Common Divisors, Payout Persistence, and Return Predictability

JOHN POWELL[†], JING SHI[‡], TOM SMITH[§] AND ROBERT WHALEY[¶]
[†]Massey University, Palmerston North, New Zealand,
[‡]Australian National University, Canberra, Australia and Jiangxi
University of Finance and Economics, Jiangxi, China
[§]Australian National University, Canberra, Australia and
[¶]Vanderbilt University, Nashville, TN, USA

ABSTRACT

In the finance and accounting literature, the use of a common divisor in the dependent and independent variables of ordinary least-squares regressions is commonplace. What goes less recognized, however, is that their use induces spurious correlation between the regression variables and invalidates standard testing procedures. This paper analyses the common divisor problem by outlining analytical results concerning the expected R^2 and providing a simulation procedure that generates test statistics from which critical values can be drawn. To illustrate the procedure, we re-investigate payout yield return predictability findings that have appeared in the literature and show that the results are spurious.

I. INTRODUCTION

Return predictability is the allure of finance. Finding mechanical rules that generate abnormal risk-adjusted returns after trading costs has been the pursuit of investors for generations. Interestingly, return predictability has also been the subject of intense academic interest. Dividend yield return predictability, for example, has been documented so widely that it has been referred to as 'one of the new facts in finance.' Although its predictability was challenged in recent years, researchers quickly recognized that an increasing proportion of firms pay out cash to shareholders through share repurchases rather than cash dividends, and, when the payout yield definition is broadened to include share repurchases, the payout yield return predictability result is restored.

The purpose of this paper is not to provide a mechanical trading rule or to contribute to the burgeoning academic literature on payout yield return predictability. Instead, it is intended to be a cautionary note. This paper provides strong evidence that the documented return predictability of payout yield is spurious. The payout yield return predictability regression model, like so many regression models in finance and accounting, has dependent and

independent variables that share a common divisor. Common divisors can induce correlation between the regression variables, when no relation actually exists, and invalidate the use of standard hypothesis testing procedures. That is not to say that ratios with common divisors should not be used in regression analysis – only that, when they are used, the correct distribution must be used to judge statistical significance.

Pearson (1897) was the first to note the spurious correlation induced by common divisors. He developed a formula for measuring the expected correlation between the ratios X/Z and Y/Z, where X, Y, and Z are independent. The formula shows that the level of spurious correlation depends on the coefficients of variation of the three variables and, depending on the first two moments of the X, Y, and Z distributions, the expected level of correlation can be quite high. Indeed, in the event X, Y, and Z are identically distributed, the expected correlation is 0.5! This paper builds on Pearson's insight by developing a simulation procedure for generating data under the null and constructing the empirical distributions of the resulting parameter estimates and test statistics from which critical values for hypothesis testing can be drawn.

The structure of the paper is as follows. Section II contains a review of the payout yield return predictability literature. The results of this literature serve only as a backdrop, since the common divisor problem is pervasive in literature extending well beyond return predictability. The payout yield return predictability regression problem is more intricate, however, and allows us to refine the Monte Carlo simulation to handle not only the spurious correlation problem induced by a common divisor but also the spurious regression problem induced by variable (i.e., payout and price level) persistence. Section III illustrates the payout yield return predictability result using annual return and payout yield data over the 78-year period 1926 through 2003. Our results, like past studies, indicate that returns are predictable when standard test procedures are applied. More important, we show the payout yield is not the key factor in predicting return: most, if not all, of the return predictability arises from using the inverse of price on the right-hand side of the regression equation. We rely on the statistics literature to show the levels of correlation between variables that can be expected when variables share a common divisor, and then outline the steps of a Monte Carlo simulation procedure that accounts for the effects and provides critical test-statistics for the purpose of hypothesis testing. We conclude by describing different subsets of the literature in finance and accounting whose conclusions may be altered by the use of this technique. Section IV modifies the simulation procedure to account for variable persistence that often appears in financial time series such as payout and price levels. Section V provides some suggestions for possible areas of investigation and generalizes the simulation framework to account for divisors that are not identical but are strongly correlated. Section VI alters the simulation procedure to take account of correlated dividend and price processes. Section VII contains a brief summary of the simulation procedure and regression results.

II. PAYOUT YIELD RETURN PREDICTABILITY LITERATURE

The literature on stock return predictability is extensive. Dividend yield is probably the best-known of the many variables that are found to have stock return forecasting power, with Cochrane (1999, p. 37) labelling it 'one of the new facts of finance.' Dividend yield is typically measured as the ratio of the previous year's dividend payments to the current share price index level. The use of annualized dividends removes the strong seasonal effects in dividend payments and reduces the autocorrelation that appears in monthly and quarterly overlapping dividend yield series. Return is measured as the rate of price appreciation plus dividend yield over a subsequent interval of time, ranging from 1 to 6 years.[1] Evidence that returns are predictable appears in Fama and French (1988a), Harvey (1989), Ferson and Harvey (1991), Whitelaw (1994), Pesaran and Timmermann (1995), Pontiff and Schall (1998), Bossaerts and Hillion (1999), Cremers (2002), Lewellen (2004), and Torous et al. (2004), to name a few. Table 1 summarizes return predictability studies that focus exclusively or almost exclusively on the dividend yield. Goyal and Welch (2003, p. 639) state 'This empirical regularity – that dividend ratios seem to predict equity return – ranks amongst the most important findings of academic finance, and it shows no signs of subsiding.' Indeed, financial time-series studies that employ business cycle variables now commonly include dividend yield as one of the control variables. Recent examples include Baker and Wurgler (2004), Bansal and Yaron (2004), Almeida et al. (2004), and Brandt and Santa-Clara (2006).

   That is not to say that there have not been challenges to the dividend yield return predictability result, only that the challenges have been few. Black and Scholes (1974) and Goetzmann and Jorion (1993), for example, document that the statistical relation between dividend yield and stock return is not strong. Bossaerts and Hillion (1999) use a number of statistical model selection criteria to examine the predictability of stock return using dividend yield and find in-sample predictability but no out-of-sample forecasting power. The most significant challenge, perhaps, is the recent recognition that the dividend yield return predictability result has waned through time. Goyal and Welch (2003) and Lettau and Ludvigson (2005) find the predictive power of the dividend yield is present in pre- but not post-1990 data. Robertson and Wright (2006) and Boudoukh et al. (2007) respond to this challenge by arguing that the value of a stock is driven by all shareholder cash flows, not just cash dividends. Share repurchases are dividend substitutes and, in recent years, have become an increasingly popular and important means of providing payout to share-holders.[2] When the shareholder payout definition is broadened to include share

---

1 Boudoukh et al. (2008) show that these results are indistinguishable since the statistics are almost perfectly correlated.
2 See, for example, Fama and French (2001), Grullon and Michaely (2002), and Liang and Sharpe (1999), among others.

Table 1 Fourteen previous studies on predictability of the dividend yield

Study	Sample data	Sample period	Frequency of data	Return horizon	Empirical finding	R^2 or $\bar{R}^2$
Shiller (1981)	SP	1872–1983	A	1	++	0.06
Rozeff (1984)	SP	1926–1982	A	1	++	0.11
Campbell and Shiller (1988)	SP VW	1871 (1926)–1986	A	1	++	NA
Fama and French (1988a)	VW EW	1927–1986	M/Q/A	M/Q/1/ 2/3/4	++	0.00–0.29
Balvers et al. (1990)	VW	1947–1987	A	1	++	0.069
Nelson and Kim (1993)	SP	1872–1986	A	1/3/10	+	0.016–0.267
Goetzmann and Jorion (1993)	SP	1927–1990	M	M/1/2/3/ 4	+	0.01–0.39
Kothari and Shanken (1997)	VW EW	1926–1991	A	1	++	0.014–0.075
Lamont (1998)	SP	1947–1994	Q	Q	++	0.05
Stambaugh (1999)	VW	1927–1996	M	M	+	NA
Goyal and Welch (2003)	VW	1926–2002	A	1	+	0.0431
Lewellen (2004)	VW EW	1946–2002	M	M	++	0.004–0.008
Lettau and Ludvigson (2005)	VW	1948–2001	A	1/2/3/4/ 5/6	+	0.08–0.23
Campbell and Yogo (2006)	VW SP	1880–2002	M/Q/A	M/Q/1	++	NA

This table presents model specifications used in 14 recent empirical studies on the predictability of the lagged dividend yield on stock index return. In the column titled 'sample data,' VW refers to CRSP value-weighted index, EW refers to CRSP equal-weighted index, and SP refers to S&P 500 (or composite) index. The column 'sample period' reports the full sample period used in each study (although some studies include sub-period results). In the column 'frequency of data,' M refers to monthly data, Q refers to quarterly data, and A refers to annual data. In the column 'return horizon,' M (Q) denotes monthly (quarterly) return, 1 denotes 1-year return, 2 denotes 2-year return, and so on. Under 'empirical finding,' '+' is used to signify a positive but insignificant relation, and '++' signifies a significant positive relation. The final column contains the reported levels of R^2 or $\bar{R}^2$. NA denotes not available. Note that in Campbell and Shiller (1988), the sample period is 1887–1986 for SP and is 1926–1986 for VW.

repurchases, the payout yield return predictability re-emerges in recent years as strongly as ever.

III. RETURN PREDICTABILITY AND COMMON DIVISORS

The payout yield return predictability results reviewed in the last section provide a convenient and useful means of illustrating and analyzing the common divisor problem. In this section, we first reproduce the current state of the literature using annual price and payout data for the CRSP value-weighted

stock market index. Next, we show that the level of price rather than the level of payout is the more likely source of predictability. Third, we provide a brief review of the statistical literature that has a bearing on the problem at hand and show the levels of correlation that one can expect given the presence of a common divisor on both sides of the regression equation. Finally, we develop a simulation procedure to compute critical values of test statistics for assessing predictability in the presence of a common divisor.

A. Evidence of predictability

To begin, we verify the current state of the literature regarding payout yield return predictability. The total return and capital return stock indexes used in this study are for the annual CRSP value-weighted index over the 78-year sample period January 1926 through December 2003. The stock index price level and dividend data are obtained from CRSP, and the series are converted from nominal to real values by dividing by the Consumer Price Index obtained from Robert Shiller's web page. The dividend yield of the index at time t equals the level of real dividends during the 12 months preceding time t, D_t, divided by the real stock index level at time t, P_t.[3] Aside from dividend yield, two other payout yield series are used. Both are from Boudoukh et al. (2007)[4] and attempt to recognize that, aside from cash dividends, shareholders can receive payout from the firm through share repurchases. Boudoukh and colleagues construct the level of repurchases in two ways. The first measure, which we denote CF_t, is defined as cash dividends plus the total expenditure on the purchase of common and preferred stocks plus any reduction in the value of the net number of preferred stocks outstanding. Since these data are only available from the cash flows statement for the period 1971 through 2003, $CF_t = D_t$ for the years 1926–1970. The second measure, which we denote TS_t, is defined by the change in treasury stock excluding repurchases that may be earmarked for employee stock option exercise activity. Since treasury stock data are available beginning only in 1984, $TS_t = D_t$ for the years 1926–1983.

In this study, the payout yield return predictability regression is specified as

$$r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1 \left(\frac{X_t}{P_t} \right) + \varepsilon_{t+1} \qquad (1)$$

where the predictor variables (X_t) are (a) the real cash dividend at the time t, D_t (b) the total real payout at time t using the statement of cash flows to measure repurchases, CF_t, and (c) the total real payout at time t using the change in treasury stock to measure repurchases, TS_t. Since our intention is to use the payout yield return predictability result only as an illustration of the common

---

3 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.

4 The payout yield series were generously provided by Michael Roberts.

Table 2 Regressions of annual real index returns on four predictor variables

Period	No. of observations	β_0	$t(\beta_0)$	β_1	$t(\beta_1)$	$\bar{R}^2$
Panel A: $r_{t+1}=\beta_0+\beta_1(D_t/P_t)+\varepsilon_{t+1}$						
1926–2003	77	−0.0571	−0.892	3.4906	2.428	0.0502
1926–1984	58	−0.2079	−2.462	6.2721	3.457	0.1097
Panel B: $r_{t+1}=\beta_0+\beta_1(CF_t/P_t)+\varepsilon_{t+1}$						
1926–2003	77	−0.1257	−1.810	4.7233	3.308	0.0789
1926–1984	58	−0.1967	−2.347	5.9174	3.452	0.0977
Panel C: $r_{t+1}=\beta_0+\beta_1(TS_t/P_t)+\varepsilon_{t+1}$						
1926–2003	77	−0.1000	−1.419	4.3623	2.829	0.0709
1926–1984	58	−0.2080	−2.453	6.2681	3.449	0.1092

This table presents OLS regression results of annual real index return on four predictor variables:

$$r_t \equiv \frac{P_{t+1}+D_{t+1}-P_t}{P_t} = \beta_0 + \beta_1\left(\frac{X_t}{P_t}\right) + \varepsilon_{t+1} \quad (1)$$

where $P_t(P_{t+1})$ is the real stock index level at time $t(t+1)$, and r_{t+1} is the annual real return on the CRSP value-weighted index at time $t+1$. The predictor variables, X_t, are as follows: (a) the real cash dividend at the time t, D_t, (b) the total real payout at time t using the statement of cash flows to measure share repurchases, CF_t, and (c) the total real payout at time t using the change in treasury stock to measure share repurchases, TS_t. The predictor variables (b) and (c) are from Boudoukh et al. (2007) and were generously provided by Michael Roberts. Where repurchase information is unavailable [before 1971 for series (b) and 1984 for series (c)], total payout is set equal to the amount of the dividend only. The sample begins in 1926 and extends through 2003. The early sub-sample is 1926–1984. $\bar{R}^2$ denotes adjusted R^2. Regressions are estimated by OLS and t-statistics are adjusted for autocorrelation and heteroskedasticity using Newey–West (1987).

divisor problem, we ignore other possible prediction variables and use simple (untransformed) returns and yields in the regression specification.

The regression model is fitted to the CRSP value-weighted index data for the full sample period 1926–2003 as well as the early sub-sample 1926–1984. These horizons are chosen to conform with Boudoukh et al. (2007). The regressions are performed using each of the three different payout yield variables, and the results are reported in Panels A–C of Table 2. As expected, Panel A shows that dividend yield return predictability has diminished through time. For the 58-year sub-period 1926–1984, the dividend yield predictor variable enters in a statistically significant manner with an adjusted R^2 of 10.97%. While the result remains significant for the full sample period 1926 through December 2003, the adjusted R^2 falls to a level of 5.02%. Recall this result was first noted by Goyal and Welch (2003) and Lettau and Ludvigson (2005). Panels B and C, on the other hand, show that the payout yield return predictability has taken its place. For the full sample, the adjusted R^2 is 5.02% using D_t as the predictor variable while it is 7.89% and 7.09% for the two total payout yield variables CF_t and TS_t, respectively. Since $CF_t=D_t$ before 1971 and $TS_t=D_t$ before 1984, it is clear that total payout yield dominates dividend yield as a predictor of returns in recent years. Recall this result was first noted by Robertson and Wright (2006) and Boudoukh et al. (2007).

B. Theoretical relation

In considering the predictive power of total payout, it is important to recognize that we are trying to uncover whether the current payout X_t tells us anything about future returns. The null hypothesis is, of course, there is no information content. Such an admission would imply that a trading rule could generate abnormal returns, at least before trading costs. Assuming current payout signals a higher level of future return, the slope coefficient in (1) should be positive, exactly the result that is documented in Table 2 as well as in the past literature. On the other hand, it is possible that current payout signals nothing about future return and that the relation between r_{t+1} and payout yield X_t/P_t is spurious, arising from the statistical correlation induced by the presence of the current price in the denominator of both regression variables.

C. Expected correlation in the presence of a common divisor

The issue of spurious correlation arising from a common divisor is not new. Pearson (1897) addressed the problem and derived an approximate formula for computing the expected level of correlation between ratios. Under the assumption that random variables X, Y, and Z are uncorrelated, Pearson shows that the expected correlation between X/Z and Y/Z is

$$r_{(X/Z,Y/Z)} = \frac{V_Z^2}{\sqrt{(V_X^2 + V_Z^2)(V_Y^2 + V_Z^2)}} \tag{2}$$

where V indicates a variable's coefficient of variation (standard deviation divided by expected value). If, in addition, the uncorrelated random variables X, Y, and Z are identically distributed, $V_X = V_Y = V_Z$ and the approximate correlation formula produces a value of 1/2, even though X, Y, and Z are independent. In the event that X and Y are constants, the correlation is 1.

Pearson's (1897) approximate formula provides an indication that spurious correlation may be severe. Kim (1999), however, argues that the formula is imprecise because it ignores higher order terms. He then goes on to derive an exact formula for the correlation $r_{(X/Z, Y/Z)}$ between ratios of uncorrelated variables with a common divisor Z:

$$r_{(X/Z,Y/Z)} = \frac{V_{1/Z}^2}{\sqrt{[V_X^2(1 + V_{1/Z}^2) + V_{1/Z}^2][V_Y^2(1 + V_{1/Z}^2) + V_{1/Z}^2]}}. \tag{3}$$

Kim's formula is more precise than Pearson's approximate formula when the coefficient of variation of the common divisor (Z) differs significantly from the coefficient of variation of the inverse of the common divisor (*1/Z*). He provides theoretical examples to demonstrate that the correlation between ratios composed of three independent variables can be arbitrarily close to one, and also documents how the spurious correlation problem can affect inferences concerning the empirical relation between birth rates and death rates in 97

countries in the year 1990. Application of equation (3) indicates that a correlation between birth rates and death rates of 0.091 would be expected by chance, even if births, deaths, and population are uncorrelated, thus providing a benchmark to assess whether birth rates are contemporaneously related to death rates.

To illustrate the magnitude of the common divisor effects in the dividend/ payout yield predictability results of Table 2, we apply the Pearson and Kim formulas. For the full sample period, we compute the coefficients of variation for the relevant variables in (2) and (3). The Z_t variable is defined as the price level P_t; hence, $1/Z_t \equiv 1/P_t$. The X_t variable is defined as D_t, CF_t, or TS_t, depending on the regression specification. The Y_t variable is defined as capital appreciation plus income, $P_{t+1}+D_{t+1}-P_t$, or price level plus income, $P_{t+1}+D_{t+1}$. Where $X_t=D_t$ and $Y_t=P_{t+1}+D_{t+1}-P_t$, the expected level of correlation between X/Z and Y/Z according to the Pearson formula is 0.2112, as shown in Panel B of Table 3. Similar orders of magnitude are obtained for $X_t=CF_t$ and $X_t=TS_t$. The expected correlations using the Kim formula are 0.1455, 0.1343, and 0.1405 for the three payout variables, respectively. Clearly spurious correlation is an issue.

The definition of Y_t reveals a weakness of the Pearson and Kim formulas in our situation: the dependent variable's numerator has as one of its elements the denominator, P_t. We know that the slope coefficient in the payout yield return predictability regression (1) (and our indicator of return predictability) is unaffected by our choice of using net return, $r_{t+1} = \frac{P_{t+1}+D_{t+1}-P_t}{P_t}$, or gross return, $1+r_{t+1} = \frac{P_{t+1}+D_{t+1}}{P_t}$, as the dependent variable. The Pearson and Kim formula values, however, are. If we define $Y_t \equiv P_{t+1}+D_{t+1}$ (and, thereby, let the intercept of the regression subsume the effects of the change in the mean), the expected correlation $r_{(X/Z, Y/Z)}$ between return and dividend yield is 0.6627 using the Pearson formula (2) and 0.5185 using the Kim formula (3), as shown in Panel B of Table 3. Whichever definition of Y_t we use, however, the situation is clear. Spurious correlation arises as a result of using a common divisor.

D. Simulation procedure

With the dependent and independent variables sharing a common divisor, standard testing procedures must be modified. While the Pearson and Kim formulas are useful in assessing the approximate amount of expected correlation when variables share a common divisor, they do not provide us with the information needed to assess the statistical significance of the relation. In this section, we outline a simulation procedure that accomplishes this task. First, we describe the steps in a simulation run, and then we explain how to determine critical values of the regression statistics (e.g., $\bar{R}^2$) for hypothesis testing. It should be noted that the simulation procedure outlined below in the remaining part of this section addresses only the common divisor problem. This is done so as to allow the reader to keep separate the common divisor problem (which complicates both cross-sectional and time-series analyses) from the variable persistence problem (which complicates time-series analyses, like payout yield return predictability, in which regression variables are highly auto-

Table 3 Summary of sample descriptive statistics for CRSP value-weighted index during the period 1926–2003

Panel A: Sample summary statistics

Variable	No. of observations	Mean	Standard deviation	Coefficient of variation
P_t	77	299.6499	217.5466	0.7260
$1/P_t$	77	0.005035	0.003097	0.6152
D_t	77	9.6876	2.7806	0.2870
CF_t	77	11.5251	5.3682	0.4658
TS_t	77	10.5283	3.9102	0.3714
$P_{t+1}+D_{t+1}-P_t$	77	19.4094	60.4347	3.1137
$P_{t+1}+D_{t+1}$	77	319.0593	228.0589	0.7148

Panel B: Formula values of expected correlation

	Formula values based on					
	Income, $Y_t \equiv P_{t+1}+D_{t+1}-P_t$			Level, $Y_t \equiv P_{t+1}+D_{t+1}$		
Formula	$X_t \equiv D_t$	$X_t \equiv CF_t$	$X_t \equiv TS_t$	$X_t \equiv D_t$	$X_t \equiv CF_t$	$X_t \equiv TS_t$
Pearson (1897)	0.2112	0.2065	0.2096	0.6627	0.5978	0.6333
Kim (1999)	0.1455	0.1343	0.1405	0.5185	0.4399	0.4812

The variables include: (a) the real price index at time t, P_t, (b) the real cash dividend at time t, D_t, (c) the total real payout at time t using the statement of cash flows to measure share repurchases, CF_t, and the total real payout at time t using the change in treasury stock to measure share repurchases, TS_t. The first two variables are from the CRSP files. D_t is the level of real annual dividends during the 12 months preceding time t. The second two variables are from Boudoukh et al. (2007) and were generously provided by Michael Roberts. Where repurchase information is unavailable [before 1971 for series (c) and 1984 for series (d)], total payout is set equal to the amount of the dividend only. The sample begins in 1926 and extends through 2003.

correlated). We address the additional complication that both the payout and price series are persistent through time in the next section of the paper.

Each simulation run has three steps. First, we generate time series for price levels and dividends assuming the series are independent. The simulated values of payout are generated using

$$\ln X_t = \mu_{\ln X} + e_t \quad \text{for } t = 1, 2, 3, \ldots, T \tag{4}$$

where $\mu_{\ln X}$ is the unconditional mean of the $\ln X_t$ series and e_t is a drawing from a standard normal distribution with mean 0 and variance $\sigma_{\ln X}^2$. The time-series specification (4) is run in log form to avoid the possibility of obtaining negative total payout. The parameters $\mu_{\ln X}$ and $\sigma_{\ln X}^2$ are estimated using the sample payout data series. Table 4 contains the means and the standard deviations of the relevant payout series. The mean and standard deviation of the log of dividend payout in the full sample, for example, are 2.2309 and 0.3227, respectively. The length of the series is set equal to $T=78$, the length of the actual annual time series of our sample, that is, 1926–2003. After the time series are generated, they are exponentiated to obtain the data in levels (i.e., $X_t = e^{\ln X_t}$). The simulated values of price levels are generated in the same way,

Table 4 Summary of sample descriptive statistics for log price and log dividend series constructed for CRSP value-weighted and S&P 500 indexes during the period 1926–2004

	Full sample: 1926–2003				Early sub-sample: 1926–1984			
	$\ln P_t$	$\ln D_t$	$\ln CF_t$	$\ln TS_t$	$\ln P_t$	$\ln D_t$	$\ln CF_t$	$\ln TS_t$
No. of observations	78	78	78	78	59	59	59	59
Mean	5.5030	2.2309	2.3544	2.2927	5.2442	2.1298	2.1545	2.1347
Standard deviation	0.6634	0.3227	0.4636	0.3931	0.4980	0.3035	0.3278	0.3090
Lag k auto-correlation								
1	0.925	0.891	0.925	0.915	0.902	0.889	0.874	0.876
2	0.859	0.837	0.875	0.864	0.803	0.818	0.817	0.806
3	0.806	0.794	0.830	0.819	0.759	0.754	0.748	0.738
4	0.749	0.758	0.780	0.775	0.728	0.706	0.711	0.699
5	0.691	0.710	0.733	0.725	0.700	0.648	0.658	0.641

The variables are the natural logarithm of the real price index at time t ($\ln P_t$), the natural logarithm of the real dividend ($\ln D_t$) at time t, the natural logarithm of the total real payout at time t using the statement of cash flows to measure share repurchases ($\ln CF_t$), and the natural logarithm of the total real payout at time t using the change in treasury stock to measure share repurchases ($\ln TS_t$). The first two variables are from the CRSP files. D_t is the level of real annual dividends during the 12 months preceding time t. The latter two variables are from Boudoukh et al. (2007). The sample begins in January 1926 and extends through December 2003.

that is,

$$\ln P_t = \mu_{\ln P} + v_t \quad \text{for } t = 1, 2, 3, \dots, T \tag{5}$$

where the parameters $\mu_{\ln P}$ and $\sigma^2_{\ln P}$ are estimated using the sample price data. At this stage, the innovations in (4) and (5) are assumed to be independent, that is, $\rho(e_t, v_t) = 0$. The estimates of the mean and standard deviation are also provided in Table 4.

The second step in the Monte Carlo simulation run is to use the simulated dividend and price series, X_t and P_t, to generate the dependent and independent variables in the return predictability regression (1). Recall that earlier we define return as

$$r_{t+1} \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t}. \tag{6}$$

Since the beginning- and end-of-period levels of price are used in the computation of (6), the length of the time-series T is reduced from 78 to 77. Similarly, the simulated dividend and price series X_t and P_t are used to generate the payout yield series, computed as

$$X_t/P_t \tag{7}$$

where $X_t = D_t, CF_t, TS_t$, depending on the measure of payout. With the simulated return and payout yield series in hand, the third and final step is to perform an

OLS regression of (1) on the simulated data and record the values of the estimated coefficients, the Newey–West t-statistics, and adjusted R^2, $\bar{R}^2$.

With the mechanics of the simulation run in hand, the focus turns to establishing critical values for hypothesis testing. In all, 10,000 simulation runs are performed for each of the three payout yield return predictability models. For each model, we compile five vectors (of length 10,000) of output results – one each for β_0, β_1, t_{β_0}, t_{β_1}, and $\bar{R}^2$. Each vector is then sorted from lowest to highest. The critical values for the coefficients, the t-statistics, and the adjusted R^2 are the 95, 97.5, and 99 percentiles (i.e., vector observations 9500, 9750, and 9900, respectively). These correspond to tests of significance at the 5%, 2.5%, and 1% probability levels. The standard $\bar{R}^2$ estimated from OLS regression is compared with this critical $\bar{R}^2$ level to assess statistical significance. The critical coefficient estimates and t-statistics are similarly used to evaluate whether the regression estimates obtained using the actual data are statistically significant.

Table 5 contains the critical values of the test statistics for testing the null hypothesis that there is no relation between return and payout yield under the condition that payout and price are not persistent through time. As can be seen from the table values, the critical adjusted R^2 at the 5% probability level is 58.32% for dividend yield, 53.18% for payout yield using the cash flow statement to measure repurchases, and 55.48% for payout yield using changes in treasury stock to measure repurchases. The adjusted R^2 values for the payout yield return predictability regressions reported in Table 2 do not begin to approach the critical values reported in Table 5. Hence, the null hypothesis that returns are unpredictable cannot be rejected.

IV. PAYOUT PERSISTENCE

The Monte Carlo simulation procedure described in the last section addresses the general case of hypothesis testing when the only complication is that the dependent and independent variables share a common divisor. While this covers a wide range of time-series and cross-sectional applications in financial economics and accounting, it does not fully capture the dynamics in the payout yield return predictability regression problem. The reason is that the time-series dynamics described in (4) and (5) are inadequate. As noted by Robertson and Wright (2006), among others, the autocorrelation structures shown in Table 4 reveal a high degree of persistence in both log of payout yield $\ln X_t$ and log of price $\ln P_t$. The usual method for capturing this persistence is to model it as an $AR(1)$ process.[5] Thus, in place of generating payout using (4) in the simulation procedure, we use

---

5 Ferson et al. (2003) show that a spurious regression problem arises when stock return time series (which have been shown to be positively autocorrelated over long periods of time) are regressed on continuous explanatory variables that are persistent (i.e., highly positively autocorrelated).

Table 5 Critical values of coefficient estimates, Newey–West *t*-statistics, and adjusted R^2 levels from Monte Carlo simulation

Probability level	No. of observations	β_0	$t(\beta_0)$	β_1	$t(\beta_1)$	$\bar{R}^2$
Panel A: $r_{t+1}=\beta_0+\beta_1(D_t/P_t)+\varepsilon_{t+1}$						
5.00%	77	−0.1657	−0.817	43.0904	8.366	0.5832
2.50%	77	−0.0965	−0.479	49.5861	9.989	0.6401
1.00%	77	−0.0182	−0.094	58.2473	12.380	0.7018
Panel B: $r_{t+1}=\beta_0+\beta_1(CF_t/P_t)+\varepsilon_{t+1}$						
5.00%	77	0.0506	0.277	30.4734	7.596	0.5318
2.50%	77	0.1156	0.634	34.6152	9.051	0.5908
1.00%	77	0.1924	1.032	40.9189	11.429	0.6570
Panel C: $r_{t+1}=\beta_0+\beta_1(TS_t/P_t)+\varepsilon_{t+1}$						
5.00%	77	−0.0570	−0.305	36.5243	7.959	0.5548
2.50%	77	0.0206	0.105	41.8446	9.405	0.6163
1.00%	77	0.1018	0.549	49.1829	11.863	0.6852
Panel D: $r_{t+1}=\beta_0+\beta_1(D_t/P_t)+\varepsilon_{t+1}$ pre-1984						
5.00%	58	−0.2574	−1.544	28.9656	8.227	0.5530
2.50%	58	−0.2033	−1.252	32.2103	9.486	0.6048
1.00%	58	−0.1403	−0.922	35.5883	11.778	0.6601

This table presents critical values from the distributions of OLS results from regressing simulated annual real index return on three simulated predictor variables:

$$r_t \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1\left(\frac{X_t}{P_t}\right) + \varepsilon_{t+1} \quad (1)$$

where $P_t(P_{t+1})$ is the simulated real stock index level at time $t(t+1)$ and r_{t+1} denotes the simulated annual real return on the CRSP value-weighted index at time $t+1$. The predictor variables, X_t, are as follows: (a) the real cash dividend at the time t, D_t, (b) the total real payout at time t using the statement of cash flows to measure repurchases, CF_t, and (c) the total real payout at time t using the change in treasury stock to measure repurchases, TS_t. The price and payout series in each simulation run are generated in log form to avoid negative values. The series are then exponentiated to get actual levels. The parameters (mean and standard deviation) of each simulated series are set equal to those obtained from the actual data series. (See Table 4.) The length of each simulation run is chosen to correspond to the sample periods 1926–2003 and 1926–1984. $\bar{R}^2$ denotes adjusted R^2. The number of observations in each simulation run is 77, and 10,000 simulation runs are performed. Regressions are estimated by OLS and *t*-statistics are adjusted for autocorrelation and heteroskedasticity using Newey–West (1987). The critical values for the adjusted R^2 are the 95, 97.5, and 99 percentiles.

$$\ln X_t = \mu_{\ln X} + \rho_{\ln X} \ln X_{t-1} + e_t \quad \text{for } t = 2, 3, \ldots, T \quad (8)$$

where the parameters $\mu_{\ln X}$, σ_e^2, and $\rho_{\ln X}$ are estimated from the actual dividend data (see also Granger et al. 2001). The simulation is started at the unconditional mean of the dividend series $\mu_{\ln X}$ and the error term, e_t, is drawn from a normal distribution with mean zero and variance $\sigma_{\ln X}^2(1-\rho_{\ln X}^2)$. Recall that the means and the standard deviations of the payout $\ln X_t$ variables are reported in Table 4. Note that the length of the time series is now $T=77$ due to the presence of a lagged value on the right-hand side of (9). Similarly, the price index series is now generated as

$$\ln P_t = \mu_{\ln P} + \rho_{\ln P} \ln P_{t-1} + v_t \quad \text{for } t = 2, 3, \ldots, T \quad (9)$$

where the parameters $\mu_{\ln B}$ σ_v^2, and $\rho_{\ln P}$ are estimated from the actual price index data. Again, the innovations between the payout (8) and price (9) processes are assumed to be independent. Table 4 shows the relevant parameter estimates for the actual $\ln P_t$ series.

The incremental effects of variable persistence on the critical values for the adjusted R^2 are shown in Table 6. Where the autocorrelation in the log dividend yield and log price variables is 0, the 95% critical value for the adjusted R^2 is 0.5808.[6] As the autocorrelation in the log dividend series increases, however, the critical adjusted R^2 rises. The intuition for this result is that, the higher $\rho_{\ln D}$, the lower the variation in the innovations to the independent variable, and, therefore, the higher the threshold for realizing a statistically meaningful relation. Conversely, as the autocorrelation in the log price series increases, the critical adjusted R^2 falls. Here, the intuition is just the opposite. As autocorrelation rises, variation in the innovations of the dependent variable falls. The percent of the total variation necessary for the relation to be deemed significant in a statistical sense falls.

Table 7 contains the critical values of the test statistics for testing the null hypothesis that there is no relation between return and payout yield when the payout and price are known to be persistent through time. The critical adjusted R^2 at the 5% probability level is 18.98% for dividend yield, 18.62% for payout yield using the cash flow statement to measure repurchases, and 18.85% for payout yield using changes in treasury stock to measure repurchases. The adjusted R^2 values for the payout yield return predictability regressions reported in Table 2 are well below the critical values reported in Table 7. In other words, once the effects of a common divisor and variable persistence are taken into account, the null hypothesis that returns are unpredictable cannot be rejected.

An inevitable response by proponents of return predictability theory is that the dependent variable is the theoretical variable of interest so the regression cannot be spurious. Unfortunately, this is not the case. Whatever the theory, a significant statistical relation can (and is likely to) occur by chance if ratios with a common denominator are used on both sides of the regression equation. The simulation procedure provides a means for determining what part of the relation is spurious and what part is not.[7] What we are arguing is not that

---

6   The value, 0.5808, is nearly identical to its corresponding value reported in Table 5, 0.5832. While, in theory, these numbers should be the same, implementation issues [e.g., the length of the time series used in each simulation run is different (76 versus 77)] cause a minor discrepancy.

7   Brett (2004) and Kim (1999) use simulation and analytical results, respectively, to provide a similar breakdown for correlation analysis. Dunlap et al. (1997, pp. 190–1) warn, based on simulation analysis, of the difficulties that had existed in providing this breakdown when ratios are used in correlation analysis: 'On the other hand, what if one were interested in the ratio variable per se and not its components, . . .? The correlation of a ratio with other variables might be meaningful and the interpretation may be relatively unambiguous as long as the other variables are not ratios constructed with components common to the first ratio. In the latter case, however, any interpretation must be tempered by the knowledge that no good method exists for estimating what proportion of the correlation of ratios sharing elements is spurious and what part is not.'

Table 6 Simulated 95% critical adjusted R^2 values for different levels of first-order auto-correlation in the log dividend and log price series

First-order auto-correlation in log price series	First-order auto-correlation in log dividend series									
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.0	0.5808	0.5694	0.5757	0.5746	0.5714	0.5751	0.5828	0.5877	0.5906	0.6006
0.1	0.5330	0.5400	0.5361	0.5380	0.5390	0.5378	0.5517	0.5432	0.5540	0.5638
0.2	0.4880	0.4952	0.4973	0.5009	0.4987	0.5001	0.4971	0.4972	0.5075	0.5207
0.3	0.4471	0.4538	0.4549	0.4568	0.4528	0.4655	0.4620	0.4621	0.4621	0.4741
0.4	0.4055	0.4055	0.4083	0.4084	0.4147	0.4161	0.4144	0.4142	0.4241	0.4274
0.5	0.3622	0.3583	0.3590	0.3610	0.3627	0.3695	0.3678	0.3741	0.3716	0.3811
0.6	0.3153	0.3094	0.3126	0.3168	0.3162	0.3173	0.3150	0.3178	0.3214	0.3298
0.7	0.2594	0.2597	0.2601	0.2642	0.2649	0.2715	0.2725	0.2780	0.2734	0.2789
0.8	0.2054	0.2110	0.2100	0.2174	0.2138	0.2205	0.2245	0.2245	0.2297	0.2288
0.9	0.1626	0.1612	0.1646	0.1708	0.1742	0.1775	0.1824	0.1873	0.1877	0.1916

This table presents critical adjusted R^2 values from the distribution of OLS results from regressing simulated annual real index return on simulated dividend yield:

$$r_t \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1\left(\frac{D_t}{P_t}\right) + \varepsilon_{t+1}$$

where $P_t (P_{t+1})$ is the simulated real stock index level at time $t(t+1)$ and r_{t+1} denotes the annual real return on the CRSP value-weighted index at time $t+1$, and D_t is the real cash dividend at the time t. The price and dividend series in each simulation run are generated in log form to avoid negative values. The series are then exponentiated to get actual levels. The parameters (mean and standard deviation) for each of the simulated log series are set equal to those obtained from actual data series. (See Table 4.) The levels of first-order autocorrelation for each series are arbitrary. The number of observations in each simulation run is 76, and 10,000 simulation runs are performed. Regressions are estimated by OLS and t-statistics are adjusted for autocorrelation and heteroskedasticity using Newey–West (1987). The critical values for the adjusted R^2 are for the 95 percentile.

Table 7 Critical values of coefficient estimates, Newey–West *t*-statistics, and adjusted R^2 levels from Monte Carlo simulations with dividend and price persistence

Probability level	No. of observations	β_0	$t(\beta_0)$	β_1	$t(\beta_1)$	$\bar{R}^2$
Panel A: $r_{t+1}=\beta_0+\beta_1(D_t/P_t)+\varepsilon_{t+1}$						
5.00%	77	0.0139	0.289	7.5902	5.099	0.1898
2.50%	77	0.0229	0.507	9.0605	5.658	0.2154
1.00%	77	0.0351	0.756	10.6699	6.420	0.2544
Panel B: $r_{t+1}=\beta_0+\beta_1(CF_t/P_t)+\varepsilon_{t+1}$						
5.00%	77	0.0125	0.282	7.3804	5.101	0.1862
2.50%	77	0.0231	0.505	8.6987	5.621	0.2151
1.00%	77	0.0341	0.776	10.6753	6.292	0.2566
Panel C: $r_{t+1}=\beta_0+\beta_1(TS_t/P_t)+\varepsilon_{t+1}$						
5.00%	77	0.0133	0.272	7.7648	5.153	0.1885
2.50%	77	0.0252	0.545	9.1843	5.675	0.2147
1.00%	77	0.0354	0.837	10.9184	6.399	0.2548
Panel D: $r_{t+1}=\beta_0+\beta_1(D_t/P_t)+\varepsilon_{t+1}$ pre-1984						
5.00%	58	−0.0016	−0.030	7.1755	4.883	0.2009
2.50%	58	0.0102	0.195	8.4059	5.301	0.2242
1.00%	58	0.0267	0.526	9.9271	5.938	0.2594

This table presents critical values from the distributions of OLS results from regressing simulated annual real index return on three simulated predictor variables:

$$r_t \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1\left(\frac{X_t}{P_t}\right) + \varepsilon_{t+1} \quad (1)$$

where $P_t(P_{t+1})$ is the simulated real stock index level at time $t(t+1)$ and r_{t+1} denotes the annual real return on the CRSP value-weighted index at time $t+1$. The predictor variables (X_t) are (a) the real dividend yield at the time t (D_t/P_t), (b) the total real payout yield at time t using the statement of cash flows (CF_t/P_t), and (c) the total real payout yield at time t using the change in treasury stock (TS_t/P_t). The price and payout series in each simulation run are generated in log form to avoid negative values and are assumed to follow individual $AR(1)$ processes. The series are then exponentiated to get actual levels. The parameters (mean, standard deviation, and autocorrelation) of each simulated series are set equal to those obtained from the actual data series. (See Table 4.) $\bar{R}^2$ denotes adjusted R^2. The number of observations in each simulation run is 76, and 10,000 simulation runs are performed. Regressions are estimated by OLS and *t*-statistics are adjusted for autocorrelation and heteroskedasticity using Newey–West (1987). The critical values for the coefficients, the *t*-statistics, and the adjusted R^2 are the 95, 97.5, and 99 percentiles. Within the simulation run, dividends and price levels are assumed to follow individual $AR(1)$ processes whose parameters match the actual data.

researchers should avoid using ratios in regression analysis, but rather that, when they do so, the correct distribution must be used to judge statistical significance. It is standard in the literature to adjust parameter estimates for biases when assessing statistical significance. Well-known bias-adjustments are used across disciplines and include such standard items as the sample variance, where $n-1$ appears in the denominator instead of n, and the sample autocorrelation function, where the bias is a function of $1/T$. Within a finance context, it is well known that overlapping returns must be adjusted for bias. Fama and French (1988b) show that when multi-period returns are regressed on

lagged multi-period returns, biases are functions of $1/T$, with the cumulative bias being as high as 0.3. In our case, the bias affects not only the parameter estimates but also the whole distribution. We correct for this bias by means of Monte Carlo simulation when assessing statistical significance.

V. CORRELATED DENOMINATORS

Payout yield return predictability has served as a convenient illustration for demonstrating how Monte Carlo simulation can be used to address the common divisor problem in regression analysis. The procedure involves generating data under the null and constructing empirical distributions of the resulting parameter estimates and test statistics. Critical values for hypothesis testing are then drawn from the empirical distributions. The procedure is general and can be used in a variety of contexts. In the payout yield return predictability analyses, the common divisor, price, is used in a time-series context. In the microstructure literature, price is used in cross-section to deflate bid/ask spread and its determinants (e.g., Bollen et al. 2004). In corporate finance, the market value of equity (e.g., Grullon and Michaely 2002; Daniel and Titman 2006) and the total assets of the firm (e.g., Allayannis and Weston 2001; Faulkender and Petersen 2005; Alti 2006) are frequently used divisors in regression analyses. The number of common shares outstanding has also been used (e.g., Collins et al. 1997). In all of these instances, Monte Carlo techniques can be used to overcome the spurious correlation problem.

In its simplest form, the common divisor problem arises when regression variables other than the constant in the regression equation are divided by a common variable to control for a common confounding influence. As a result, regression coefficient estimates will be biased, and, more importantly, a significant overall relation between the dependent and independent regression variables may appear even when all of the component variables making up the numerator and the denominator of the dependent and independent regression variables are uncorrelated. Kronmal (1993) outlines a number of cases in which incorrect inferences are made as a result of spurious regression results. Among them are empirical studies examining the cross-sectional relation between weight size and body size and the question of whether females lose lung capacity at a faster rate than males as they age. Aside from Monte Carlo simulation, the problem can sometimes be addressed directly by suppressing the intercept in the regression and including the inverse of the divisor as an independent variable (e.g., Bollen et al. 2004). There remain a host of other situations where such a direct method cannot be used, however. In the remaining part of this section, we extend the Monte Carlo framework to include situations where the divisors on the two sides of the regression equation are not the same but are correlated.

Regressions using correlated divisors appear frequently in the finance and accounting literature. Fama and French (1992), for example, regress stock

returns on the book-to-market and earnings yield ratios.[8] Presumably, to control for spurious correlation, they measure return and the book-to-market ratio 6 months apart; however, the denominators on the two sides of the equation are undoubtedly highly correlated. Faulkender and Petersen (2005) regress the leverage ratio expressed as debt over market value of assets against a market to book ratio variable whose denominator is the book value of assets. Alti (2006) includes regressions where the ratio dependent variable's denominator is assets and the independent variables are ratios with lagged assets as the denominator. In situations such as these, we can modify the simulation procedure to handle correlated divisors in a fairly straightforward fashion. To illustrate, consider the regression model,

$$\left(\frac{Y_t}{P'_t}\right) = \beta_0 + \beta_1 \left(\frac{X_t}{P''_t}\right) + \varepsilon_t \tag{10}$$

where P'_t and P''_t are the simulated quasi-real stock index levels at time t, and Y_t and X_t are uncorrelated simulated series for the dependent and independent variable numerators, respectively. $\ln(P'_t)$ and $\ln(P''_t)$ are assumed to have a bivariate normal distribution with the same mean and standard deviation and with correlation ρ. The mean and standard deviation of each simulated log price and log payout series are set equal to those obtained from the actual data series (see Table 4). Different levels of cross-correlation ρ are assumed. All price and dividend series in each simulation run are generated in log form to avoid negative values. The dynamics can be modelled using (5) and (6) or (9) and (10), depending on the characteristics of the actual data. We use (5) and (6) for illustrative purposes only. The series are then exponentiated to get actual levels.

Table 8 shows the effects of having correlated divisors. Note that, where the correlation is perfect (i.e., $\rho = +1$), the adjusted R^2 values are 0.5767, 0.6308, and 0.6989, at probability levels of 5, 2.5, and 1%, respectively. These correspond to the adjusted R^2 values of 0.5832, 0.6401, and 0.7018 reported in Table 5. The values should be identical in theory and are different only due to numerical implementation issues. At the other extreme, where the divisors are independent (i.e., $\rho = 0$), the threshold critical adjusted R^2 values are the same as those applied in conventional tests. Between the extremes, the critical values rise monotonically with the level of correlation. At a correlation of 0.5, the critical value of the adjusted R^2 is 0.2399 at the 5% percent level. In other words, the effect of correlated divisors, particularly in cross-sectional work where variable persistence is not an issue, can be large.

VI. CORRELATED NUMERATORS

Up to this point, we test the null hypothesis of no return predictability under the assumption that the innovations to the payout and share price processes are

---

8 Others include Pontiff and Schall (1998), Kothari and Shanken (1997), and Lewellen (2004).

Table 8 Simulated critical adjusted R^2 values for different levels of correlation between divisors in dependent and independent variables in OLS regression

Correlation, ρ	Critical adjusted R^2		
	5%	2.5%	1%
0.0	0.0377	0.0602	0.1061
0.1	0.0617	0.0937	0.1460
0.2	0.0984	0.1410	0.2025
0.3	0.1376	0.1879	0.2538
0.4	0.1897	0.2410	0.3102
0.5	0.2399	0.2941	0.3738
0.6	0.3066	0.3685	0.4456
0.7	0.3675	0.4295	0.5126
0.8	0.4406	0.4998	0.5891
0.9	0.5097	0.5757	0.6468
1.0	0.5767	0.6308	0.6989

This table presents critical values from the adjusted R^2 distribution of OLS results from regressing simulated annual real index return on real dividend yield:

$$\left(\frac{Y_t}{P'_t}\right) = \beta_0 + \beta_1 \left(\frac{X_t}{P''_t}\right) + \varepsilon_t$$

where P'_t and P''_t are the simulated quasi-real stock index levels at time t, and Y_t and X_t are uncorrelated simulated series for the dependent and independent variable numerators, respectively. $\ln(P'_t)$ and $\ln(P''_t)$ are assumed to have a bivariate normal distribution with the same mean and standard deviation and with correlation ρ. The mean and standard deviation of each simulated log price and log payout series are set equal to those obtained from the actual data series. (See Table 4.) Different levels of cross-correlation ρ are assumed. All price and dividend series in each simulation run are generated in log form to avoid negative values. The series are then exponentiated to get actual levels. The number of observations in each simulation run is 76, and 10,000 simulation runs are performed. Regressions are estimated by OLS and t-statistics are adjusted for autocorrelation and heteroskedasticity using Newey–West (1987). The critical values for the adjusted R^2 are the 95, 97.5, and 99 percentiles.

independent. Recall the dynamics are modelled by (4) and (5) or (8) and (9). But, more broadly, there are a variety of contexts in which there exists a relation between Y_t and X_t, and we must separate the effects of the economic relation in the numerators from the spurious correlation in the denominators. In this section, we modify the simulation procedure that we used earlier to generate critical cut-off values to assess statistical significance to incorporate correlation between the innovations of the Y_t and X_t processes. To maintain our return predictability illustration, we first provide a theoretical justification for correlation between dividends and share price.

The Lucas (1978) single tree exchange economy capital asset pricing model posits economic relations between dividends and share prices and between dividends and share price changes. In the Lucas exchange economy, there is a single tree that provides a consumption good (i.e., dividends) each period. During the period, individuals either consume the dividends they receive or use them to buy additional shares. Intertemporal consumption optimization, combined with market clearing, determines the equilibrium share price. If all

individuals in the model economy have logarithmic utility functions, the equilibrium share price P_t at time t is

$$P_t = cD_t \tag{11}$$

where D_t is the dividend paid at time t and c is a constant that incorporates the rate of time preference and the expected dividend growth.

To see the relation between dividend and share price changes (returns), assume dividends follow the autoregressive process [recall equation (8)]:

$$\ln D_{t+1} = \mu_{\ln D} + \rho_{\ln D} \ln D_t + \varepsilon_{t+1}. \tag{12}$$

Under this assumption, the equilibrium share price change process is

$$\ln P_{t+1} - \ln P_t = \mu_{\ln P} + (\rho_{\ln D} - 1) \ln D_t + \varepsilon_{t+1}. \tag{13}$$

If dividends follow a random walk (i.e., $\rho_{\ln D}=1$), share price also follows a random walk and returns are independent. If dividends do not follow a random walk (i.e., $\rho_{\ln D}<1$), returns are autocorrelated.

In the Lucas (1978) single tree, logarithmic utility exchange economy, the equilibrium dividend yield, δ_t/P_t, equals a constant c. Consequently, regressing simulated returns on a constant dividend yield is a pointless exercise. It is, however, possible to capture the spirit of a Lucas model by simulating correlated price and dividend series where the dividend process follows either a random walk or an autocorrelated process. In the former case, returns are independent. In the latter case, returns are autocorrelated. In both cases, the simulation procedure can be modified to incorporate correlation between the simulated series. A bivariate normal distribution is used to generate the logarithmic dividend series [see equation (12)] and the logarithmic price series. We allow the degree of correlation to range from 0.5 to 0.99. The higher the correlation, the closer the correspondence to the Lucas (1978) logarithmic utility model (see equations (11) and (13) for the Lucas (1978) logarithmic utility model price and price change equations). Simulated cut-offs for these correlation values are reported in Table 9. Panel A contains the results for the case in which dividends follow a random walk, and Panel B contains the results of the simulations in which dividend autocorrelation level is set at $\rho=0.925$ (i.e., the level present in the actual data).

The results are consistent with expectations. When dividends (and, hence, share price) follow a random walk, the resulting simulated cut-offs for statistical inference are very close to the simulated cut-offs that are estimated using the assumption of independent but highly persistent dividend and price processes. Where the correlation between the dividend and price process is 0.99, for example, the simulated 95% critical value for the adjusted R^2 level in the regression predictability regression is 0.2386, very close to the 0.1916 level reported in the bottom right-hand corner of Table 6, where the dividend and price processes are highly persistent with autocorrelation of 0.9 but independent. When the dividend process follows an *AR*(1) process, share returns become autocorrelated. The autocorrelation, in turn, transmits the correlation between the dividend and share price processes temporally, making it more

Table 9 95% critical values of coefficient estimates, Newey–West *t*-statistics, and adjusted R^2 levels from Monte Carlo simulations when dividends follow random walk and $AR(1)$ processes

Correlation between price and dividend	No. of observations	β_0	$t(\beta_0)$	β_1	$t(\beta_1)$	$\bar{R}^2$
Panel A: Dividends follow a random walk						
0.50	76	0.576	3.505	3.138	19.172	0.8211
0.60	76	0.440	2.995	2.766	19.518	0.8083
0.70	76	0.385	2.659	2.602	19.368	0.7856
0.80	76	0.378	2.412	2.525	18.058	0.7411
0.90	76	0.457	2.370	2.553	14.516	0.6484
0.95	76	0.595	2.224	2.658	11.258	0.5190
0.99	76	1.265	2.162	3.319	6.064	0.2386
Panel B: Dividends follow an AR(1) process						
0.50	76	1.068	3.619	4.190	11.409	0.6499
0.60	76	0.969	3.234	3.738	9.867	0.5567
0.70	76	0.979	2.876	3.652	8.454	0.4600
0.80	76	1.155	2.619	3.767	6.848	0.3388
0.90	76	1.584	2.499	4.136	5.394	0.221
0.95	76	2.257	2.320	4.856	4.303	0.1474
0.99	76	5.296	2.232	7.752	3.131	0.0684

This table presents critical values from the distributions of OLS results from regressing simulated annual real index return on simulated dividend yield variable:

$$r_t \equiv \frac{P_{t+1} + D_{t+1} - P_t}{P_t} = \beta_0 + \beta_1 \left(\frac{D_t}{P_t} \right) + \varepsilon_{t+1}$$

where $P_t(P_{t+1})$ is the simulated real stock index level at time $t(t+1)$ and r_{t+1} denotes the annual real return on the CRSP value-weighted index at time $t+1$. The predictor variable is the real dividend yield at the time t (D_t/P_t). The price and dividend yield series in each simulation run are generated in log form to avoid negative values. In Panel A, dividends are assumed to follow a random walk process, and, in Panel B, dividends are assumed to follow an $AR(1)$ process with a first-order autocorrelation set at 0.925 (the level observed in the actual data). $\bar{R}^2$ denotes adjusted R^2. The number of observations in each simulation run is 76, and 10,000 simulation runs are performed. Regressions are estimated by OLS and *t*-statistics are adjusted for autocorrelation and heteroskedasticity using Newey–West (1987).

difficult to discern between the competing explanations of return predictability. This reduction in the power of the test is reflected in the simulated cut-offs in Panel B of Table 9, which are lower than those in Panel A, particularly at higher levels of correlation between the dividend and price processes.[9] While the application of the Lucas model in this context is somewhat contrived, it is useful in showing how the simulation procedure can be useful in creating

---

9 Table 9 indicates the size of the test under the null of no return predictability as well as size under the alternative of return predictability; hence, the power of the test can be inferred from the size under the null and the alternative.

appropriate cut-off levels for statistical inference for a broad range of problems involving common divisors.

VII. SUMMARY

This paper offers a technique for hypothesis testing when regression variables share a common divisor. The fact that common divisors induce spurious correlation has been noted in the literature for more than 100 years. Pearson (1897) developed a formula for measuring the expected correlation between the ratios X/Z and Y/Z, where X, Y, and Z are independent. The formula shows that the level of spurious correlation depends on the coefficients of variation of the three variables and, depending on the first two moments of the X, Y, and Z distributions, the expected level of correlation can be quite high. In the event X, Y, and Z are identically distributed, the expected correlation is 0.5! This paper builds on Pearson's insight by developing a simulation procedure for generating data under the null and constructing the empirical distributions of the resulting parameters estimates and test statistics from which critical values for hypothesis testing can be drawn.

To illustrate the technique, we re-examine the payout yield return predictability phenomenon that has appeared in the literature. So strong is the reported evidence that the relation has been called one of the new facts of finance. Using Monte Carlo simulation, we show how to address the common divisor problem in regression analysis. We also show how variable persistence such as that found in price and payout time series can also be integrated. Once these effects are modelled, the critical adjusted R^2 values are higher than those reported in the literature, suggesting that the documented statistical correlation in the payout yield return predictability regressions is spurious. We also show how to modify the Monte Carlo simulation procedure to develop critical cut-off levels for statistical inference when denominators on the two sides of the regression equation are not the same but are correlated and when the numerators of the equation are linked theoretically.

Tom Smith
School of Finance and Applied Statistics
Australian National University
Canberra
Australian Capital Territory 0200
Australia
tom.smith@anu.edu.au

REFERENCES

Allayannis, G., and J. Weston (2001), 'The Use of Foreign Currency Derivatives and Firm Market Value', *Review of Financial Studies*, 14, 243–75.

Almeida, H., M. Campello, and M. Weisbach (2004), 'The Cash Flow Sensitivity of Cash', *Journal of Finance*, 59, 1777–804.

Alti, A. (2006), 'How Persistent is the Impact of Market Timing on Capital Structure', *Journal of Finance*, 61, 1681–710.

Baker, M., and J. Wurgler (2004), 'A Catering Theory of Dividends', *Journal of Finance*, 59, 1125–6.

Balvers, R., T. Cosimano, and B. McDonald (1990), 'Predicting Stock Returns in an Efficient Market', *Journal of Finance*, 45, 1109–28.

Bansal, R., and A. Yaron (2004), 'Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles', *Journal of Finance*, 59, 1481–509.

Black, F., and M. Scholes (1974), 'The Effects of Dividend Yield and Dividend Policy on Common Stock Prices and Returns', *Journal of Financial Economics*, 1, 1–22.

Bollen, N., T. Smith, and R. Whaley (2004), 'Modeling the Bid/Ask Spread: Measuring the Inventory-holding Premium', *Journal of Financial Economics*, 72, 97–141.

Bossaerts, P., and P. Hillion (1999), 'Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?', *Review of Financial Studies*, 12, 405–28.

Boudoukh, J., R. Michaely, M. Richardson, and M. Roberts (2007), 'On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing', *Journal of Finance*, 62, 877–915.

Boudoukh, J., M. Richardson, and R. Whitelaw (2008), 'The Myth of Long Horizon Predictability', *Review of Financial Studies*, 21, 1577–1605.

Brandt, M., and P. Santa-Clara (2006), 'Dynamic Portfolio Selection by Augmenting the Asset Space', *Journal of Finance*, 61, 2187–218.

Brett, M. (2004), 'When is a Correlation Between Non-Independent Variables "Spurious"?', *Oikos*, 105, 647–56.

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

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

Cochrane, J. (1999), 'New Facts in Finance', *Economic Perspectives*, 23, 36–58.

Collins, D., E. Maydew, and I. Weiss (1997), 'Changes in the Value-Relevance of Earnings and Book Values Over the Past Forty Years', *Journal of Accounting and Economics*, 24, 39–67.

Cremers, K. (2002), 'Stock Return Predictability: A Bayesian Model Selection Perspective', *Review of Financial Studies*, 15, 1223–49.

Daniel, K., and S. Titman (2006), 'Market Reactions to Tangible and Intangible Information', *Journal of Finance*, 61, 1605–43.

Dunlap, W., J. Dietz, and J. Cortina (1997), 'The Spurious Correlation of Ratios that Have Common Variables: A Monte Carlo Examination of Pearson's Formula', *The Journal of General Psychology*, 124, 182–93.

Fama, E., and K. French (1988a), 'Dividend Yields and Expected Stock Returns', *Journal of Financial Economics*, 22, 3–25.

Fama, E., and K. French (1988b), 'Permanent and Temporary Components of Stock Prices', *Journal of Political Economy*, 96, 246–73.

Fama, E., and K. French (1992), 'The Cross-Section of Expected Stock Returns', *Journal of Finance*, 47, 427–66.

Fama, E., and K. French (2001), 'Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?', *Journal of Financial Economics*, 60, 3–43.

Faulkender, M., and M. Petersen (2005), 'Does the Source of Capital Affect Capital Structure?', *Review of Financial Studies*, 19, 45–79.

Ferson, W., and C. Harvey (1991), 'The Valuation of Economic Risk Premium', *Journal of Political Economy*, 99, 385–415.

Ferson, W., S. Sarkissian, and T. Simin (2003), 'Spurious Regressions in Financial Economics?', *Journal of Finance*, 58, 1393–41.

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

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

Granger, C. W. J., N. Hyung, and Y. Jeon (2001), 'Spurious Regressions with Stationary Series', *Applied Economics*, 33, 899–904.

Grullon, G., and R. Michaely (2002), 'Dividends, Share Repurchases, and the Substitution Hypothesis', *Journal of Finance*, 57, 1649–84.

Harvey, C. (1989), 'Time-Varying Conditional Covariances in Tests of Asset Pricing Models', *Journal of Financial Economics*, 24, 289–317.

Kim, J. (1999), 'Spurious Correlation Between Ratios with a Common Divisor', *Statistics and Probability Letters*, 44, 383–6.

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

Kronmal, R. (1993), 'Spurious Correlation and the Fallacy of the Ratio Standard Revisited', *Journal of Royal Statistical Society*, 156(Series A), 379–92.

Lamont, O. (1998), 'Earnings and Expected Returns', *Journal of Finance*, 53, 1563–87.

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

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

Liang, J., and S. Sharpe (1999), 'Share Repurchases and Employee Stock Options and their Implications for S&P 500 Share Retirements and Expected Returns', Board of Governors of the Federal Reserve System Discussion Paper.

Lucas, R. E. Jr. (1978), 'Asset Prices in an Exchange Economy', *Econometrica*, 46, 1429–45.

Nelson, C., and M. Kim (1993), 'Predictable Stock Returns: The Role of Small Sample Bias', *Journal of Finance*, 48, 641–6.

Newey, W. K., and K. D. West (1987), 'A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica*, 55, 703–708.

Pesaran, M., and A. Timmermann (1995), 'Predictability of Stock Returns: Robustness and Economic Significance', *Journal of Finance*, 50, 1201–28.

Pontiff, J., and L. Schall (1998), 'Book-to-Market Ratios as Predictors of Market Returns', *Journal of Financial Economics*, 49, 141–60.

Robertson, D., and S. Wright (2006), 'Dividends, Total Cashflows to Shareholders and Predictive Return Regression', *Review of Economics and Statistics*, 88, 91–9.

Rozeff, M. (1984), 'Dividend Yields are Equity Risk Premiums', *Journal of Portfolio Management*, 11, 68–75.

Shiller, R. (1981), 'Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?: Reply', *American Economic Review*, 73, 236–7.

Stambaugh, R. (1999), 'Predictive Regressions', *Journal of Financial Economics*, 54, 375–421.

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

Whitelaw, R. (1994), 'Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns', *Journal of Finance*, 49, 515–41.