
JOHN G. POWELL*

JING SHI

TOM SMITH

DIVIDEND PERSISTENCE AND RETURN PREDICTABILITY

Abstract

Evidence of dividend yield return predictability has been presented so widely and consistently that the result has tended to be generally accepted. This paper shows that return predictability of the dividend yield is a spurious result that is due to dividend persistence and finds that standard dividend behaviour explanatory models are also affected by the spurious regression problem. A simulation procedure is utilized to take account of a spurious correlation that compounds the spurious regression problem when the dependent and independent variables in a time series regression are ratios composed of common component variables. The paper's results therefore imply that extreme care should be taken when using ratios as predictor or explanatory variables in time series regression. The paper introduces a reformulated Lintner first difference dividend behaviour model that is not subject to spurious regression.

This version: November 2, 2004

Please do not quote, copy or cite without permission of the authors.

*Corresponding author. Department of Finance Banking and Property, Massey University, Private Bag 11 222, Palmerston North, New Zealand. Email: J.Powell@massey.ac.nz. Fax 646 350 5651, Phone 646 356 9099 ext 2333.

Powell is from Department of Finance Banking and Property, Massey University, Palmerston North, New Zealand; Smith and Shi are from the School of Finance and Applied Statistics, The Australian National University, Canberra, ACT 0200, Australia. Smith's Email: Tom.Smith@anu.edu.au, and Shi's Email Jing.Shi@anu.edu.au

DIVIDEND PERSISTENCE AND RETURN PREDICTABILITY

1. INTRODUCTION

Evidence of dividend yield return predictability has been presented so widely and consistently that the result has tended to be generally accepted, with debate instead shifting to whether the predictability is due to risk and return considerations or other explanations such as market inefficiency. This paper shows that return predictability of the dividend yield is a spurious result that is due to dividend persistence, and it finds that standard dividend behaviour explanatory models are also strongly affected by the spurious regression problem. The paper demonstrates how the spurious regression problem is compounded when the dependent and independent variables in a regression equation are ratios constructed from common component variables, as happens with dividend yield return predictability and dividend behaviour regression models. A simulation procedure is utilized to take account of this problem, and the paper derives a reformulated dividend explanatory model to indicate how this econometric problem can be avoided.

A first hint that return predictability of the dividend yield is spurious can be obtained by examining how the dividend yield predicts returns relative to other persistent variables. A standard approach to test for dividend yield return predictability is to regress the time $t+1$ stock index rate of return (r_{t+1}) against the time t dividend yield (D_t/P_t) using the regression equation

$$r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1 \left[\frac{D_t}{P_t} \right] + \varepsilon_{t+1}, \quad (1)$$

where D_t is the level of real annual dividends during the twelve months preceding time t and P_t is the real stock index level at time t . The adjusted R^2 statistic for regression equation (1) for the annual CRSP value-weighted NYSE Index during the time period 1927 to 1996 is close to 3%, for instance, and the results are usually interpreted in relation to a correlation between the current level of dividends and subsequent returns (see, e.g., Campbell, Lo and MacKinlay, 1997).¹ It can be noted, however, that if variation in dividends plays an important role in the return predictability relationship then substituting a constant level of dividends (c) for the dividend term (D_t) in the numerator of the dividend yield variable (D_t/P_t) in regression equation (1) should reduce or eliminate return predictability in the following altered regression equation

$$r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_2 \left[\frac{c}{P_t} \right] + \varepsilon_{t+1}, \quad (2)$$

where c is the unconditional average of the dividend level during the sample period. Surprisingly, the adjusted R^2 for equation (2) actually rises to above 4% for the value-weighted NYSE Index. This unexpected result is a first indication that a spurious effect might be occurring and is a very strong hint of a spurious regression problem.

To explore this possibility, the persistence properties of the independent variables in regression equations (1) and (2) are examined. They are found to be highly persistent, and this persistence combines with return autocorrelation to create spurious return predictability (see Ferson, Sarkissian and Simin, 2003a; Foster, Smith and Whaley,

¹ Lewellen (2004) points out that dividend yield return predictability could also be due to temporary mispricing, an interpretation that will be discussed in Section 3.

1997).² The spurious effect of regressing returns against a persistent dividend yield variable is then shown to be strongly reinforced by a spurious correlation that is due to returns on the left hand side of equation (1) and the dividend yield on the right hand side being ratios that are constructed from the same underlying variables (the share index level and the dividend level). Spurious correlation due to common regression variable components therefore combines with the persistence properties of the dividend yield to account for the apparent dividend yield return predictability, rather than (as widely argued) any property of the dividend level that is related to risk, return or mispricing.

Given the strong persistence properties of dividends, it is not surprising to find that standard dividend explanatory models such as Lintner (1956) and Marsh and Merton (1987) are also strongly affected by the spurious regression problem, since the dependent and independent variables in these models are all very persistent (see also Ferson, Sarkissian and Simin, 2003a). To correct this problem, the paper demonstrates how the Lintner behavioural model of dividends can be reformulated and extended using Marsh and Merton (1987) so that the model is not subject to spurious regression.

The structure of the paper is as follows. Section 2 contains a review of the dividend yield return predictability and dividend behaviour literature. The spurious regression literature is then reviewed to indicate the potential problems that might be present with the return predictability and dividend behaviour literature. Section 3 shows that something “funny” is going on in the return predictability literature by using other highly persistent explanatory variables to improve on standard dividend yield return

² Campbell, Lo and MacKinlay (1997) demonstrate that stock index returns are autocorrelated, for instance, and Ferson, Sarkissian and Simin (2003a) point out that even if returns are not highly persistent, underlying expected returns are persistent, so there is a risk of spurious regression when testing return predictability.

predictability results, and then shows that all the results are spurious. Section 4 demonstrates how standard dividend behaviour regression equations are also subject to spurious regression. Section 5 corrects this problem by reformulating the Lintner (1956) dividend model. Section 6 provides a brief conclusion to the paper.

2. LITERATURE REVIEW

2.1 *Dividend Yield Return Predictability*

The literature on stock return predictability is extensive, with the dividend yield arguably being the best-known of the many different variables that are found to have forecasting power for stock returns. The dividend yield is usually measured as the ratio of the previous year's dividend payments to the current share price index level. This process removes the influence of strong seasonal effects in dividend payments, but it also artificially raises autocorrelation in monthly and quarterly overlapping dividend yield series.

Many studies find that dividend yields predict a substantial amount of the cross-sectional and times series variation of stock returns (see, e.g., Litzenberger and Ramaswamy, 1982; Fama and French, 1988; Harvey, 1989; Ferson and Harvey, 1991; Whitelaw, 1994; Pesaran and Timmermann, 1995; Pontiff and Schall, 1998; Bossaerts and Hillion, 1999; Cremers, 2002; and Lewellen, 2004). Table 1 summarizes these results. Kothari and Shanken (1997) present evidence that the dividend yield also tracks variation in expected real one-year stock returns. Litzenberger and Ramaswamy (1982) report a positive but non-linear association between stock returns and dividend yields. Fama and French (1988), Campbell and Shiller (1988a) and Poterba and Summers (1988)

find that dividend yields are positively related to the future returns of long-term bonds as well as stocks. Lewellen (2004) finds evidence in favour of return predictability using the natural logarithm of the dividend yield once a small sample bias correction is improved to more precisely account for a coefficient estimate bias that is induced by a strong correlation between the dividend yield slope coefficient and the dividend yield's autocorrelation. Campbell and Yogo (2003) find that the Lewellen (2004) bias-corrected test has poor power relative to Bonferroni inequality probability confidence interval tests when persistence does not equal unity, but never-the-less find weak evidence in support of dividend yield return predictability using a Uniformly Most Powerful (UMP) test. The Campbell and Yogo (2003) UMP test subtracts from the dividend yield dependent variable the innovations in the yield that are correlated with returns in order to reduce the independent variable's noise and consequently to increase the power of the return predictability regression coefficient significance test.

[Table 1 about here]

In contrast, only a few studies indicate that there is not a strong statistical relationship between dividend yields and stock returns, including Black and Scholes (1974) and Goetzmann and Jorion (1993, 1995). Bossaerts and Hillion (1999) use a number of statistical model selection criteria to examine the predictability of stock returns using dividend yields and find in-sample predictability but no out-of-sample forecasting power. Goyal and Welch (2003) find the predictive power of the dividend yield is present in pre- but not post-1990 data. Stambaugh (1999) finds that return predictability disappears when the bias induced by a correlation between the regression error and innovations in the autocorelated dividend yield regressor is accounted for using

Bayesian autocorrelation priors. Ferson, Sarkissian and Simin (2003a) indicate that the dividend yield's predictive power for monthly returns is questionable when account is taken of the spurious regression problem combined with data mining.

Dividend yields have also been used to predict long horizon returns (e.g. Fama and French, 1988; Campbell, Lo and MacKinlay, 1997). Fama and French (1988) estimate regressions of returns on the lagged dividend yield using post-war NYSE index data for return horizons from one month to four years. They observe that the dividend yield explains a significant proportion of multiple year returns, and the explanatory power of the dividend yield increases with the return horizon. Hodrick (1992) uses three alternative methods of conducting inference and measurement for long-horizon forecasting and finds that changes in dividend yields forecast significantly persistent changes in expected stock returns.

Theoretical explanations for the predictive power of the dividend yield have been developed using the dividend discount model (see, e.g., Campbell and Shiller, 1988a and 1988b; Fama and French, 1988; Donaldson and Kamstra, 1996; and Campbell, Lo and MacKinlay, 1997). Campbell and Shiller (1988a, 1988b) and Fama and French (1988) argue that the dividend discount model implies that a high current dividend level relative to the share price index level predicts some combination of either higher expected future returns or lower future dividends. A high current dividend yield therefore forecasts higher future returns if expected future dividends are held constant (e.g., under the perhaps strong assumption of "all else being equal"), thus providing a theoretical basis for dividend yield return predictability.

Although the dividend yield return predictability literature is vast and theoretical explanations for predictability are also well-known, this literature has not been directly connected to the equally well-known literature on behavioural explanations of dividends (see, e.g., Lintner, 1956; Marsh and Merton, 1987). A review of the dividend behaviour literature provides insights as to why dividend yield return predictability regression models are likely to be strongly affected by dividend persistence.

2.2 *Dividend Behaviour*

The starting point for a discussion of behavioural models of dividends is the Lintner (1956) speed of adjustment model (see, eg., Marsh and Merton, 1987). Lintner (1956) argues that corporate managers feel that they have a duty to pay out a proportion of earnings to shareholders, but are reluctant to increase dividends too quickly in reaction to an increase in earnings in case the earnings increase turns out to be temporary and the dividend increase subsequently has to be reversed. The Lintner (1956) model of dividends therefore implies that the current change in dividends is equal to a target dividend payout minus last period's dividend (D_{t-1}) times a speed of adjustment factor, plus a constant. The target payout is equal to the current level of trailing annual earnings (E_t) times a long run payout ratio target. The current dividend level (D_t) does not adjust instantaneously to the target payout level, thus avoiding the situation where an increase in earnings is only temporary and would have to be reversed in future years. Lintner's dividend model (equation (2) in Lintner, 1956) therefore states that the time t dividend is given by

$$D_t = \theta_0 + \theta_1 E_t + \theta_2 D_{t-1} + \varepsilon_t, \quad (3)$$

where t is in years, ε_t is the error term, $\theta_1/(1-\theta_2)$ is the long-run target payout ratio, and $(1-\theta_2)$ is a speed-of adjustment factor whose value will be closer to zero the more slowly that firms adjust their dividends to their long-run target level. Lintner (1956) reports an R^2 in excess of 90% when testing this model (regression equation (3)) using aggregate dividends and earnings data. Grullon and Michaely (2002) use the Lintner (1956) model to generate dividend forecast errors for companies in order to test whether share repurchases are substitutes for dividends, thus updating the Lintner (1956) results for individual companies, but they do not test the Lintner (1956) model using aggregate dividends (see also Fama, 1974).³

It can be noted that current time t earnings have not yet been observed when the current dividend level is decided upon and declared, since companies report their earnings after the end of the quarter whereas dividends are decided upon and announced prior to their payment each quarter. This problem can be eliminated in the Lintner model by lagging earnings by either a quarter or a year so that the time t dividend level choice is modelled only in relation to information that is observable at time t :

$$D_t = \theta_0 + \theta_1 E_{t-1} + \theta_2 D_{t-1} + \varepsilon_t. \quad (4)$$

Underlying the Lintner model is the idea that dividends would not be adjusted to changes in earnings that are only temporary, so Marsh and Merton (1987) introduce the concept of permanent earnings into dividend behaviour models and argue that dividends will only be determined in relation to permanent earnings. They further argue that, in an efficient market, the current share price index level is equal to the present value of all

³ The average adjusted R^2 they obtain of 45.7% is not directly comparable with the Lintner (1956) results because Grullon and Michaely (2002) estimate the Lintner model using changes in dividends (equation (1) in Lintner, 1956), not the dividend level regression model that Lintner (1956) tests (see equation (3) above), but the parameter estimates obtained are similar to the Lintner (1956) results.

future permanent earnings. The rate of change of share prices can then be shown to be equal to the rate of change of expected permanent earnings when it is assumed that the long-run discount rate is constant, thus implying a log linear relationship between dividend and price changes that leads to regression equation (11) in Marsh and Merton (1987):

$$\log\left(\frac{D_{t+1}}{D_t}\right) + \frac{D_t}{P_{t-1}} = \psi_0 + \psi_1 \log\left(\frac{P_t + D_t}{P_{t-1}}\right) + \psi_2 \log\left(\frac{D_t}{P_{t-1}}\right) + \varepsilon_{t+1} . \quad (5)$$

The Marsh and Merton (1987) model explains roughly half as much of the variation in aggregate dividends as does Lintner's equation (3) above (see Lintner, 1956; Marsh and Merton, 1987; as well as the Results section).

An alternative explanation of aggregate dividend behaviour is provided by Shiller's (1983) trend-autoregressive model.⁴ The model implies that deviations in aggregate dividends follow a trend-autoregressive process whereby half of the deviation in dividends from trend disappears within three years, with the underlying trend being explained by ongoing (not anticipated) earnings growth. Changes in current dividends are therefore determined by a time trend as well as past deviations from trend in Shiller's model, rather than anticipated earnings growth, but empirical tests indicate that the trend-autoregressive model provides (at most) half as much explanatory power as the Marsh and Merton (1987) model (see Marsh and Merton, 1987).

⁴ For an extensive theoretical and empirical review of dividend policy and dividend behaviour at the individual firm level, see Allen, Bernardo and Welch (2000).

2.3 *Spurious Regression*

Behavioural models of dividends share in common the use of lagged dividend terms to explain current dividend choices, a feature which implies that dividends are highly persistent. The recent literature on spurious regression points out that the use of persistent independent variables can lead to spurious regression results when the dependent variable is also at least partially persistent, since error terms in the regression equation inherit autocorrelation from the persistent dependent variable (Ferson, Sarkissian, and Simin, 2003a). This autocorrelation in the error term leads to biased standard error estimates and can therefore indicate a significant relationship when none exists, especially when data sets are mined for potentially significant explanatory variables (Ferson, Sarkissian, and Simin, 2003a; Foster, Smith and Whaley, 1997). Behavioural models of dividends are therefore likely candidates for spurious regression due to their strong persistence properties.

Return predictability regressions that use the dividend yield as a predictor variable will also be likely candidates for spurious regression because the numerator of the independent variable, the dividend level, is highly persistent and will consequently make the dividend yield variable highly persistent as well. It is therefore likely that persistence might be responsible for the apparent return predictability of the dividend yield, rather than any property of dividends related to risk, return or mispricing, and in this situation other similarly persistent variables might also possess apparent return predictability, as will be illustrated below.

3. DIVIDEND YIELD RETURN PREDICTABILITY

A starting point for examining whether dividend yield return predictability is a spurious result is to reproduce standard dividend yield return predictability results and compare them to the return predictability results that can be obtained using other persistent explanatory variables. The total return and capital return stock indices used in the study are the annual CRSP equally-weighted and value-weighted New York Stock Exchange Indices and the annual Standard and Poor Composite Index for the time period 1927 to 1996.⁵ This facilitates comparison of the results with those obtained in recent studies (see Table 1), and the data sets are also easily accessible. The CRSP indices are obtained from John Campbell's web page, and the Standard and Poor Composite Index data set is obtained from Global Financial Data (and is also available at Robert Shiller's web page). All series are converted from nominal to real values by dividing by the Consumer Price Index (CPI) obtained from Robert Shiller's web page. The dividend yield on the index at time t equals the level of real dividends during the twelve months preceding time t (D_t) divided by the real stock index level at time t (P_t). Panel A of Table 2 reproduces standard dividend yield return predictability results by regressing annual time $t+1$ rates of return against a constant as well as the annual time t dividend yield (see, e.g., Fama and French, 1988):

$$r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1 \left[\frac{D_t}{P_t} \right] + \varepsilon_{t+1}. \quad (1)$$

⁵ Numerous dividend yield return predictability studies also use monthly or quarterly returns with overlapping dividend yield observations. The use of overlapping observations introduces excess dividend yield autocorrelation. This well-known overlapping observation problem is avoided with the use of annual data.

The Fama and French (1988) results and the results of other studies are fairly closely replicated in Panel A of Table 2, with the regression adjusted R^2 ranging from 1.4% for the S&P Composite Index to 3.3% for the value-weighted NYSE Index, and 1.5% for the equally-weighted NYSE Index.⁶

[Table 2 about here]

The return predictability results documented in Panel A of Table 2 are often explained in terms of a theoretical relationship between current dividend yields and subsequent returns that is derived from the dividend discount model (see, e.g., Campbell and Shiller, 1988a and 1988b; Fama and French, 1988). It can be noted that this theoretical justification for return predictability relies upon a high current dividend level relative to the share index level predicting higher future returns, so substituting a constant level of dividends (c) for the dividend term (D_t) in the numerator of the dividend yield variable (D_t/P_t) in regression equation (1) should reduce or eliminate return predictability in the following altered regression equation:

$$r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_2 \left[\frac{c}{P_t} \right] + \varepsilon_{t+1}, \quad (2)$$

where c is the unconditional average of the dividend level during the sample period. The results for regression equation (2) are outlined in Panel B of Table 2. Surprisingly, the adjusted R^2 actually rises to above 3.4% for the S&P Composite Index and 6.8% for the Equally-Weighted NYSE Index. The results for this dividend yield from a constant dividend variable (c/P_t) are therefore so strong that they provide a first indication that something “funny” is going on in previous dividend yield return predictability studies

⁶ Regressions are estimated using Ordinary Least Squares and t-statistics are estimated using Newey-West (1987) standard errors.

since, interestingly, the dividend yield from a constant dividend variable (c/P_t) has higher explanatory power than does the standard dividend yield variable (D_t/P_t). This indicates that dividends are unlikely to be providing the explanatory power in return predictability regressions.

To further emphasize the point, and to check for robustness against a reverse time trend that might be present in the dividend yield from a constant dividend variable, a completely smooth “pseudo-dividend” series (M_t) is created that grows at a constant rate through time. The series (M_t) is referred to as a “pseudo-dividend” series because it grows at the same overall rate as the actual dividend series during the sample time period 1927 to 1996, yet it is a completely non-stochastic series. A “pseudo-dividend” yield equal to the real pseudo-dividend (M_t) divided by the real stock index series (P_t) is then used as the independent variable in a return predictability regression (see Panel C of Table 2). The pseudo-dividend yield ratio has at least as high a level of power for explaining returns as the actual dividend yield, yet the numerator of the ratio is non-stochastic and therefore cannot account for the correlation between the ratio and subsequent returns. This further indicates that it is very unlikely that the explanatory power of the dividend yield is due to correlation between dividends and subsequent price levels.

The Table 2 results instead suggest that the observed empirical relationship between subsequent returns and the dividend yield might be spurious. To investigate this possibility, the persistence properties of the dividend, dividend yield, price index, dividend yield from a constant dividend, and return series are investigated in Table 3. Table 3 confirms that the real dividend series are highly persistent, as are the dividend

yield series. The dividend yield from a constant dividend series are even more persistent than the dividend yield series, thus indicating that persistence might have an important influence on return predictability, since predictability increases when more persistent variables are used as predictor variables in Table 2. Table 3 indicates that returns are sufficiently autocorrelated to suggest that regressing subsequent rates of return against highly persistent dividend yield explanatory variables could lead to spurious regression.⁷

[Table 3 about here]

To follow up this possibility, a simulation procedure is utilized that provides the cut-off R^2 that would be obtained from a regression for which the dependent and independent variables are uncorrelated but have the same autocorrelation properties as the actual data, thus taking account of the potential for spurious regression (see also Foster, Smith and Whaley, 1997; Ferson, Sarkissian and Simin, 2003a). To obtain the cut-off R^2 using simulation, the moments and the serial correlation properties of the regression variables are first estimated for each data series, as described in the Appendix. Uncorrelated dependent and independent variables with the same serial correlation properties and sample moments are then simulated for a time period equal to the sample length (1927 to 1996), and a regression is run on these simulated series. The process is repeated 1,000 times, and the adjusted R^2 s are recorded for each regression and ranked from lowest to highest. The 95th percentile adjusted R^2 is then reported as the 5% cut-off R^2 and is compared to the actual adjusted R^2 obtained using the original data to assess the overall significance of the estimated regression relationship (see Foster, Smith and Whaley, 1997; Ferson, Sarkissian and Simin, 2003a).

⁷ Interestingly, Table 3 indicates the contemporaneous correlation between real dividend yields and real returns is actually negative.

The fifth column of Table 2 reports the cut-off R^2 that is obtained using the simulation procedure so that it can be compared with the adjusted R^2 reported in the fourth column of Table 2. For instance, the cut-off R^2 reported for the equally-weighted NYSE Index in the fifth column of Panel A of Table 2 indicates that an R^2 of 4.04% would be expected to be obtained by regressing subsequent returns against uncorrelated independent variables that are equally as persistent as the dividend yield variable (D_t/P_t), whereas an adjusted R^2 of only 1.54% is actually obtained. The adjusted R^2 of 1.54% does not exceed the critical cut-off R^2 level of 4.04%, thus implying that the equally-weighted dividend yield return predictability regression adjusted R^2 is insignificant (see also Foster, Smith and Whaley, 1997; Ferson, Sarkissian and Simin, 2003a). The cut-off R^2 levels reported in Table 2 therefore indicate that the dividend yield return predictability regression results are due to spurious regression since the reported adjusted R^2 never exceeds the cut-off R^2 (see the fifth column of Panel A of Table 2).

The simulation procedure treats the variables in regression equation (1),

$$r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1 \left[\frac{D_t}{P_t} \right] + \varepsilon_{t+1}, \quad (1)$$

as if they are independent. This independence assumption is not appropriate since returns on the left hand side of equation (1) and the dividend yield on the right hand side both come from the same underlying variables (the share index level P and the dividend level D). The simulation procedure is therefore modified to recognize the dependency of both the return and dividend yield variables on the share index and dividend levels (see the Appendix for details of the modified simulation procedure). Rather than simulating the

return series r_{t+1} as the dependent variable using the properties of the return series and simulating an uncorrelated dividend yield series (D_t/P_t) using the properties of the dividend yield series to calculate the cut-off R^2 , uncorrelated dividend (D_t) and share index (P_t) series are instead simulated using the estimated properties of these series. Dividend yield (D_t/P_t) and return $(r_{t+1} \equiv (P_{t+1} + D_{t+1} - P_t)/P_t)$ observations are then calculated using the simulated dividend and share index values, a regression is run using the constructed simulated variables, and the modified cut-off R^2 is then reported using the 95th percentile adjusted R^2 obtained from the modified simulation regression procedure.

The modified simulation procedure leads to a considerably higher simulated cut-off R^2 (see the modified cut-off R^2 reported in the final column of Table 2). The modified cut-off R^2 for the Equally-Weighted NYSE Index dividend yield regression rises to 64.09% from 4.04%, for instance, and the modified cut-off R^2 for the two value-weighted indices rise to above 75% from just over 4% (see the final column of Panel A of Table 2). The modified cut-off R^2 s reported in Table 2 therefore make it very clear-cut that dividend yield return predictability is a spurious result since the actual adjusted R^2 s obtained for the dividend yield regressions which range from 1.4% to 3.32% are an order of magnitude smaller than the modified cut-off R^2 levels. The increase in the modified cut-off R^2 s that is apparent in Table 2 is caused primarily by the presence of the highly persistent price index common denominator on both sides of regression equation (1).

The modified simulation procedure highlights the influence of the dividend yield and rates of return variables sharing a common denominator (the share price index level P_t), an effect that can create a spurious correlation even when all variables that

make up the numerators and the common denominator in a regression are independent (Pearson, 1897; Kronmal, 1993). This spurious correlation problem arises when all variables except for the constant in a “true” regression equation are divided by a common variable, often in an (incorrect) attempt by the researcher to control for a “common confounding influence”. Kronmal (1993, page 381), summarizing Friedlander (1980), outlines how this leads to a biased least squares estimate of the independent coefficient when the “true” regression constant is non-zero (see also Pearson, 1897; Tanner, 1949; Neyman, 1952; and Friedlander, 1980). More importantly, it can also lead to a significant estimate of the overall relationship between the dependent and independent regression variables even when all of the component variables making up the numerators and the denominator of the dependent and independent regression variables are uncorrelated. Intuitively, the common denominator on both sides of the regression model can introduce a correlation between the independent and the dependent variable if the regression constant is not divided by the same “control” variable.

The Table 2 spurious regression return predictability results can also be related to the persistent regressor coefficient bias literature (see, e.g., Stambaugh, 1999; Lewellen, 2004). Table 2 focuses on the significance of the overall relationship between dividend yields and subsequent returns to determine whether the observed relationship is spurious, and does not focus on whether the dividend yield coefficient t-statistics are biased, since the paper does not focus on the coefficients *per se*.⁸ Still, it can be useful to examine how the results relate to recommendations made in the return predictability literature to overcome the spurious regression and coefficient bias problems.

⁸ Ferson, Sarkissian and Simin (2003b) point out that the return predictability coefficient bias literature does not address the interaction between spurious regression and data mining.

To make this comparison, it can first be noted that if high dividend yields really do predict higher subsequent returns, then an increase in the dividend yield should also forecast an increase in subsequent returns. Ferson, Sarkissian and Simin (2003b) suggest that testing the relationship between returns and changes in predictor variables away from their trailing moving average can overcome the spurious regression problem, since detrending predictor variables will create a less persistent independent variable. Subsequent returns can therefore be regressed against changes in the dividend yield from its trailing moving average, with a twelve month moving average lag length being utilized (as recommended by Ferson, Sarkissian and Simin, 2003b). Lewellen (2004) implements a somewhat related regression model to test for return predictability, since he uses the natural logarithm of the dividend yield as the regression independent variable; the natural logarithm of the dividend yield can be interpreted as a change in the dividend yield away from a dividend yield of one. Campbell and Yogo (2003) also recommend detrending dividend yields when testing for return predictability.⁹ Subsequent real returns are therefore regressed against the detrended dividend yield using the regression model

$$r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1 X_t + \varepsilon_{t+1}, \quad (6)$$

where r_{t+1} denotes the annual real index return at time $t+1$, X_t denotes the stochastic detrended dividend yield calculated as

$$X_t = \frac{D_t}{P_t} - \frac{1}{\tau} \sum_{j=1, \dots, \tau} \frac{D_{t-j}}{P_{t-j}}, \quad (7)$$

⁹ Campbell and Yogo (2003) recommend subtracting from the dividend yield independent variable the innovations in the yield that are correlated with returns to obtain a less noisy independent variable, thus eliminating some of the noise and increasing the power of the test.

D_t is the level of real annual dividends during the twelve months preceding time t , and P_t is the real stock index level at time t . While different numbers of lags could be used in the detrending, a 12-month lag is used, as recommended by Ferson, Sarkissian and Simin (2003b).

Results for return predictability regression model (6) are reported in Table 4. Interestingly, the results imply that an increase in the dividend yield actually foreshadows lower, not higher, returns (although the overall relationship is clearly insignificant). Regression model (6) has not been tested before, but the Table 4 results can be compared to Lewellen's (2004) finding that the return predictability regression coefficient of the natural logarithm of the dividend yield is actually negative within sub-periods for the monthly CRSP equally-weighted and value-weighted Indices when estimated using the Stambaugh (1999) bias-adjustment.¹⁰ The Table 4 results again reinforce the Table 2 findings that the observed relationship between dividend yields and subsequent returns is spurious, and they also make it extremely unlikely that predictor coefficient bias adjustments would lead to a significant overall relationship, since bias adjustments do not normally reverse the sign of the coefficients. The Table 4 results, combined with the Table 2 results, further suggest that a spurious correlation effect influences the estimated relationship between dividend yields and subsequent returns, since the reversal of the dividend yield regression coefficient sign between Table 2 and Table 4 is consistent with

¹⁰ Lewellen (2004) reports very low adjusted R^2 s (less than .014) when regressing subsequent monthly returns against the natural logarithm of the dividend yield, and tends to find a significant dividend yield coefficient estimate only when imposing the assumption of unitary dividend yield serial correlation, an assumption that is clearly violated in Table 3 for annual data.

spurious correlation effects outlined in Kronmal (1993).¹¹ The Table 4 results also cast severe doubt on the temporary mispricing explanation of return predictability, since the Table 4 results imply that an increase in mispricing would predict higher, not lower, subsequent returns.

[Table 4 about here]

The results of Tables 2 and 4, taken together, imply that the spurious effect of regressing returns against a highly persistent explanatory variable such as the dividend yield is strongly reinforced by a spurious correlation effect. To further explore the origins of this spurious return predictability effect, models of dividend behaviour are also tested for spurious regression.

4. DIVIDEND BEHAVIOUR RESULTS

The persistence properties of dividend yields that contribute to spurious return predictability imply that dividend behaviour models are also likely to be subject to spurious regression, since the dependent and independent variables in the Linter (1956) and Marsh and Merton (1987) dividend behaviour models are all very persistent (see also Ferson, Sarkissian and Simin, 2003a). These dividend behaviour models are examined in Panels A and B of Table 5. The Table 5 results indicate that spurious regression plays an extremely important role in these standard dividend behaviour regression models which regress highly persistent dividends against lagged dividends and other terms. Panel A of

¹¹ Kronmal (1993) reports how storks can be responsible for either a decrease or an increase in births, depending on how the dependent variable births is regressed against the independent variable storks in alternative incorrectly specified regression models that are subject to spurious correlation!

Table 5 reveals that spurious regression is very apparent in the results for the Marsh and Merton (1987) model

$$\log\left(\frac{D_{t+1}}{D_t}\right) + \frac{D_t}{P_{t-1}} = \psi_0 + \psi_1 \log\left(\frac{P_t + D_t}{P_{t-1}}\right) + \psi_2 \log\left(\frac{D_t}{P_{t-1}}\right) + \varepsilon_{t+1} . \quad (5)$$

The results indicate that current log returns appear to be very important in explaining subsequent dividend changes, but the modified simulation cut-off R^2 s are much higher than the adjusted R^2 s in Panel A of Table 5, thus implying that the results are due to spurious regression.

[Table 5 about here]

The Lintner (1956) model of the time t dividend level choice,

$$D_t = \theta_0 + \theta_1 E_{t-1} + \theta_2 D_{t-1} + \varepsilon_t , \quad (4)$$

is tested using the S&P composite index only due to earnings data availability (see also Arnott and Asness, 2003).¹² Panel B of Table 4 reveals that the estimated independent variable coefficients for regression model (4) are very similar to those found in Lintner (1956), even though the data set is extended by four decades. The results imply that lagged earnings and (especially) lagged dividends explain almost all of the variation in the current dividend level. The R^2 exceeds 90%, as in Lintner (1956), and slightly exceeds the cut-off R^2 , thus implying that the original Lintner model appears to somewhat survive the spurious regression problem. The very high cut-off R^2 reported for the Lintner model in Panel B of Table 5 suggests that the spurious regression problem has an influence on the Lintner model regression, even if it is not solely responsible for the

¹² The Standard and Poor's Composite index earnings data set is obtained from Robert Shiller's web site.

results, thus indicating the potential need for a reformulation of the original Lintner model.

A step towards a reformulation of the Lintner regression model is already provided in Lintner (1956), where the Lintner dividend behaviour theoretical model is originally presented in terms of changes in dividends, not the dividend level as in regression model (4) above (see equation (1) in Lintner, 1956, and see also Grullon and Michaely, 2002; Fama, 1974). Restating the Lintner dividend change model (Lintner (1956) equation (1)) in terms of information that is observable at time t leads to the following regression model:

$$D_t - D_{t-1} = \lambda_0 + \lambda_1 E_{t-1} + \lambda_2 D_{t-1} + \varepsilon_t \quad (8)$$

Results for the Lintner (1956) dividend change regression model (8) are presented in Panel C of Table 5 and are consistent with the results for the Lintner (1956) dividend level regression model (4) presented in Panel B. Having changes in dividends rather than the dividend level as the dependent variable in the Lintner regression model greatly reduces the adjusted R^2 (from 91.8% down to 8.6%) as well as the cut-off R^2 (from 81.3% to 8%) in Panel C, so the adjusted R^2 once again slightly exceeds the cut-off R^2 . The following section takes further steps to reformulate the Lintner dividend behaviour model entirely in terms of first differences on both sides of the regression equation, and also incorporates the “permanent earnings” innovation of Marsh and Merton (1987) into the reformulated dividend behaviour model.

5. AN ALTERNATIVE DIVIDEND BEHAVIOUR MODEL SPECIFICATION

A source of persistence in the Lintner (1956) dividend level model (equation (4)) that is likely to have a very important effect on the model's time series regression properties is a common time trend in both the dependent and independent regression variables:

$$D_t = \theta_0 + \theta_1 E_{t-1} + \theta_2 D_{t-1} + \varepsilon_t, \quad (4)$$

A standard solution to this problem is a reformulation entirely in terms of first differences (see also equation (1) in Lintner, 1956). To derive the Lintner model in terms of first differences, first note that equation (4) also implies:

$$D_{t-1} = \theta_0 + \theta_1 E_{t-2} + \theta_2 D_{t-2} + \varepsilon_{t-1}. \quad (9)$$

Equation (4) minus equation (9) leads to a reformulated first-difference Lintner model

$$D_t - D_{t-1} = \theta_1 (E_{t-1} - E_{t-2}) + \theta_2 (D_{t-1} - D_{t-2}) + (\varepsilon_t - \varepsilon_{t-1}). \quad (10)$$

A second potential improvement to the Lintner model follows from the Lintner (1956) argument that dividends would not be adjusted to changes in earnings that are only temporary. This led Marsh and Merton (1987) to introduce the concept of permanent earnings and to argue that dividends are determined in relation to permanent earnings. They further argue that, in an efficient market, the current share price index level is equal to the present value of all future permanent earnings. By assuming that the long-run discount rate is relatively constant, then "permanent earnings" divided by the long-run discount rate is proportionate to the current share index level, so by adjusting dividends through time in reaction to permanent earnings then managers are adjusting dividends to the share index level. The Marsh and Merton permanent earnings innovation can

therefore be directly introduced into the Lintner model by substituting the price index level P in for earnings E in the reformulated Lintner first-difference regression model (10):

$$D_t - D_{t-1} = \theta_1(P_{t-1} - P_{t-2}) + \theta_2(D_{t-1} - D_{t-2}) + (\varepsilon_t - \varepsilon_{t-1}). \quad (11)$$

Results for the reformulated Lintner first-difference regression model (11) are reported in Table 6.^{13, 14} A notable result from the table is a more reasonable adjusted R^2 for the Lintner first-difference model (no longer 90%) which now greatly exceeds the cut-off R^2 , thus implying a lessening of the influence of spurious regression in the re-specified model. The results imply that changes in the aggregate level of dividends are explained by lagged share price index innovations and, in the case of the Equal Weighted NYSE index, lagged dividend innovations. The results are therefore consistent with the predictions of Marsh and Merton (1987). The right side of Table 6 also reproduces the results using detrended real data to ensure that the results are not sensitive to any remaining time trends in the first difference regression variables (the detrending follows Marsh and Merton, 1986). The results are very similar, with innovations in detrended prices explaining subsequent changes in detrended dividends, so the Lintner dividend first-difference regression model (11) appears to provide a correctly specified dividend behaviour model.

[Table 6 about here]

¹³ Note that the reformulated Lintner first difference regression model (11) nests the constant growth dividend model $P = kE/(r-g)$, where k is the (constant) dividend payout ratio from earnings, r is the constant discount rate, and g is the constant growth rate. It can also be noted that the results in Table 6 are insensitive to the inclusion of an intercept term in regression model (11); results not reported.

¹⁴ The error term in equation (9), $\varepsilon_t - \varepsilon_{t-1}$, is accommodated using a Heteroskedasticity and Autocorrelation Consistent Covariance estimation procedure within Generalized Method of Moments.

6. CONCLUSION

It is now known that spurious regression is a very serious problem when highly persistent variables are used in a time series regression model to predict or explain dependent variables that are at least partially persistent. The spurious regression problem is compounded when the dependent and independent variables in a regression equation are ratios constructed from common component variables. Both of these problems are present in dividend yield return predictability and dividend behaviour regression models, thus explaining why standard dividend return predictability and dividend behaviour regression results are spurious. Return predictability of the dividend yield is shown to be due to the persistence properties of the dividend yield as well as a spurious correlation that results from regression variables being constructed from common component variables, rather than (as widely argued) any property of the dividend level that is related to risk, return or mispricing. The paper's results therefore strongly reinforce the Kronmal (1993) message that extreme care should be taken when using ratios in regression analysis, especially in time series regressions when highly persistent ratios are used as predictor or explanatory variables.

Standard dividend behaviour models are also strongly affected by the spurious regression problem since the dependent and independent variables in these models are all very persistent. A reformulation of the Lintner (1956) dividend model in terms of first differences provides a dividend behaviour explanatory model that is far less subject to spurious regression, and also directly incorporates the Marsh and Merton (1987) permanent earnings explanation of dividend behaviour. Results for the reformulated Lintner first difference dividend model imply that changes in the aggregate level of

dividends are explained by lagged share price innovations as well as lagged dividend innovations, as predicted by Marsh and Merton (1987).

REFERENCES

Allen, F., Bernardo, A.E., and I. Welch, 2000, A Theory of Dividends Based on Tax Clienteles, *Journal of Finance* 55, 2499-2536.

Arnott, Robert D., and Clifford S. Asness, 2003, Surprise! Higher Dividends = Higher Earnings Growth, *Financial Analysts Journal* 59, 70-87.

Black, Fischer, and Myron Scholes, 1974, The Effects Of Dividend Yield And Dividend Policy On Common Stock Prices And Returns, *Journal of Financial Economics* 1, 1-22.

Bossaerts, Peter, and Pierre Hillion, 1999, Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?, *Review of Financial Studies* 12, 405-428.

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

Campbell, John Y., and Robert J. Shiller, 1988b, Stock Prices, Earnings and Expected Dividends, *Journal of Finance* 43, 661-676.

Campbell, John Y., Lo, Andrew W., and A. Craig MacKinlay, 1997, *The Econometrics and Financial Markets*, Princeton University Press, New Jersey.

Campbell, John Y., and Motohiro Yogo, 2003, Efficient Tests of Stock Return Predictability, Working Paper.

Cremers, K.J.M., 2002, Stock Return Predictability: A Bayesian Model Selection Perspective, *Review of Financial Studies* 15, 1223-1249.

Donaldson, R. Glen and Mark Kamstra, 1996, A New Dividend Forecasting Procedure That Rejects Bubbles In Asset Prices: The Case Of 1929's Stock Crash, *Review of Financial Studies* 9, 333-383.

Fama, Eugene F., 1974, The Empirical Relationships between the Dividends and Investment Decisions of Firms, *American Economic Review* 64, 304-318.

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

Ferson, Wayne E., and Campbell R. Harvey, 1991, The Valuation of Economic Risk Premium, *Journal of Political Economy* 99, 385-415.

- Ferson, Wayne E., Sarkissian, Sergei, and Timothy T. Simin, 2003a, Spurious Regressions in Financial Economics?, *Journal of Finance* 58, 1393-1413.
- Ferson, Wayne E., Sarkissian, Sergei, and Timothy T. Simin, 2003b, Is Stock Return Predictability Spurious?, *Journal of Investment Management* 1, 1-10.
- Friedlander, L.J., 1980, *A Study of the Correlation between Ratio Variables*, PhD Dissertation, University of Washington.
- Foster, F. Douglas, Tom Smith and Robert E. Whaley, 1997, Assessing The Goodness-Of-Fit Of Asset Pricing Models: The Distribution Of The Maximal R-Squared, *Journal of Finance* 52, 591-607.
- Goetzmann, William N., and Philippe Jorion, 1993, Testing The Predictive Power Of Dividend Yields, *Journal of Finance* 48, 663-679.
- Goetzmann, William N., and Philippe Jorion, 1995, A Longer Look At Dividend Yields, *Journal of Business* 68, 483-508.
- Goyal, Amit, and Ivo Welch, 2003, Predicting Equity Premium with Dividend Ratios”, *Management Science* 49, 639-654.
- Grullon, G. and R. Michaely, 2002, Dividends, Share Repurchases, and the Substitution Hypothesis, *Journal of Finance* 57, 1649-1684.
- Harvey, Campbell R., 1989, Time-Varying Conditional Covariances in Tests of Asset Pricing Models, *Journal of Financial Economics* 24, 289-317.
- Hodrick, Robert J., 1992, Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, *Review of Financial Studies* 5, 357-386.
- 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.
- Kronmal, Richard A., 1993, Spurious Correlation and the Fallacy of the Ratio Standard Revisted, *Journal of Royal Statistical Society* 156, Series A, 379-392.
- Lewellen, Jonathon, 2004, Predicting Returns With Financial Ratios, *Journal of Financial Economics*, forthcoming.
- Lintner, John, 1956, Distribution of Incomes of Corporations among Dividends, Retained Earnings, and Taxes, *American Economic Review* 46, 97-113.
- Litzenberger, Robert H., and Krishna Ramaswamy, 1979, The Effect of Personal Taxes and Dividends on Capital Asset Prices: Theory and Empirical Evidence, *Journal of Financial Economics* 7, 163-195.

Marsh, Terry A., and Robert C. Merton, 1986, Dividend Variability and Variance Bound Tests for the Rationality of Stock Market Prices, *American Economic Review* 76, 483-498.

Marsh, Terry A., and Robert C. Merton, 1987, Dividend Behavior for the Aggregate Stock Market, *Journal of Business* 60, 1-40.

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

Neyman, Jerzy, 1952, *Lectures and Conferences on Mathematical Statistics and Probability*, 2nd edition, pp. 143-154, US Department of Agriculture, Washington D.C.

Pearson, Karl, 1897, Mathematical Contributions to the Theory of Evolution - On a Form of Spurious Correlation which may Arise when Indices are Used in the Measurement of Organs, *Proceedings of the Royal Society* 60, 489-498.

Pesaran, M. Hashem, and Allan Timmermann, 1995, Predictability Of Stock Returns: Robustness And Economic Significance, *Journal of Finance* 50, 1201-1228.

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

Poterba, James M., and Lawrence H. Summers, 1988, Mean Reversion In Stock Prices: Evidence And Implications, *Journal of Financial Economics* 22, 27-60.

Shiller, Robert J., 1983, Do Stock Prices Move Too Much To Be Justified By Subsequent Changes In Dividends?: Reply, *American Economic Review* 73, 236-237.

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

Tanner, John. M., 1949, Fallacy of Per-weight and Per-surface Area Standard, and their relation to Spurious Correlation, *Journal of Applied Physiology* 2, 1-15.

Whitelaw, Robert F., 1994, Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns, *Journal of Finance* 49, 515-541.

APPENDIX

A1 *Cut-off R² Simulation Procedure*

The cut-off R² simulation procedure provides the cut-off R² that is obtained by regressing simulated dependent and independent variables that are uncorrelated but have the same autocorrelation properties as the actual dependent and independent variable data series. The simulated independent variable series X_t is generated as

$$X_t = \alpha_x + \rho_x X_{t-1} + e_t \quad \text{for } t = 2, 3, \dots, n, \quad (\text{A1})$$

where α_x is the intercept and ρ_x is the first order autocorrelation coefficient. The unconditional mean of the independent variable series (μ_x) is

$$\mu_x = \frac{\alpha_x}{1 - \rho_x}$$

and the variance σ_x^2 is

$$\sigma_x^2 = \frac{\sigma_e^2}{1 - \rho_x^2}.$$

The parameters that calibrate the simulation, μ_x , σ_x^2 and ρ_x , are taken from the actual data. The simulation is started at the unconditional mean μ_x and the error term is generated from a normal with mean zero and variance $\sigma_x^2 (1 - \rho_x^2)$.

Similarly, the uncorrelated dependent variable series is generated as

$$r_t = \alpha_r + \rho_r r_{t-1} + e_t \quad \text{for } t = 2, 3, \dots, n \quad (\text{A2})$$

with

$$\mu_r = \frac{\alpha_r}{1 - \rho_r}$$

and

$$\sigma_r^2 = \frac{\sigma_e^2}{1 - \rho_r^2},$$

where the parameters μ_r , σ_r^2 and ρ_r are taken from the actual data. The dependent variable simulation is started at the unconditional mean μ_r and the error term for process A(2) is generated independently of process A(1) from a normal with mean zero and variance $\sigma_r^2(1 - \rho_r^2)$.

The simulated dependent and independent variable series are stored, and an ordinary least squares regression is run. The process is repeated 1,000 times. The adjusted R^2 s are recorded for each regression and ranked from lowest to highest. The 95th percentile adjusted R^2 is then reported as the 5% cut-off R^2 .

A2 *Modified Cut-off R^2 Simulation Procedure*

To obtain the modified cut-off R^2 , the cut-off R^2 simulation procedure is modified to recognize the dependency of both returns (r_{t+1}) and the dividend yield variable ($X_t \equiv D_t / P_t$) on the same underlying component variables (the share index level P and the dividend level D) in regression model

$$r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1 \left[\frac{D_t}{P_t} \right] + \varepsilon_{t+1}. \quad (1)$$

The modified cut-off R^2 simulation procedure simulates the dividend series (D_t) and price index series (P_t) independently prior to constructing the dependent and independent variable return and dividend yield series. The simulated dividend series D_t is generated as

$$D_t = \alpha_D + \rho_D D_{t-1} + e_t \quad \text{for } t = 2, 3, \dots, n, \quad (\text{A3})$$

with

$$\mu_D = \frac{\alpha_D}{1 - \rho_D}$$

and

$$\sigma_D^2 = \frac{\sigma_e^2}{1 - \rho_D^2},$$

where the parameters μ_D , σ_D^2 and ρ_D are taken from the actual dividend data. The simulation is started at the unconditional mean of the dividend series μ_D and the error term is generated from a normal with mean zero and variance $\sigma_D^2(1 - \rho_D^2)$.

The uncorrelated price index series is generated as

$$P_t = \alpha_P + \rho_P P_{t-1} + e_t \quad \text{for } t = 2, 3, \dots, n \quad (\text{A4})$$

with

$$\mu_P = \frac{\alpha_P}{1 - \rho_P}$$

and

$$\sigma_P^2 = \frac{\sigma_e^2}{1 - \rho_P^2},$$

where the parameters μ_P , σ_P^2 and ρ_P are taken from the actual price index series. The simulation is started at the unconditional mean μ_P and the error term for process A(4) is generated independently of process A(3) from a normal with mean zero and variance $\sigma_P^2(1 - \rho_P^2)$.

A dividend yield series ($X_t \equiv D_t / P_t$) is subsequently created from the simulated dividend series D_t and simulated price index series P_t :

$$X_t = \left[\frac{D_t}{P_t} \right] \text{ for } t = 1, \dots, n.$$

A return series ($r_{t+1} \equiv (P_{t+1} + D_{t+1} - P_t) / P_t$) is created from the same simulated dividend series D_t and simulated price index series P_t :

$$r_{t+1} = \frac{P_{t+1} + D_{t+1} - P_t}{P_t} \text{ for } t = 1, 2, \dots, n-1.$$

The dividend yield series (X_t) and the return series (r_{t+1}) are then stored. An ordinary least squares regression is run with the stored dividend yield series (X_t) as the independent variable, and the stored return series (r_{t+1}) as the dependent variable. The process is repeated 1,000 times, with the adjusted R^2 s being recorded for each regression and ranked from lowest to highest. The 95th percentile adjusted R^2 is then reported as the 5% modified cut-off R^2 .

Table 1

Model specifications used in 10 previous empirical studies of the dividend yield

This table presents model specifications used in a number of recent empirical studies on the dividend yield. Under I/P/S, I refers to a well know index of stocks (e.g. NYSE or S&P index), P refers to a stock portfolio based on financial characteristics of the data (e.g. size or industry), and S refers to individual stock data. For M/Q/A, M refers to monthly data, Q refers to quarterly data and A refers to annual data. For variables used in the studies, 1 denotes lagged return, 2 denotes credit spread between the yields of investment grade and below investment grade bonds, 3 denotes yield on a short-term T-Bill, 4 denotes term spread between the yields on long-term government bonds and the short term T-bill, 5 denotes yield on a long-term bond, 6 denotes yield spread between the yields on commercial paper and the short-term T-bill, 7 denotes January Dummy, 8 denotes book-to-market ratio, and 9 denotes systematic risk. '+' ('-') is used to signify a positive (negative) but insignificant relation, and '++' ('--') a significant positive (negative) relation. Note that some studies only report R² while others report adjusted R².

Study	Litzenberger & Ramaswamy (1979)	Fama & French (1988)	Harvey (1989)	Ferson & Harvey (1991)	Nelson & Kim (1993)	Nelson & Kim (1993)	Whitelaw (1994)	Kothari & Shanken (1997)	Pontiff & Schall (1998)	Bossaerts & Hillion (1999)
I/P/S	S	P	I	I	I	I	I	I	I	I
Sample period	36-77	27-86	41-87	64-86	72-86	72-86	53-89	26-91	26-94	56-95
Freq. of Data (M/Q/A)	M	M/Q/A	M	M	A	A	M	M	M	M
Dep Var.	excess return	nominal and real returns	excess return	Excess return	Real return	excess return	Excess return	Index return	Index return	Excess return
Dividend Yield	++	++	++	++	++	+	++	++	-	++
1			+	+						
2			++	+			++		++	
3				--			--		-	--
4			++	+					+	
5										++
6							-			
7			+	++						+
8									+	
9	+									
R ² or adj. R ²	NA	0.00-0.22	0.075	0.14	0.039	0.016	0.082	0.075	0.02	0.074

Table 2
Regressions of Annual Real Index Returns on Three Predictors

This table presents OLS regressions of annual real index return on three predictors:

$$r_{t+1} = \beta_0 + \beta_1 X_t + \varepsilon_{t+1}$$

where r_{t+1} denotes annual real index return at time $t+1$ on the annual CRSP equally-weighted and value-weighted New York Stock Exchange indices and the annual Standard and Poor Composite Index for the time period 1927 to 1996. The CRSP indices are obtained from John Campbell's web page, and the Standard and Poor Composite Index data set is obtained from Global Financial Data (and is also available at Robert Shiller's web page). X_t variously denotes the time t real dividend yield (D_t / P_t), the dividend yield from a constant dividend (c / P_t), or the real pseudo dividend yield (M_t / P_t). D_t is the level of real annual dividends during the twelve months preceding time t , P_t is the real stock index level at time t , c is the unconditional average of the dividend level over the sample period, and the pseudo-dividend series (M_t) is a non-stochastic series growing at a constant growth rate through time and is the sum of real quarterly pseudo-dividends in the past year. Regressions are estimated by OLS and figures in parentheses are t-statistics using Newey-West (1987) standard errors. *** denotes significance at the 1% level, ** denotes significance at the 5% level and * denotes significance at the 10% level. The fifth column of the table reports the cut-off R^2 that is obtained using a simulation procedure where dependent and independent variables are uncorrelated but have the same autocorrelation properties as the actual data. The final column of the table shows the modified cut-off R^2 which recognizes the dependency of both the return and dividend yield variables on the share index and dividend levels.

Panel A: $r_{t+1} = \beta_0 + \beta_1(D_t / P_t) + \varepsilon_{t+1}$					
Stock Index	Intercept β_0	Slope β_1	Adjusted R^2	Cut-off R^2	Modified Cut-off R^2
Equally-Weighted NYSE Index	-0.021 (-0.156)	3.849 (1.304)	1.54%	4.04%	64.09%
Value-Weighted NYSE Index	-0.059 (-0.875)	3.396 (2.041)**	3.32%	4.47%	79.30%
S&P Composite Index	-0.01 (-0.167)	2.288 (1.513)	1.40%	4.13%	76.18%
Panel B: $r_{t+1} = \beta_0 + \beta_1(c / P_t) + \varepsilon_{t+1}$					
Stock Index	Intercept β_0	Slope β_1	Adjusted R^2	Cut-off R^2	Modified Cut-off R^2
Equally-Weighted NYSE Index	0.073 (2.632)**	0.538 (1.618)	6.82%	4.88%	65.46%
Value-Weighted NYSE Index	0.003 (0.078)	1.675 (2.309)**	4.02%	4.38%	96.33%
S&P Composite Index	0.011 (0.296)	1.541 (2.096)**	3.43%	3.92%	94.34%
Panel C: $r_{t+1} = \beta_0 + \beta_1(M_t / P_t) + \varepsilon_{t+1}$					
Stock Index	Intercept β_0	Slope β_1	Adjusted R^2	Cut-off R^2	Modified Cut-off R^2
Equally-Weighted NYSE Index	-0.132 (-2.403)**	5.301 (4.076)***	20.14%	5.04%	62.34%
Value-Weighted NYSE Index	-0.007 (-0.122)	1.318 (1.831)*	3.60%	4.06%	89.86%
S&P Composite Index	0.011 (0.188)	1.194 (1.529)	2.14%	4.34%	89.69%

Table 3
Descriptive Statistics

Panel A presents summary statistics on the real annual dividend (D_t), real dividend yield (D_t/P_t), real index return (r_t), real price index (P_t) and the dividend yield from a constant dividend (c/P_t) on the annual CRSP equally-weighted and value-weighted New York Stock Exchange indices and the annual Standard and Poor Composite Index for the time period 1927 to 1996. D_t is the level of real annual dividends during the twelve months preceding time t , P_t is the real stock index level at time t , r_t denotes annual real index return at time t , and c is the unconditional average of the dividend level over the sample period. Panel B reports autocorrelations of the variables up to 6 lags, and correlations among the variables are presented in Panel C.

	Equally-Weighted NYSE Index					Value-Weighted NYSE Index					S&P Composite Index				
	D_t	D_t/P_t	r_t	P_t	c/P_t	D_t	D_t/P_t	r_t	P_t	c/P_t	D_t	D_t/P_t	r_t	P_t	c/P_t
Panel A: Descriptive Statistics															
Mean	0.888	0.04	0.135	24.997	0.109	0.189	0.043	0.089	4.983	0.051	0.217	0.044	0.091	5.741	0.051
Median	0.693	0.036	0.157	19.994	0.044	0.192	0.041	0.117	4.598	0.041	0.228	0.041	0.114	5.214	0.042
Maximum	2.506	0.072	1.387	88.451	0.885	0.297	0.089	0.571	12.852	0.133	0.32	0.097	0.528	15.942	0.122
Minimum	0.041	0.016	-0.474	1.004	0.010	0.078	0.022	-0.386	1.426	0.015	0.113	0.02	-0.356	1.778	0.014
Std. Dev.	0.683	0.013	0.299	21.595	0.156	0.059	0.013	0.204	2.579	0.028	0.061	0.015	0.199	3.08	0.029
Panel B: Autocorrelation															
1	0.94	0.578	0.048	0.897	0.854	0.94	0.674	-0.011	0.868	0.881	0.926	0.699	0.007	0.857	0.893
2	0.875	0.204	-0.18	0.818	0.656	0.87	0.397	-0.211	0.77	0.752	0.843	0.446	-0.227	0.75	0.784
3	0.814	0.088	-0.109	0.774	0.517	0.806	0.299	-0.062	0.736	0.669	0.788	0.365	0.033	0.72	0.716
4	0.77	0.12	-0.205	0.702	0.422	0.766	0.257	-0.081	0.676	0.614	0.748	0.282	-0.084	0.655	0.657
5	0.737	0.156	-0.004	0.628	0.354	0.731	0.264	0.039	0.585	0.603	0.709	0.247	-0.015	0.56	0.632
6	0.698	0.198	-0.065	0.573	0.304	0.688	0.258	0.005	0.496	0.571	0.655	0.256	0.016	0.48	0.593
Panel C: Correlations															
D_t	1					1					1				
D_t/P_t	-0.255	1				-0.517	1				-0.561	1			
r_t	-0.098	-0.508	1			-0.053	-0.429	1			-0.104	-0.41	1		
P_t	0.961	-0.414	-0.018	1		0.875	-0.778	0.135	1		0.872	-0.796	0.138	1	
c/P_t	-0.584	0.322	-0.102	-0.555	1	-0.853	0.826	-0.193	-0.873	1	-0.869	0.846	-0.182	-0.87	1

Table 4
Regressions of Annual Real Index Returns on Stochastic Detrended Lagged Dividend Yield Variables

This table presents OLS regressions of annual real index return on a stochastic detrended dividend yield variable, as recommended by Ferson *et al* (2003b):

$$r_{t+1} = \beta_0 + \beta_1 X_t + \varepsilon_{t+1}$$

where r_{t+1} denotes annual real index return at time $t+1$ on the annual CRSP equally-weighted and value-weighted New York Stock Exchange indices and the annual Standard and Poor Composite Index for the time period 1927 to 1996. The CRSP indices are obtained from John Campbell's web page, and the Standard and Poor Composite Index data set is obtained from Global Financial Data (and is also available at Robert Shiller's web page). X_t denotes the stochastic detrended dividend yield and is calculated as:

$$X_t = \frac{D_t}{P_t} - \frac{1}{\tau} \sum_{j=1, \dots, \tau} \frac{D_{t-j}}{P_{t-j}}$$

where D_t is the level of real annual dividends during the twelve months preceding time t and P_t is the real stock index level at time t . While different numbers of lags could be used in the detrending, a 12-month lag is used, as recommended by Ferson *et al* (2003b). Regressions are estimated by OLS and figures in parentheses are t-statistics using Newey-West (1987) standard errors. *** denotes significance at the 1% level, ** denotes significance at the 5% level and * denotes significance at the 10% level. The final column of the table reports the modified cut-off R^2 which recognizes the dependency of both the return and dividend yield variables on the share index and dividend levels.

Stock Index	Intercept β_0	Slope β_1	Adjusted R^2	Modified Cut-off R^2
Equally-Weighted NYSE Index	0.131 (4.358)***	-5.602 (-1.158)	3.92%	41.77%
Value-Weighted NYSE Index	0.084 (4.249)***	-2.259 (-0.937)	-0.10%	42.23%
S&P Composite Index	0.086 (4.408)***	-2.664 (-1.384)	0.87%	40.74%

Table 5
Results for the Marsh and Merton (1987) and Lintner (1956) Dividend Models

This table presents regression results for the Marsh and Merton (1987) and Lintner (1956) dividend behaviour models. The Marsh and Merton (1987) model is tested using the annual CRSP equally-weighted and value-weighted New York Stock Exchange indices and the annual Standard and Poor Composite Index for the time period 1927 to 1996, and the Lintner (1956) models are tested using the annual Standard and Poor Composite Index for the time period 1927 to 1996. Panel A of the table reports results for the Marsh and Merton (1987) model

$$\log(D_{t+1}/D_t) + D_t/P_{t-1} = \psi_0 + \psi_1[(P_t + D_t)/P_{t-1}] + \psi_2 \log(D_t/P_{t-1}) + \varepsilon_{t+1}$$

where D_t is the level of real annual dividends during the twelve months preceding time t and P_t is the real stock index level at time t . Panel B reports results for Lintner (1956) model 1

$$D_t = \theta_0 + \theta_1 E_{t-1} + \theta_2 D_{t-1} + \varepsilon_t,$$

where E_t denotes the real level of earnings during the twelve months preceding time t . The model is tested using the S&P composite index only due to earnings data availability. Panel C reports results for Lintner (1956) model 2

$$D_t - D_{t-1} = \lambda_0 + \lambda_1 E_{t-1} + \lambda_2 D_{t-1} + \varepsilon_t.$$

The model is tested using the S&P composite index only due to earnings data availability. Regressions are estimated by OLS and figures in parentheses are t-statistics using Newey-West (1987) standard errors. *** denotes significance at the 1% level, ** denotes significance at the 5% level and * denotes significance at the 10% level. While the cut-off R^2 is reported in Panels B and C, Panel A shows the modified cut-off R^2 which recognizes the dependency of the dependant and independent variables in the equation.

Panel A: Marsh and Merton (1987) Model					
$\log(D_{t+1}/D_t) + D_t/P_{t-1} = \psi_0 + \psi_1[(P_t + D_t)/P_{t-1}] + \psi_2 \log(D_t/P_{t-1}) + \varepsilon_{t+1}$					
Stock Index	ψ_0	ψ_1	ψ_2	Adjusted R^2	Modified Cut-off R^2
Equally-Weighted NYSE Index	-0.311 (-0.979)	0.493 (3.494)***	-0.110 (-1.152)	42.13%	59.61%
Value-Weighted NYSE Index	-0.046 (-0.278)	0.321 (2.799)***	-0.025 (-0.507)	28.88%	42.17%
S&P Composite Index	-0.037 (-0.234)	0.357 (3.193)***	-0.022 (-0.461)	32.44%	43.31%
Panel B: Lintner (1956) Model 1					
$D_t = \theta_0 + \theta_1 E_{t-1} + \theta_2 D_{t-1} + \varepsilon_t$					
Stock Index	θ_0	θ_1	θ_2	Adjusted R^2	Cut-off R^2
S&P Composite Index	0.019 (2.053)**	0.073 (1.816)*	0.788 (7.776)***	91.79%	81.25%
Panel C: Lintner (1956) Model 2					
$D_t - D_{t-1} = \lambda_0 + \lambda_1 E_{t-1} + \lambda_2 D_{t-1} + \varepsilon_t$					
Stock Index	λ_0	λ_1	λ_2	Adjusted R^2	Cut-off R^2
S&P Composite Index	0.019 (2.053)**	0.073 (1.816)*	-0.212 (-2.091)**	8.64%	7.98%

Table 6
Results on a Reformulated Lintner (1956) Dividend First-Difference Model

This table presents results for the reformulated Lintner (1956) dividend first-difference regression model on the annual CRSP equally-weighted and value-weighted New York Stock Exchange indices and the annual Standard and Poor Composite Index for the time period 1927 to 1996:

$$D_t - D_{t-1} = \theta_1(P_{t-1} - P_{t-2}) + \theta_2(D_{t-1} - D_{t-2}) + (\varepsilon_t - \varepsilon_{t-1})$$

where D_t is the level of real annual dividends during the twelve months preceding time t and P_t is the real stock index level at time t . Both real and detrended real data are examined. Regressions are estimated by OLS and t-statistics are estimated using Newey-West (1987) standard errors. In the simulation procedure, the error term $(\varepsilon_t - \varepsilon_{t-1})$ is accommodated using a Heteroskedasticity and Autocorrelation Consistent Covariance estimation procedure within Generalized Method of Moments. *** denotes significance at the 1% level, ** denotes significance at the 5% level and * denotes significance at the 10% level. The cut-off R^2 is obtained using a simulation procedure where dependent and independent variables are uncorrelated but have the same autocorrelation properties as the actual data. Note that the modified cut-off R^2 is not required due to the properties of the dependant and independent variables in the model.

	Real Data		Detrended Real Data	
	Coeff.	t-stat	Coeff.	t-stat
<i>Equally-Weighted NYSE Index</i>				
$(P_{t-1} - P_{t-2})$	0.004	2.794***	0.012	2.326**
$(D_{t-1} - D_{t-2})$	0.406	2.769***	0.059	0.531
Adj. R^2 (%)	16.65%		26.30%	
Cut-off R^2 (%)	2.48%		6.71%	
<i>Value-Weighted NYSE Index</i>				
$(P_{t-1} - P_{t-2})$	0.006	2.334**	0.015	3.108***
$(D_{t-1} - D_{t-2})$	0.144	1.257	0.111	0.803
Adj. R^2 (%)	11.52%		30.71%	
Cut-off R^2 (%)	5.89%		6.06%	
<i>S&P Composite Index</i>				
$(P_{t-1} - P_{t-2})$	0.008	3.101***	0.021	3.709***
$(D_{t-1} - D_{t-2})$	0.150	1.012	0.093	0.638
Adj. R^2 (%)	22.49%		40.16%	
Cut-off R^2 (%)	5.02%		6.81%	