 in absolute magnitude). Similarly, the bottom part of Panel B reports that the bias from the estimated coefficients on the log dividend yield, predicting future total returns under the null that $\beta_r = 0$, is also very small, less than 0.007 at each horizon. This bias is far smaller than the biases that Stambaugh (1999) and Lewellen (2004) find for forecasting returns using dividend yields in levels. Using monthly NYSE data from 1927-1996, Stambaugh finds that at a one-period horizon, the bias in β_r is 0.07, so the bias is one order of magnitude less than Stambaugh's estimate. While there are several differences between the system (11) and Stambaugh's setup (for example (11) is a multivariate system and Stambaugh includes only a univariate endogenous regressor), the main reason why my system has such smaller bias is that I use dividend yields in logs, while Stambaugh uses dividend yields in levels.

Stambaugh (1999) considers the system:

$$\begin{aligned} y_t &= \alpha + \beta_x x_{t-1} + u_t \\ x_t &= \theta + \phi x_{t-1} + v_t \end{aligned} \quad (22)$$

where y_t are monthly excess NYSE returns, x_t is the monthly NYSE dividend yield in levels and:

$$\text{cov} \left(\begin{bmatrix} u_t \\ v_t \end{bmatrix}, [u_t \ v_t] \right) = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}.$$

Stambaugh finds that an approximation for the bias in $\hat{\beta}_x$ under the assumption of normality is:

$$E(\hat{\beta}_x - \beta_x) \approx -\frac{\sigma_{uv}}{\sigma_v^2} \left(\frac{1 + 3\phi}{T} \right). \quad (23)$$

The term in brackets on the RHS is the upward bias of the autocorrelation $\hat{\phi}$ in small samples.

Stambaugh computes that the bias $E(\hat{\beta}_x - \beta_x) = 0.07$ using $T = 840$ corresponding to monthly NYSE data from 1927-1996, where $\phi = 0.972$, $\sigma_u^2 \times 10^4 = 30.05$, $\sigma_v^2 \times 10^4 = 0.108$ and $\sigma_{uv} \times 10^4 = -1.621$. The ratio of $-\sigma_{uv}/\sigma_v^2$ is 15, which magnifies the small autocorrelation bias. If log dividend yields are used instead of dividend yields in levels, then $\phi = 0.983$ and the autocorrelation bias is almost unchanged. However, now $\sigma_u^2 \times 10^4 = 30.10$, $\sigma_v^2 \times 10^4 = 32.33$, $\sigma_{uv} \times 10^4 = -29.74$ and the ratio $-\sigma_{uv}/\sigma_v^2 = 0.92$, making the bias negligible (0.004). This is approximately the same size of the bias in the system (11) reported in Table 4. The difference between dividend yields in levels and dividend yields in logs is not due to a lower ϕ (in fact, ϕ is higher with log dividend yields), nor due to a lower correlation between u and v (in fact, $\text{corr}(u, v) = -0.90$ with level dividend yields and $\text{corr}(u, v) = -0.95$ with log dividend yields).

The reason that the Stambaugh bias is not a concern when log dividend yields are used is that σ_v , the standard deviation of shocks to the regressor variable, is the same order of magnitude as σ_u , the standard deviation of shocks to returns. This is not the case when dividend yields in levels are used, where σ_v is one order of magnitude smaller than σ_u . This does not mean that the log transformation ‘‘increases’’ the variance of shocks to the regressor variable, it merely transforms them to be the same scale. Rather, σ_v for logs divided by the average log dividend yield is actually much smaller than σ_v for levels divided by the average level of the dividend yield. Intuitively, the log transformation reduces the magnitude of the skewness and kurtosis making the log dividend yield more normally distributed than the level dividend yield. Furthermore, since the annual autocorrelation of the log dividend yield is fairly low over annual horizons (0.761 in the full sample, listed in Table 1), accounting for the bias

in the autocorrelation of the regressor variable as Lewellen (2004) would result in negligible distortions to the predictability coefficient. The bottom line is that if log dividend yields are used as instruments at annual horizons, the bias in the predictive regressions is not a concern.

4.5 Small Sample Distributions of Wald Tests

I now turn to examining the small sample distribution of the Wald test statistics testing $\text{cov}(dy, \sum \rho^{j-1} g_{t+j}) = 0$ (equation (9)) and $\text{cov}(dy, \sum \rho^{j-1} r_{t+j}) = 0$ (equation (10)). To test the former (latter), I simulate data from (11) constraining $\beta_g = 0$ ($\beta_r = 0$). In each simulated sample, I compute the test statistics by re-estimating the parameters in (11), and use their moment conditions to form the covariance matrix in the calculation of the Wald test statistics. To compute the Wald statistic, I use the Hodrick (1992) 1B covariance estimator. I use a set of parameters for the data-generating process estimated over the period 1927-1990, and the set of parameters in Table 3, which are estimated using the whole sample.

Table 5 reports the small sample 5% and 10% cut-off levels, along with the nominal critical values. It is striking that the small sample critical values are so dissimilar to the asymptotic critical values. In particular, using an asymptotic critical value of 3.841, corresponding to a 5% level, an econometrician would fail to reject that $\text{cov}(dy, \sum \rho^{j-1} g_{t+j}) = 0$ (critical value = 0.219) and would also fail to reject that $\text{cov}(dy, \sum \rho^{j-1} r_{t+j}) = 0$ (critical value = 1.452) over the full sample. Hence, the asymptotic critical values are almost useless and lead to severe under-rejections. The Wald statistics for both the dividend and return variance decompositions have very long right hand tails, and values of over 100 were recorded in the simulations. However, the return variance decomposition Wald statistic is much more skewed than the Wald statistic for dividend growth, leading to much higher critical values. The reason is that shocks to returns and the dividend yield have a high negative correlation (-88.0% in Table 3), whereas shocks to dividend growth and the dividend yield have a low positive correlation (25.8%). When a negative shock hits returns, the dividend yield is also very likely to decline, leading to a spurious relationship in the small sample even though returns are not predictable by dividend yields under the null.

4.6 Summary of Small Sample Results

To summarize, the small sample size and power analysis yields the following results. First, the size distortions from Newey-West (1987) and Hodrick (1992) 1A standard errors are large,

lead to over-rejections of the null of predictability and are more severe for the regressions involving returns than the regressions involving dividend growth. In contrast, Hodrick (1992) 1B t-statistics have good size properties. Second, Hodrick 1B standard errors yield the most powerful t-statistics, relative to Newey-West and Hodrick 1-B standard errors. Third, the bias in the regressions using log dividend yields is negligible. Finally, we must use a small sample distribution to correctly interpret the significance levels of Wald statistics. These results suggest that of the three t-statistics, the Hodrick (1992) 1B t-statistic is the best choice for robust statistical inference and that there is no need to worry about the bias terms when using log dividend yields as predictive instruments.

4.7 Re-Interpreting the Predictive Regressions

Armed with the knowledge that Hodrick 1B t-statistics are the best choice for gauging the evidence of predictability, I can re-interpret the predictability regressions in Table 2 only with Hodrick 1B t-statistics. Over the full sample, at short horizons below three years, particularly for the monthly frequency, there is strong evidence that log dividend yields predict future dividend growth. In contrast, both with monthly and annual frequencies, there is no evidence for predictability of total returns by log dividend yields at any horizon.

Prior to 1990, there is much stronger evidence of predictability of both dividend growth and returns by log dividend yields. Future dividend growth is now predictable up to a four-year (two-year) forecasting horizon with monthly (annual) data. Returns are predictable at longer horizons (3-5 years), but are not predictable at the one-year horizon. Hence, without the 1990s, there is even stronger evidence of cashflow growth predictability.

5 Explaining the Variation of the Market Dividend Yield

5.1 Variance Decompositions Implied from Predictive Regressions

Table 6 reports the variance decompositions in Proposition 3.1 using the data-generating process (11). Using annual horizons in Panel A over the full sample, 88.0% of the variation in dividend yields is due to dividend growth and only 12.0% is due to returns. Omitting the 1990s increases the proportion due to discount rates to 35.0%. With monthly data in Panel B, the proportions due to cashflows are slightly lower, 68.8% over the whole sample and 54.2% over the sample without the 1990s. The reason why these point estimates are so high is that the

one-year regression coefficients for dividend growth are much larger (more than twice the size over the full sample) than the coefficients for total returns, in absolute magnitude.

Below the proportional variance decomposition estimates, I report Hodrick (1992) 1B standard errors. Since the variance decompositions only involve a ratio of the regression coefficients, from Proposition 3.1, the standard errors can be computed by using a bivariate regression system of dividend growth and returns onto log dividend yields. The variance decompositions for dividend growth are highly significant, but the variance decompositions for returns are not significant, due to the large standard errors associated with the one-year regressions with returns.

5.2 Variance Decompositions Implied from Wald Tests

In the right-hand side of Table 6, I conduct Wald tests for the significance of the variance decompositions $\text{cov}(dy, \sum \rho^{j-1} g_{t+j})$ and $\text{cov}(dy, \sum \rho^{j-1} r_{t+j})$. These tests also involve the parameters ρ_i of the dividend yield regression onto lagged dividend yields, lagged dividend growth and past return, which the predictive coefficient-based inference in Proposition 3.1 does not. In computing the covariance matrix I use only the regression coefficients with Hodrick 1B standard errors. Using asymptotic critical values, Wald tests cannot reject the hypothesis that the covariances are not equal to zero for either dividend growth or discount rates, with both annual and monthly frequencies, and including or omitting the 1990s. However, this inference is very misleading because of the extreme distortions of the Wald statistic in small samples (see Table 5). With small-sample adjusted critical values, the Wald tests for the variance decomposition of dividend growth are all significant in both sample periods, while the variance decomposition for returns are highly insignificant.

In summary, the variance decompositions implied from one-period predictive regressions assign a very large proportion to dividend growth, 88% over the full sample at an annual frequency. This attribution is significant both with robust standard errors from predictive regressions and also using Wald tests correcting for small sample distributions.

5.3 Variance Decompositions Implied from Weighted Long-Horizon Regressions

To use weighted long-horizon regressions (16) and (17) and to decompose the variance of the log dividend yield into percentage attributions due to dividend growth and discount rates, I first

compute ρ , and hold this constant in the regressions. The coefficient ρ is 0.960, estimated over the 1927-1990 sample, and 0.969, estimated over the full 1927-2000 sample. Naturally, the one-period weighted regressions (16) and (17) are exactly the same as the un-weighted $k = 1$ regressions in (18), reported in Table 2, since the weight on the one-year return is $\rho^0 = 1$.

I report the weighted long-horizon regressions of (16) and (17) in Table 7. Let us focus discussion on Panel A reporting the annual frequency regressions. The coefficients from the regressions of weighted future g on dy are negative, and absolute values of these coefficients represent the percentage variance of the log dividend yield attributable to dividend growth at horizon k , in years. For example, over 1927-2000 at a one-year (two-year) horizon, dividend growth accounts for 12.9% (18.8%) of the variance of the dividend yield. The entries for the regression of weighted future discount rates on dy can be directly read as the percentage variance decomposition. For example, over the full sample, discount rates account for 1.8% (10.6%) of variation in the dividend yield over one-year (two-year) horizons.

As the horizon increases from one to five years, the magnitude of the coefficients for weighted g tend to decrease, from 12.9% to 9.5% over the full sample. At the same time, the coefficients for weighted r increase, from 1.8% to 38.0%. At the five-year horizon, this still leaves 52.5% ($100\% - 9.5\% - 38.0\%$) attributable to $\text{cov}(dy_{t+5}, dy_t)$ from (14). The 1927-1990 sample increases these proportions dramatically. At the one-year (five-year) horizon, 23.2% (12.0%) of the variation of the dividend yield is due to the variation in dividend growth and 12.5% (63.0%) is attributable to the variation in returns. Hence, I can explain the dividend yield variation over the 1990s with much lower precision than over the sample excluding the 1990s.

Table 7 also reports the variance decompositions at the 15-year horizon, which corresponds to the horizon used in Cochrane (1992). Over the full sample, 24.0% of the dividend yield variance can be attributed to dividends and 90.0% to discount rates. From 1927-1990, these point estimates are very similar: the variance decomposition is 24.0% dividend growth and 87.0% returns. Cochrane (1992) uses NYSE data from 1926-1988, and finds the decomposition to be 8.4% (59.7%) dividends and 94.8% (55.6%) returns for the value-weighted (equal-weighted) index. The CRSP data counts more small firms, so my point estimates of the cashflow attributions are slightly larger than Cochrane's numbers. These numbers are very different from the estimates implied by the one-period regressions, because at the one-year horizon, dividends are much more predictable than returns by log dividend yields.

At the 15-year horizon, there are effectively less than five independent observations in the sample, and the long horizon makes size distortions a serious issue. When Hansen-

Hodrick (1980) standard errors (Hodrick (1992) 1A t-statistics) used by Fama and French (1988) and others are computed, one would conclude that the large variance decomposition to returns is highly significant (t -1A = 4.45 and 4.26 over 1927-1990 and 1927-2000, respectively). Also, if Newey-West (1987) standard errors are employed, as in Cochrane (1992), one would also conclude that the returns significantly account for almost all of the variation in log dividend yields (t -NW = 4.07 and 4.01 over 1927-1990 and 1927-2000, respectively). However, the Monte Carlo analysis shows that these t-statistics substantially over-reject. In contrast, using Hodrick (1992) 1B standard errors, the variance attributions to both dividend growth and returns are highly insignificant. Even over the period omitting the 1990s we cannot make any statistically significant statement about the long-term variance decompositions from weighted long-horizon regressions. I note that the use of monthly frequency data in Panel B does not change the inference very much. In particular, at the 15-year horizon, the point estimate of the percentage variance decomposition assigns 21% to dividend growth and 90% to discount rates, but with Hodrick 1B t-statistics these results are not statistically significant.

Nevertheless, the increasing point estimates of the long-horizon discount rate regressions suggest that variations in returns contribute more to long-term variations in the log dividend yield than variations in cashflows. Note that over the 1927-1990 period, t -1B statistics are significant for horizons 2-5 years for the return variance decompositions. According to the point estimates, dividend growth variation, although highly significantly, is a short-term transitory effect, while shocks to expected returns have more permanent effects. However, with robust t-statistics, we can make no formal statistically significant statement about the variance attribution at long horizons, with or without the 1990s. The only thing we can say for certain with weighted long-horizon regressions is that dividend growth significantly accounts for substantially more of the variation of log dividend yields at short horizons than returns.

6 Conclusion

High dividend yields significantly predict low future dividend growth. The predictability of dividend growth at short horizons (1-2 years) dominates the estimates of predictability of expected returns from dividend yields. Dividend growth predictability is even stronger when the 1990s are excluded from the sample. The strong short horizon predictability of dividend growth means that variance decompositions implied by one-period regressions assign over 85% of the variance of log dividend yields to dividend growth. This high variance decomposition

to dividends is highly significant using inference from regression coefficients computed with robust standard errors and Wald tests with critical values from small sample distributions.

Long-horizon regressions, and weighted-long horizon regressions, suggest that dividend growth predictability is strongest at short horizons (1-3 years) and is weak at long horizons. Over 1927-1990 there is evidence that expected return predictability is stronger at long horizons (3-5 years) and is insignificant at short horizons. Over the whole sample from 1927-2000, there is no evidence of the predictability of expected returns at any horizon. While the point estimates of the variance decompositions from weighted-long horizon regressions suggest that expected returns drive most of the long-term variation in log dividend yields, with robust covariance estimates the standard errors are so large that no significant attribution is possible. However, at short horizons, movements in dividend growth account for over twice the amount of variation in log dividend yields than movements in discount rates, and this short-term attribution to dividend growth is highly significant.

Although this study has focused only on U.S. data, the findings indicate that examining the importance of dividend growth and discount rate components in explaining the dynamics of dividend yields could be even more relevant for Asian-Pacific markets. Table 8 reports the capital gain and income components of total returns in local currencies, where the income yield is given in equation (1) over a recent sample from January 1993 to June 2011. Income returns account for 21% of the U.S. total return over this period. This is towards the low end; in this basket of Asian-Pacific markets, Australia, China, Hong Kong, Japan, New Zealand, Singapore, and Thailand all have income returns accounting for more than one quarter of total returns. In Japan, total returns have entirely been driven by dividends. Given the even larger importance of dividends as a proportion of total returns in these other Asian-Pacific markets, examining how dividend yields are affected by future expectations of dividend growth versus returns is a fruitful area for future research.

Appendix

A Proof of Proposition 3.1

Partition the companion matrix A as:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} [\rho_1] & [\rho_2 \ \rho_3] \\ \begin{bmatrix} \beta_g \\ \beta_r \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{pmatrix} \quad (\text{A-1})$$

I evaluate the variance decomposition $e'_i A(I - \rho A)^{-1} \Sigma_X e_1$, where $i = 2, 3$. Using a standard formula for the inverse of a partitioned matrix, I can write $(I - \rho A)^{-1}$ as:

$$(I - \rho A)^{-1} = \begin{pmatrix} D & -DA_{12} \\ -A_{21} & I + A_{21}DA_{12} \end{pmatrix}, \quad (\text{A-2})$$

where $D = 1/(\rho_1 - \rho_2\beta_g - \rho_3\beta_r)$. Hence, the bottom two rows of $A(I - \rho A)^{-1}$ can be written as:

$$\begin{pmatrix} \dots \\ D\beta_g & -D\beta_g\rho_2 & -D\beta_g\rho_3 \\ D\beta_r & -D\beta_r\rho_2 & -D\beta_r\rho_3 \end{pmatrix}.$$

Since $\Sigma_X e_1 = (\sigma_{X,1}^2 \ \sigma_{X,21} \ \sigma_{X,31})'$, multiply out the terms to compute the variance decomposition of log dividend yields due to g as:

$$-e'_2 A(I - \rho A)^{-1} \Sigma_X e_1 = -\beta_g \times \frac{\sigma_{X,1}^2 - \rho_2 \sigma_{X,21} - \rho_3 \sigma_{X,31}}{\rho_1 - \rho_2\beta_g - \rho_3\beta_r}, \quad (\text{A-3})$$

and the variance decomposition of log dividend yields due to r as:

$$e'_3 A(I - \rho A)^{-1} \Sigma_X e_1 = \beta_r \times \frac{\sigma_{X,1}^2 - \rho_2 \sigma_{X,21} - \rho_3 \sigma_{X,31}}{\rho_1 - \rho_2\beta_g - \rho_3\beta_r}. \quad (\text{A-4})$$

The terms (A-3) and (A-4) share the same coefficient $\frac{\sigma_{X,1}^2 - \rho_2 \sigma_{X,21} - \rho_3 \sigma_{X,31}}{\rho_1 - \rho_2\beta_g - \rho_3\beta_r}$, which cancels when taking ratios.

References

- [1] Ammer, J., and M. Jianping, 1996, "Measuring International Economic Linkages with Stock Market Data," *Journal of Finance*, 51, 5, 1743-1763.
- [2] Ang, A., and G. Bekaert, 2007, "Stock Return Predictability: Is it There?" *Review of Financial Studies*, 20, 651-707.
- [3] Ang, A., and J. Liu, 2001, "A General Affine Earnings Valuation Model," *Review of Accounting Studies*, 6, 397-425.
- [4] Ang, A., and J. Liu, 2007, "Risk, Return, and Dividends," *Journal of Financial Economics*, 81, 1-38.
- [5] Bakshi, G., and Z. Chen, 2005, "Stock Valuation in Dynamic Economies," *Journal of Financial Markets*, 8, 111-151.
- [6] Bansal, R., and A. Yaron, 2004, "Risk for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481-1509.
- [7] Bansal, R., R. Dittmar, and C. Lundblad, 2005, "Consumption, Dividends, and the Cross-Section of Equity Returns," *Journal of Finance*, 60, 1639-1722.
- [8] Bekaert, G., E. Engstrom, and S. Grenadier, 2010, "Stock and Bond Returns with Moody Investors," *Journal of Empirical Finance*, 17, 876-894.
- [9] Bekaert, G., and R. Hodrick, 2001, "Expectations Hypothesis Tests," *Journal of Finance*, 56, 1357-1394.
- [10] Berkowitz, J., and L. Giorgianni, 2001, "Long-Horizon Exchange Rate Predictability?" *Review of Economics and Statistics*, 83, 1, 81-91.
- [11] van Binsbergen, J. H., and R. S. J. Koijen, 2010, "Predictive Regressions: A Present-Value Approach," *Journal of Finance*, 65, 1439-1471.
- [12] Burnside, C., and M. Eichenbaum, 1996, "Small Sample Properties of GMM-based Wald Tests," *Journal of Business and Economic Statistics*, 14, 294-308.
- [13] Campbell, J. Y., 1991, "A Variance Decomposition for Stock Returns," *Economic Journal*, 101, 405, 157-179.
- [14] Campbell, J. Y., 2001, "Why Long Horizons? A Study of Power Against Persistent Alternatives," *Journal of Empirical Finance*, 8, 459-491.
- [15] Campbell, J. Y., and J. Ammer, 1993, "What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns," *Journal of Finance*, 48, 1, 3-37.
- [16] Campbell, J. Y., and R. J. Shiller, 1988a, "Stock Prices, Earnings and Expected Dividends," *Journal of Finance*, 43, 3, 661-676.
- [17] Campbell, J. Y., and R. J. Shiller, 1988b, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 3, 195-228.
- [18] Campbell, J. Y., and R. J. Shiller, 1989, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 3, 195-228.
- [19] Campbell, J. Y., and S. Thompson, 2008, "Predicting Excess Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, 21, 1509-1531.

- [20] Campbell, J. Y., and M. Yogo, 2006, "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81, 27-60.
- [21] Chen, L., 2009, "On the Reversal of Return and Dividend Growth Predictability: A Tale of Two Periods," *Journal of Financial Economics*, 92, 128-151.
- [22] Chen, L., and X. Zhao, 2009, "Return Decomposition," *Review of Financial Studies*, 22, 5213-5249.
- [23] Cochrane, J. H., 1992, "Explaining the Variance of Price-Dividend Ratios," *Review of Financial Studies*, 5, 243-280.
- [24] Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton.
- [25] Cochrane, J. H., 2011, "Presidential Address: Discount Rates," *Journal of Finance*, 66, 1047-1108
- [26] Fama, E., and F. French, 1988, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3-26.
- [27] Goyal, A., and I. Welch, 2008, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455-1508.
- [28] Hansen, L. P., J. Heaton, and N. Li., 2008, "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116, 260302.
- [29] Hansen, L., J. Heaton, and A. Yaron, 1996, "Finite-Sample Properties of some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262-280.
- [30] Hansen, L., and R. Hodrick, 1980, "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88, 829-853.
- [31] Hodrick, R., 1992, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies*, 5, 3, 357-386.
- [32] Lettau, M., and S. Van Nieuwerburgh, 2008, "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, 21, 1607-1652.
- [33] Lettau, M., and J. Watcher, 2007, "Why Is Long-Horizon Equity Less Risky? A Duration-Based Explanation of the Value Premium," *Journal of Finance*, 62, 5592.
- [34] Lewellen, J., 2004, "Predicting Returns with Financial Ratios," *Journal of Financial Economics*, 74, 209-235.
- [35] Mamaysky, H., 2002, "On the Joint Pricing of Stocks and Bonds: Theory and Evidence," working paper, Yale University.
- [36] Mark, N. C., and D. Sul, 2002, "Asymptotic Power Advantages of Long-Horizon Regressions," working paper, Ohio State University.
- [37] Nelson, C., and M. Kim, 1993, "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance*, 48, 2, 641-661.
- [38] Newey, W., and K. West, 1987, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.
- [39] Richardson, M., and T. Smith, 1991, "Tests of Financial Models in the Presence of Overlapping Observations," *Review of Financial Studies*, 4, 2, 227-254.

- [40] Richardson, M., 1993, "Temporary Components of Stock Prices: A Skeptic's View," *Journal of Business and Economic Statistics*, 11, 2, 199-207.
- [41] Shiller, R. J., 1981, "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71, 3, 421-436.
- [42] Shiller, R.J., and A. E. Beltratti, 1992, "Stock Prices and Bond Yields: Can Their Comovements be Explained in Terms of Present Value Models?" *Journal of Monetary Economics*, 30, 25-46.
- [43] Stambaugh, R. F., 1999, "Predictive Regressions," *Journal of Financial Economics*, 54, 375-421.
- [44] Valkanov, R., 2003, "Long-Horizon Regressions: Theoretical Results and Applications," *Journal of Financial Economics*, 68, 201-232.
- [45] Vuolteenaho, T., 2002, "What Drives Firm-Level Stock Returns?" *Journal of Finance*, 57, 1, 233-264.

Table 1: Summary Statistics

	Jan 1927 - Dec 1990			Jan 1927 - Dec 2000		
	Mean	Std	Auto	Mean	Std	Auto
Log Dividend Yield	-3.173	0.286	0.668	-3.269	0.382	0.761
Dividend Growth	0.044	0.133	0.300	0.042	0.124	0.302
Total Equity Return	0.094	0.204	0.102	0.103	0.196	0.091

The table shows summary statistics for data on the CRSP value-weighted market portfolio. All data is at an annual frequency. All returns and growth rates are continuously compounded. Auto denotes the autocorrelation.

Table 2: Long-Horizon Predictive Regressions

	Hor k	Jan 1927 - Dec 1990				Jan 1927 - Dec 2000			
		Coef	t-NW	t-1A	t-1B	Coef	t-NW	t-1A	t-1B
Panel A: Univariate Regressions at an Annual Frequency									
g on dy	1	-0.232	-2.33*	-2.76**	-2.76**	-0.129	-1.79	-2.12*	-2.12*
	2	-0.150	-1.81	-1.70	-2.33*	-0.095	-1.46	-1.37	-1.74
	3	-0.090	-1.51	-1.40	-1.79	-0.065	-1.29	-1.17	-1.35
	4	-0.046	-1.25	-1.38	-1.19	-0.037	-1.09	-1.14	-0.89
	5	-0.019	-0.94	-0.94	-0.54	-0.016	-0.82	-0.79	-0.41
r on dy	1	0.125	1.14	1.27	1.27	0.018	0.23	0.27	0.27
	2	0.167	2.63**	2.32*	1.96*	0.054	0.82	0.76	0.81
	3	0.151	3.73**	4.10**	2.15**	0.062	1.09	1.03	0.96
	4	0.147	4.73**	5.72**	2.26**	0.075	1.49	1.36	1.16
	5	0.135	5.34**	4.76**	2.25**	0.080	1.90	1.67	1.27
Panel B: Univariate Regressions at a Monthly Frequency									
g on dy	1	-0.201	-2.71**	-2.15*	-6.07**	-0.100	-1.89	-1.50	-4.35**
	2	-0.127	-1.93	-1.60	-4.71**	-0.073	-1.45	-1.19	-3.30**
	3	-0.071	-1.50	-1.37	-3.23**	-0.046	-1.18	-1.05	-2.19*
	4	-0.037	-1.20	-1.21	-2.01*	-0.027	-1.02	-0.97	-1.34
	5	-0.015	-0.83	-0.85	-0.92	-0.012	-0.72	-0.68	-0.59
r on dy	1	0.179	2.03*	1.72	1.49	0.051	0.83	0.69	0.63
	2	0.181	3.84**	4.22**	1.82	0.062	1.19	1.03	0.79
	3	0.166	4.57**	5.61**	2.08*	0.066	1.33	1.12	0.89
	4	0.155	5.31**	4.90**	2.37*	0.073	1.49	1.22	1.04
	5	0.145	6.31**	5.33**	2.50**	0.080	1.85	1.54	1.18

The table reports regression coefficients (coef) b in the predictive regressions: $\frac{1}{k}z_{t+k} = a + b dy_t + \epsilon_{t,t+k}$ for annual log dividend yields dy_t predicting $z = g, r$, dividend growth and total equity returns, respectively, over horizon (hor) k in years. Data on the CRSP market value-weighted portfolio is used. T-statistics computed using Newey-West (1987) standard errors are denoted as t-NW, Hodrick (1992) 1-A standard errors as t-1A, and Hodrick (1992) 1-B standard errors as t-1B. T-statistics significant at the 5% (1%) level are denoted by * (**).

Table 3: Estimation of the Predictability System

	dy	g	r
Predictability Coefficients			
dy Regression			
	ρ_1	ρ_2	ρ_3
dy	0.953 (0.184)	0.469 (0.117)	0.228 (0.061)
dy Predicting g			
	β_g		
g	0.129 (0.061)		
dy Predicting r			
	β_r		
r	0.018 (0.067)		
Covariance of Shocks			
	dy	g	r
	0.042	[0.258]	[-0.880]
	g	0.006	[0.125]
	r	-0.035	0.038

The table reports estimates of the predictability coefficients of the system (11), where dy is the log-dividend yield, g is dividend growth, and r denotes total returns. Standard errors for each equation are computed using Hodrick (1992) 1-B standard errors and are reported in parentheses. Using Hodrick 1-B standard errors, a joint estimation of the null against the alternative of an unrestricted VAR is not rejected (p-value = 0.07). The covariance matrix lists correlations in square brackets. The estimation is conducted at the annual frequency with a one-year horizon from Dec 1927 to Dec 2000.

Table 4: Size and Power of the T-Statistics

	Hor	Nominal Size = 5%			Nominal Size = 10%		
		t-NW	t-1A	t-1B	t-NW	t-1A	t-1B
Panel A: H0: $\beta_g = 0$, H1: $\beta_g \neq 0$							
Empirical Size	1	0.067	0.057	0.057	0.102	0.102	0.102
	2	0.099	0.080	0.052	0.173	0.133	0.103
	3	0.143	0.102	0.048	0.212	0.167	0.100
	4	0.171	0.135	0.043	0.235	0.199	0.105
	5	0.192	0.159	0.044	0.265	0.228	0.105
Size-Adjusted Power	1	0.945	0.950	0.950	0.977	0.974	0.974
	2	0.902	0.907	0.933	0.956	0.956	0.959
	3	0.859	0.866	0.909	0.918	0.917	0.939
	4	0.771	0.773	0.875	0.868	0.857	0.913
	5	0.702	0.711	0.839	0.790	0.787	0.881
Bias of $\hat{\beta}_g \times 1000$	1	-1.587					
	2	-1.269					
	3	-1.194					
	4	-1.054					
	5	-1.009					
Panel B: H0: $\beta_r = 0$, H1: $\beta_r \neq 0$							
Empirical Size	1	0.080	0.076	0.076	0.139	0.127	0.127
	2	0.133	0.102	0.073	0.207	0.173	0.134
	3	0.170	0.129	0.066	0.247	0.202	0.128
	4	0.191	0.165	0.065	0.279	0.235	0.123
	5	0.214	0.186	0.065	0.303	0.268	0.112
Size-Adjusted Power	1	0.319	0.327	0.327	0.444	0.447	0.447
	2	0.303	0.294	0.312	0.420	0.423	0.433
	3	0.273	0.249	0.281	0.410	0.406	0.414
	4	0.269	0.241	0.271	0.385	0.364	0.392
	5	0.239	0.200	0.253	0.365	0.360	0.385
Bias of $\hat{\beta}_r \times 1000$	1	6.598					
	2	6.358					
	3	5.992					
	4	5.768					
	5	5.542					

Note to Table 4

Panel A examines empirical size and power of Newey-West (1987) (t-NW), Hodrick (1992) 1A (t-1A), and Hodrick (1992) 1B (t-1B) t-statistics under the system of Table 3 for a small sample of 73 years. We simulate 10,000 draws for each exercise. In Panel A, I examine the null of no predictability of dividend growth ($\beta_g = 0$) versus the alternative of $\beta_g = -0.129$, holding all other parameters constant. To compute empirical size, in each small sample simulated under the null, I run the regression of cumulated dividend growth at various horizons (hor) in years onto dy and record the percentage number of rejections of the null of the t-statistics corresponding to nominal size levels of 5% and 10% significance. To compute power, I simulate under the alternative $\beta_g = -0.129$ and record the percentage number of rejections of the null hypothesis using the size-adjusted t-statistic from the simulations in Panel A. The bias of the estimated $\hat{\beta}_g$ coefficients is also recorded. In Panel B, I run a similar exercise with the null of no predictability of total expected returns ($\beta_r = 0$) versus the alternative of $\beta_y = 0.018$, holding all other parameters constant. All tests are two-sided tests.

Table 5: Critical Values of Wald Statistics

Nominal Size	Jan 1927 - Dec 1990		Jan 1927 - Dec 2000	
	5%	10%	5%	10%
$\text{cov}(dy, \sum \rho^{j-1} g_{t+j}) = 0$	0.202	0.137	0.219	0.104
$\text{cov}(dy, \sum \rho^{j-1} r_{t+j}) = 0$	2.837	1.987	1.452	1.058
Nominal	3.841	2.706	3.841	2.706

The table lists small sample critical values of the Wald statistics, testing if the variance decomposition due to growth rates, $\text{cov}(dy, \sum \rho^{j-1} g_{t+j})$ in equation (9), or the variance decomposition due to returns $\text{cov}(dy, \sum \rho^{j-1} r_{t+j})$ in equation (10) is equal to zero. To test $\text{cov}(dy, \sum \rho^{j-1} g_{t+j}) = 0$, I simulate the data-generating process (11) under the null that dividend growth is unpredictable, $\beta_g = 0$, estimate the parameters in (11) from the small sample, and compute the Wald statistic (9). Similarly, to test $\text{cov}(dy, \sum \rho^{j-1} r_{t+j}) = 0$, I simulate (11) under the null that returns are unpredictable, $\beta_r = 0$, and record the Wald statistic (10). I report the Wald statistics corresponding to nominal size levels of 5% and 10% significance levels. The nominal size levels are the 90% and 95% cumulative distribution critical values from a χ_1^2 distribution. The Wald statistics are computed using 10,000 simulations with a covariance matrix using only the moment conditions from the regression coefficients in (9) with Hodrick (1992) 1B standard errors. For the first period January 1927 to December 1990, the process (11) is estimated only on this subsample, where as for January 1927 to December 2000, the parameter estimates in Table 3 are used. In each case, to compute the small sample distribution of the Wald statistics, I only change the parameters β_g or β_r and leave the other parameters set at their estimated values.

Table 6: Variance Decomposition of the Market Log Dividend Yield

Variance Decomposition			Wald Tests		
	% Dividend Growth	% Discount Rates	$\text{cov}(dy, \sum \rho^{j-1} g_{t+j}) = 0$	$\text{cov}(dy, \sum \rho^{j-1} r_{t+j}) = 0$	
Panel A: Annual Frequency					
1927-1990	Estimate	0.650	0.350	1.85	1.05
	Std Err	0.257	0.242	0.174	0.306
			Adjusted p-val	0.000	0.233
			Wald statistic		
			Asymptotic p-val		
1927-2000	Estimate	0.880	0.120	2.77	0.81
	Std Err	0.333	0.333	0.096	0.808
			Adjusted p-val	0.012	0.819
			Wald statistic		
			Asymptotic p-val		
			Adjusted p-val		
Panel B: Monthly Frequency					
1927-1990	Estimate	0.542	0.458	2.10	1.46
	Std Err	0.228	0.228	0.147	0.227
			Adjusted p-val	0.000	0.148
			Wald statistic		
			Asymptotic p-val		
			Adjusted p-val		
1927-2000	Estimate	0.688	0.312	0.16	0.14
	Std Err	0.304	0.304	0.688	0.707
			Adjusted p-val	0.046	0.655
			Wald statistic		
			Asymptotic p-val		
			Adjusted p-val		

Under “Variance Decomposition,” the table reports the variance decomposition due to dividend growth and total expected returns using Proposition 3.1 from the process (11), based on the univariate predictive regression estimates. Under “Wald Tests,” the table reports Wald tests of $\text{cov}(dy, \sum \rho^{j-1} g_{t+j}) = 0$ (equation (9)) and $\text{cov}(dy, \sum \rho^{j-1} r_{t+j}) = 0$ (equation (10)), using only the moments from the coefficients in (11) with Hodrick (1992) 1B standard errors to compute the covariance matrix. The adjusted p-values use the small sample distributions from Table 5.

Table 7: Weighted-Long Horizon Regressions

		Jan 1927 - Dec 1990				Jan 1927 - Dec 2000			
	Hor k	Coef	t-NW	t-1A	t-1B	Coef	t-NW	t-1A	t-1B
Panel A: Weighted Regressions at the Annual Frequency									
g on dy	1	-0.232	-2.33*	-2.76**	-2.76**	-0.129	-1.79	-2.12*	-2.12*
	2	-0.296	-1.82	-1.71	-2.31*	-0.188	-1.47	-1.38	-1.72
	3	-0.270	-1.53	-1.41	-1.79	-0.195	-1.31	-1.19	-1.34
	4	-0.196	-1.31	-1.39	-1.25	-0.152	-1.14	-1.16	-0.92
	5	-0.120	-1.11	-1.09	-0.67	-0.095	-0.95	-0.90	-0.49
	15	-0.240	-1.51	-2.02	-0.67	-0.240	-1.57	-2.18*	-0.66
r on dy	1	0.125	1.14	1.27	1.27	0.018	0.23	0.27	0.27
	2	0.326	2.59**	2.29*	1.91	0.106	0.81	0.75	0.79
	3	0.580	3.67**	4.01**	2.07*	0.180	1.09	1.03	0.93
	4	0.556	4.68**	5.66**	2.13*	0.288	1.51	1.38	1.11
	5	0.630	5.24**	4.63**	2.10*	0.380	1.93	1.70	1.21
	15	0.870	4.07**	4.45**	1.24	0.900	4.01**	4.26**	1.08
Panel B: Weighted Regressions at the Monthly Frequency									
g on dy	1	-0.201	-2.71**	-2.15*	-6.07**	-0.100	-1.89	-1.50	-4.35**
	2	-0.252	-1.94	-1.61	-4.67**	-0.144	-1.46	-1.20	-3.28**
	3	-0.213	-1.54	-1.39	-3.25**	-0.138	-1.21	-1.07	-2.21*
	4	-0.156	-1.27	-1.26	-2.13*	-0.112	-1.07	-1.10	-1.42
	5	-0.095	-1.03	-1.04	-1.14	-0.075	-0.88	-0.81	-0.73
	15	-0.210	-1.39	-1.66	-1.08	-0.210	-1.51	-2.01	-0.89
r on dy	1	0.179	2.03*	1.72	1.49	0.051	0.83	0.69	0.63
	2	0.356	3.80**	4.18**	1.78	0.122	1.19	1.02	0.78
	3	0.480	4.57**	5.64**	2.01*	0.192	1.33	1.12	0.87
	4	0.588	5.32**	4.90**	2.25*	0.280	1.51	1.24	1.00
	5	0.680	6.26**	5.30**	2.34*	0.380	1.88	1.57	1.12
	15	0.885	4.31**	4.37**	1.40	0.900	4.72**	4.89**	1.12

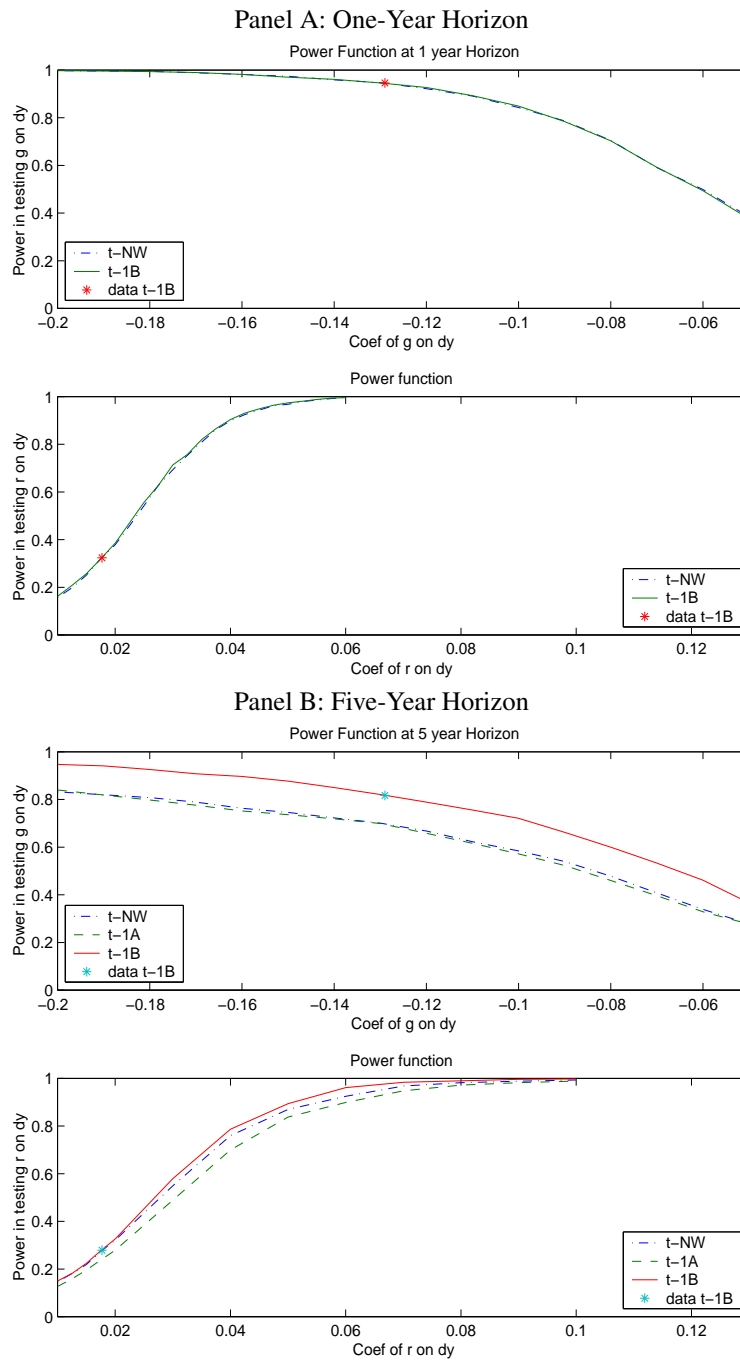
The table reports regression coefficients (Coef) β in the predictive regressions: $\sum_{j=1}^k \rho^{j-1} z_{t+j} = a + \beta dy_t + \epsilon_{t,t+k}$ for annual log dividend yields dy_t predicting weighted $z = g, r$, dividend growth and total equity returns and excess returns, respectively, over horizon (hor) k in years. The coefficient ρ is 0.960 over 1927-1990 and 0.969 over 1927-2000. Data on the CRSP market value-weighted portfolio is used. T-statistics computed using Newey-West (1987) standard errors are denoted as t-NW, Hodrick (1992) 1-A standard errors as t-1A, and Hodrick (1992) 1-B standard errors as t-1B. T-statistics significant at the 5% (1%) level are denoted by * (**).

Table 8: Capital Gain and Income Components of Total Returns

	Capital Gain	Income Return	Total Return	Proportion Income
Australia	0.0647	0.0382	0.1030	37%
China	0.0428	0.0231	0.0659	35%
Hong Kong	0.0959	0.0336	0.1295	26%
India	0.1488	0.0159	0.1648	10%
Indonesia	0.1985	0.0262	0.2247	12%
Japan	-0.0032	0.0113	0.0081	140%
South Korea	0.1391	0.0167	0.1558	11%
Malaysia	0.0932	0.0249	0.1182	21%
New Zealand	0.0261	0.0509	0.0770	66%
Philippines	0.0797	0.0198	0.0996	20%
Singapore	0.0689	0.0240	0.0929	26%
Taiwan	0.0872	0.0210	0.1082	19%
Thailand	0.0851	0.0286	0.1137	25%
United States	0.0733	0.0194	0.0927	21%

The table reports capital gain and income components of total returns in local currency using MSCI country returns from January 1993 to June 2011. The income component is given by equation (1). I use monthly frequency MSCI arithmetic returns and annualize by multiplying the means by 12. The last column, the Proportion Income, is the Income Return divided by the Total Return.

Figure 1: Power Function of T-Statistics



The top panel shows power of the Newey-West (1987) (t-NW) t-statistic and Hodrick (1992) 1B (t-1B) t-statistics at a one-year horizon for the regressions involving dividend growth (top plot) and expected returns (bottom plot). Note that at the one-year horizon, the Hodrick (1992) 1A t-statistics are identical to the 1B t-statistics. The bottom panel shows the power of the various t-statistics at the five-year horizon for dividend growth (top plot) and expected returns (bottom plot). On each x -axis we report different coefficients of the one-year predictability of dividend growth or expected returns by dividend yields, in the data-generating process (11). The power corresponding to the coefficients estimated from data over January 1927 to December 2000 are labeled with asterisks.