OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 70, 2 (2008) 0305-9049
doi: 10.1111/j.1468-0084.2007.00499.x

Miller and Modigliani, Predictive Return Regressions and Cointegration*

Piergiorgio Alessandri†, Donald Robertson‡ and Stephen Wright§

†*Bank of England, London EC2R 8AH, UK*
(e-mail: piergiorgio.alessandri@bankofengland.co.uk)
‡*University of Cambridge, Cambridge CB3 9DD, UK*
(e-mail: donald.robertson@econ.cam.ac.uk)
§*Birkbeck College, University of London, London W1E 7HX, UK*
(e-mail: s.wright@bbk.ac.uk)

Abstract

This paper investigates the use of alternative measures of dividend yields to predict US aggregate stock returns. Following Miller and Modigliani [*Journal of Business* (1961), Vol. 34, pp. 411–433] we construct a cashflow yield that includes both dividend and non-dividend cashflows to shareholders. Using a data set covering the course of the 20th century, we show in a cointegrating vector autoregression framework that this measure has strong and stable predictive power for returns. The weak predictive power of standard measures of the dividend yield is explained by the strong rejection of the implied cointegrating and causality restrictions on the impact of non-dividend cashflows.

I. Introduction

There is a large body of research that claims to find evidence that the dividend yield predicts stock returns (e.g. Fama and French, 1988; Campbell and Shiller, 1988, 1998; Jegadeesh, 1990; Pesaran and Timmerman, 1995). In recent years, however, there have been a number of papers (e.g. Goetzmann and Jorion, 1993; Nelson and Kim, 1993; Kirby, 1997; Bossaerts and Hillion, 1999; Foster, Smith and Whaley, 1997; Stambaugh, 1999; Goyal and Welch, 2003; Campbell and Yogo, 2006) that have cast

*This paper was part of Piergiorgio Alessandri's doctoral thesis. It represents the views of the author and not necessarily those of the Bank of England or the Monetary Policy Committee members.
JEL Classification numbers: C32, C53, E44, G10, G14.

doubt on the earlier evidence of predictability, attributing it to data mining or other statistical problems.

This paper is consistent with the revisionist literature, in casting further doubt on the predictive power of standard measures of the dividend yield. However, this does not necessarily destroy evidence of return predictability. Instead, we conclude that the dividend yield has been, in effect, mismeasured – especially in recent years. In Miller and Modigliani (1961), stock market value depends on total cashflows to shareholders, not just the dividend component. We show that an alternative 'cashflow yield', that includes both dividends and non-dividend cash transfers to shareholders, has a more powerful, and more stable ability to predict returns than standard dividend yield measures.

We show that both approaches can be nested in a common econometric framework. Most past tests of predictability have relied on regressions of the form

$$r_t = \alpha + \phi x_{t-1} + u_t \tag{1}$$

where r_t is the log return, and x_t is some predictor variable, most commonly the (conventional) dividend yield. Since Stambaugh (1999), it has become increasingly common to take account of the time-series properties of the predictor variable, usually taken to be of the form

$$x_t = \theta + \rho x_{t-1} + v_t. \tag{2}$$

Both r_t and x_t are usually assumed stationary (although x_t may be highly persistent). We show that these two equations arise from imposing restrictions on a cointegrating vector autoregression (VAR), in which x_t is some combination of underlying non-stationary variables. A key element in our approach is to test these restrictions.

We start from a VAR representation of three non-stationary series: log-real market value (v_t), log-real aggregate dollar dividends (d_t), and a third series (n_t), that captures the impact of non-dividend cashflows. We show that, subject to alternative testable restrictions on cointegrating relations, the VAR in $[v_t\, d_t\, n_t]'$ can be reparameterized into two systems in $[r_t\, x_t\, s_t]'$, where x_t is either our new measure of the cashflow dividend yield or the standard measure. The third variable, s_t, is an additional model-specific variable that absorbs the remaining features of the VAR. We then test a sequence of additional restrictions that produce alternative systems of the form (1) and (2) that are nested in the underlying VAR.

We show that the conventional 'narrow' dividend yield fails at the first hurdle, as the required cointegrating restriction on the non-dividend component of cashflows is very strongly rejected. The tighter restrictions implied by equations (1) and (2) are even more strongly rejected. The very high persistence of the conventional dividend yield that has been associated with bias in return prediction equations (Nelson and Kim, 1993; Stambaugh, 1999; Campbell and Yogo, 2006) appears to arise from the imposition of invalid restrictions.

By contrast, the Miller–Modigliani-consistent cashflow yield can easily be imposed as a cointegrating relation. The remaining restrictions implied by equations (1) and (2) can also be imposed at no significant cost in terms of goodness of fit. Any bias on the coefficient on the cashflow dividend yield in the predictive regression for returns is minor, as, in contrast to the standard narrow dividend yield, the cashflow yield is much less persistent.

In redefining dividends in terms of total cashflow, our work is related to a number of papers that have investigated non-dividend cashflows in other contexts. Most firm-level studies (e.g. Liang and Sharpe, 1999; Fama and French, 2001; Grullon and Michaely, 2002) have focussed on the role of repurchases. While repurchases have indeed grown very rapidly over the past decades, to levels similar to conventional dividends, a number of authors (Bagwell and Shoven, 1989; Ackert and Smith, 1993; Mehra, 1998; Franklin and Michaely, 2003) have noted that their quantitative significance at an aggregate level has in recent years often been outweighed by the impact of cash-financed Mergers and Acquisitions (M & A), that also result in a net flow of cash to shareholders. These have, however, received distinctly less attention in econometric research (a point stressed in Franklin and Michaely's (2003) recent comprehensive review article). The only previous treatment of M&A comparable with our own that we are aware of is in Ackert and Smith (1993).

More recently, Jiang and Lee (2005), Lee and Rui (2007) and Boudoukh *et al.* (2007) all examine time-series properties, but focus primarily on repurchases. Boudoukh *et al.* provide results similar to our own on predictability, using a yield measure for quoted companies, adjusted for repurchases. They also show that their measure of the payout yield helps to explain the cross section of expected returns. The results in this paper complement those of Boudoukh *et al.*, as well as our own earlier research (Robertson and Wright, 2006). In contrast to Boudoukh *et al.*, we have a much longer sample, include M&A, and we focus on the derivation and testing of predictive regressions via cointegrating and causality restrictions in a VAR, that other research on predictive regressions has simply taken as given.

The structure of the paper is as follows. In section II, we examine the link between the stationarity of alternative yield measures and the predictability of returns; section III describes our data set; section IV sets out the general econometric framework in which we operate; sections V and VI describe the tests of the restrictions implied by the two alternative yield measures; section VII compares the resulting predictive regressions for returns; section VIII concludes the paper. Appendices provide details of algebraic manipulations and test procedures.

II. Miller and Modigliani (1961) and the predictability of returns

The aggregate stock market return

Consider (following Miller and Modigliani, 1961) two alternative ways of looking at the aggregate stock market return. The first (commonly used) definition is the return to an investor who holds a single share in the notional company represented by an aggregate stock price index:

$$1+R_{t+1} \equiv \frac{P_{t+1}+D_{t+1}/E_t}{P_t} = \frac{P_{t+1}}{P_t}(1+X^{\mathrm{n}}_{t+1}) \tag{3}$$

where P_t is some aggregate stock price index, measured at the end of period t; D_t is total dividend payments in dollars by the companies in the index during the course of period t; E_t is an index of the number of shares outstanding at the end of period t, defined such that $E_t \equiv V_t/P_t$ where V_t is the value of the stock market; and

$$X^{\mathrm{n}}_t \equiv \frac{D_t/E_{t-1}}{P_t} \tag{4}$$

is the conventional 'narrow' dividend yield: the ratio of dividends per share to the share price. Thus, in familiar fashion, the return consists of the capital appreciation on the notional single share, boosted by the contribution of the per-share dividend yield. We shall define all variables in real terms, although all the above identities will hold equally in nominal terms.

Following Miller–Modigliani, in this discrete time setting dividends are assumed to be paid out, over the course of period $t+1$, to those holding equity at the end of period t, hence the timing convention in defining dividends per share. By implication, net new issues are deemed to take place at the end of the period, hence at the ex-dividend price: $I_{t+1}=P_{t+1}(E_{t+1}-E_t)$, where I_t denotes net corporate equity issues. In Miller and Modigliani's original setting, the net flow of new issues was assumed to be a positive amount, implying a transfer from shareholders to corporations. However, in more recent years, as we shall see, once other cashflows are netted off, this figure has often been a negative amount, and hence a transfer from corporations to shareholders. Thus, let C_t be the total net cashflow from the corporate sector to the non-corporate sector, defined by $C_t \equiv D_t - I_t$. Using the definitions of E_t, I_t and C_t, the return in equation (3) can also be expressed in the following way:

$$1+R_{t+1} = \frac{V_{t+1}+C_{t+1}}{V_t} = \frac{V_{t+1}}{V_t}(1+X^{\mathrm{c}}_{t+1}) \tag{5}$$

where the cashflow dividend yield is defined as

$$X^{\mathrm{c}}_t \equiv \frac{C_t}{V_t} = \frac{D_t - I_t}{V_t}. \tag{6}$$

Definitions and identities are summarized in Table 1. The expression in equation (5) is most easily interpreted as defining the return to a notional representative investor who owns the entire stock market, as the numerator and denominator are, respectively, the wealth of such an investor at the end of periods $t+1$ and t. In this representation, any transfer of cash from the corporate to the non-corporate sector has the same impact on returns.

The equivalence between dividends and repurchases is often discussed, but the role of other flows has received less attention.[1] For an individual firm, new issues are, of course, *not* precisely equivalent to dividends, as there is no necessity that those

[1] The excellent survey by Franklin and Michaely (2003) provides a notable counter-example.

TABLE 1

Definitions and identities

Net equity issues	$I_t = P_t(E_t - E_{t-1})$
Total net cashflow	$C_t \equiv D_t - I_t$
Narrow dividend yield	$X_t^{\mathrm{n}} \equiv \frac{D_t/E_{t-1}}{P_t}$
Cashflow dividend yield	$X_t^{\mathrm{c}} \equiv \frac{C_t}{V_t}$
Return on equity	$1 + R_{t+1} \equiv \frac{P_{t+1} + D_{t+1}/E_t}{P_t}$
Alternative expressions for the return:	
$1 + R_{t+1} = \frac{V_{t+1} + C_{t+1}}{V_t} = \frac{P_{t+1}}{P_t}(1 + X_{t+1}^{\mathrm{n}}) = \frac{V_{t+1}}{V_t}(1 + X_{t+1}^{\mathrm{c}})$	

Note: P_t, share price; D_t, dividend; E_t, number of shares.

purchasing new issues are existing shareholders. However, once we deal with the aggregate, the representative investor who owns the entire market is the only possible investor – thus, new issues subtract from the representative investor's wealth in a precisely equivalent way to that in which dividends and repurchases add to it. Similarly, cash-financed mergers and acquisitions return cash to shareholders in aggregate (in contrast to equity-financed M&A, that leave the aggregate amount of equity unchanged, but simply switch ownership).

The definition of the return in equation (5) also gives rise, when solved forward for V_t, to the Miller–Modigliani result that dividends, *per se*, are irrelevant to firm value, as, via the identity of a firm's sources and uses of funds, C_t must equal net profit less investment.[2] In a Miller–Modigliani world, therefore, any switch between dividends and non-dividend cashflows, for a given net cashflow, will have no impact on stock market value. On the other hand, any change in the flow of future net cashflows, $\{C_{t+i}\}_{i=1}^{\infty}$, *will* change market value, irrespective of its source.

Dividend yields and return predictability

For our econometric analysis, we work in a log-linear framework, letting lower case letters denote logarithms. To do so, we exploit log-linearized approximations of two identities. First, we derive a measure of the log impact of net new issues on total cashflows to shareholders by defining

$$N_t \equiv \frac{D_t}{C_t} = \frac{D_t}{D_t - I_t}.$$

Total cashflows in logarithms are then given by

$$c_t = d_t - n_t. \tag{7}$$

Using equation (7), together with the definition of E_t and I_t, we show in Appendix A that we can derive the log-linearized approximation

[2]Interpreted either as physical investment for a notional 100% equity firm or, more generally, as total investment (i.e. net of financial asset accumulation or decumulation) for a geared firm.

$$\Delta e_t = \Delta v_t - \Delta p_t \approx \delta n_t \tag{8}$$

where $\delta \equiv \exp(\bar{x}^c)$ and $\bar{x}^c$ is the mean of the (log) cashflow yield. Exploiting this identity between the level of non-dividend cashflows and the rates of change of prices and market value, we can derive two parallel log-linear approximations for the log return, $r_{t+1} \equiv \ln(1+R_{t+1})$, in terms of alternative measures of the dividend yield:

$$\begin{aligned} r_{t+1} &\approx \varphi + \Delta p_{t+1} + (1-\rho)x^n_{t+1} \\ &\approx \varphi + \Delta v_{t+1} - \delta n_{t+1} + (1-\rho)(d_{t+1} + \delta n_{t+1} - v_{t+1}) \end{aligned} \tag{9}$$

$$\begin{aligned} r_{t+1} &\approx \varphi + \Delta v_{t+1} + (1-\rho)x^c_{t+1} \\ &\approx \varphi + \Delta v_{t+1} + (1-\rho)(d_{t+1} - n_{t+1} - v_{t+1}) \end{aligned} \tag{10}$$

where $\rho \equiv [1+\exp(\bar{x}^c)]^{-1}$ (hence $\delta \equiv (1-\rho)/\rho$) and $\varphi \equiv \ln(1+\exp(\bar{x}^c)) - (1-\rho)\bar{x}^c$.

Equation (9) is the familiar Campbell and Shiller (1988) approximation that decomposes the log return into the log price change and the narrow dividend yield. Equation (10) provides an alternative decomposition in terms of the log change in market value and the cashflow yield.[3] By exploiting $c_t \equiv d_t - n_t$ and equation (8), both can be expressed in terms of the same three variables v_t, d_t and n_t.

Following Campbell and Shiller (1980), the alternative expressions for log returns in equations (9) and (10) can each be solved forward (subject to the appropriate transversality conditions) to yield expressions for the two alternative yields as forward sums:

$$x^n_t \approx d_t - v_t + \delta n_t \approx -\frac{\varphi}{1-\rho} + \sum_{i=1}^{\infty} \rho^{i-1}[r_{t+i} - \{\Delta d_{t+i} - \delta n_{t+i-1}\}] \tag{11}$$

$$x^c_t \equiv d_t - v_t - n_t \approx -\frac{\varphi}{1-\rho} + \sum_{i=1}^{\infty} \rho^{i-1}[r_{t+i} - \{\Delta d_{t+i} - \Delta n_{t+i}\}]. \tag{12}$$

The term in curly brackets in equation (11) is the growth rate of dividends per share in period $t+i$. This reformulation of the standard Campbell–Shiller forward sum brings out the potentially important feature that future growth rates of dividends per share involve terms in the future *level* of n_t. By contrast, the expression in equation (12) shows that the current value of the log cashflow dividend yield is approximately equal to a linear combination of future log returns and future growth of log cashflow.

The predictability literature has, directly or indirectly, drawn on certain key features of such forward sum specifications. As both are simply log-linearized dynamic accounting identities, the time-series properties of the left-hand side must mirror those of the right-hand side. Cochrane (2001) forcefully makes the point that if the left-hand side of equations such as (11) or (12) varies at all then it *must* covary with elements on the right-hand side and consequently each measure of the dividend yield

[3]Here, we exploit the convenient empirical property that $\bar{n}$ and hence $\overline{\Delta e}$ are insignificantly different from zero. Hence $\overline{x^c}$, which feeds into the linearization parameters, is trivially different from $\overline{x^n}$, which would determine the linearization parameter in the standard Campbell–Shiller approach.

must forecast at least one of the elements on the right-hand side. However, this strong statement depends crucially on an assumption of stationarity of whatever is on the left-hand side of the forward sum (see the discussion in Cochrane, 2001, pp. 399–400).[4]

Research on the predictive power of the dividend yield has usually assumed both that the narrow dividend yield is stationary, and that movements in the dividend yield are associated with movements in future returns rather than future dividend payments. The cashflow yield x_t^c will be stationary as long as r_t is stationary and d_t and n_t contain at most one unit root each. Two of these assumptions – stationarity of returns and unit root in dividends – are quite common and have received good empirical support. By contrast, from equation (11), for the narrow yield x_t^n to be stationary, our measure of the impact of net new issues n_t is also required to be stationary. There is no theoretical reason to expect this to be the case, as even a stable growth rate of total cashflows does not require stability of the share of the non-dividend component of total cashflow. We show that there have indeed been large and possibly permanent shifts in n_t. We argue that this significantly undermines the assumption that x_t^n is stationary, and therefore its ability to predict returns, and may help to explain many of the econometric problems that have been identified in past critiques of the predictability literature.

III. Data

Our data come from a data set that relates to the total non-financial US corporate sector over 1900–2000. The data are described in full in Wright (2004). V_t, D_t and I_t are based on data from the Federal Reserve's Flow of Funds Tables, and Bureau of Economic Analysis (1977) data where these exist, and such historical sources as are available in earlier periods. Starting from 1946, I_t is available in the Flow of Funds data (Table R102, line 11). In recent years, this series has been consistently negative, implying net corporate purchases, due to the combined impact of repurchases and cash-financed mergers and acquisitions. Before 1946, data on these last two components are not available; however, they are likely to be empirically negligible.[5] Wright (2004) constructs a series for new issues for this earlier period with data from various sources (Miller, 1963; *Historical Statistics*; and, for the first decade of the 20th century, various editions of the *Commercial and Financial Chronicle*). V_t, D_t and I_t are deflated by the consumer price index (also taken from Wright, 2004) to derive real values. The series for P_t and R_t can be derived from the core series above using the expressions in section II.

[4]To see this, multiply by $x_t^n - Ex_t^n$ in the case of equation (11) or by $x_t^c - Ex_t^c$ in the case of equation (12), and then take unconditional expectations, thus expressing the variance of the relevant yield as a sum of covariances. Without stationarity, the resulting unconditional variances and covariances do not exist.

[5]Franklin and Michaely (2003) note that before 1983 repurchases were barely legal, and as a result very uncommon. In the period of overlap with Fed data, the alternative sources for new issues that we rely on in earlier periods yield very similar figures, suggesting that the omission of cash-financed acquisitions before 1946 is not empirically significant either.

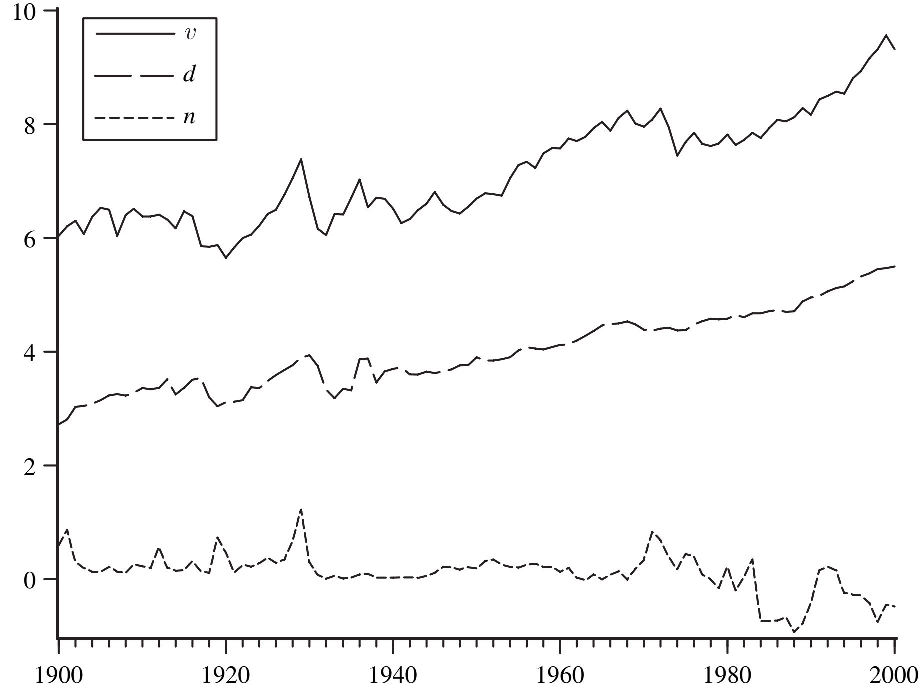

Figure 1. Logarithms of equity value (v_t), dividends (d_t) and a measure of new share issues (n_t)

Figure 1 shows the logarithms of the basic series, namely v_t, d_t and n_t. Figure 2 shows our measures of the approximated log-narrow yield x_t^{n} and the log-cashflow yield x_t^{c}, as defined in equations (11) and (12). The exact log-narrow yield $x_t^{\mathrm{n}} \equiv d_t - v_t + \Delta e_t$ is graphically undistinguishable from the approximated one: the correlation between the two is 0.9999. Cashflow and narrow yields have a very similar mean (3.97% and 3.51% respectively), but at times distinctively different properties. It is noteworthy that these differences were not just evident in the past two decades. In roughly the first two-thirds of the century, the difference between the two series reflected distinct surges in new issues at certain periods (most strikingly in 1929 and in the early 1970s) that lowered the cashflow yield significantly. In other periods, by contrast (most notably the 1930s and the early 1940s), new issues essentially collapsed to zero, such that the two yields were nearly identical. However, in the last two decades of the century, there was a distinct shift, with the difference between the two yields switching sign, as firms engaged in significant levels both of repurchases and geared acquisitions, that more than offset new issues. The impact of the implied adjustment to the dividend yield in recent years is distinctly more significant than in estimates based solely on data for repurchases, as in e.g. Liang and Sharpe (1999) and Fama and French (2001). While there are data coverage differences, the primary explanation is the impact of cash-financed acquisitions in the Fed

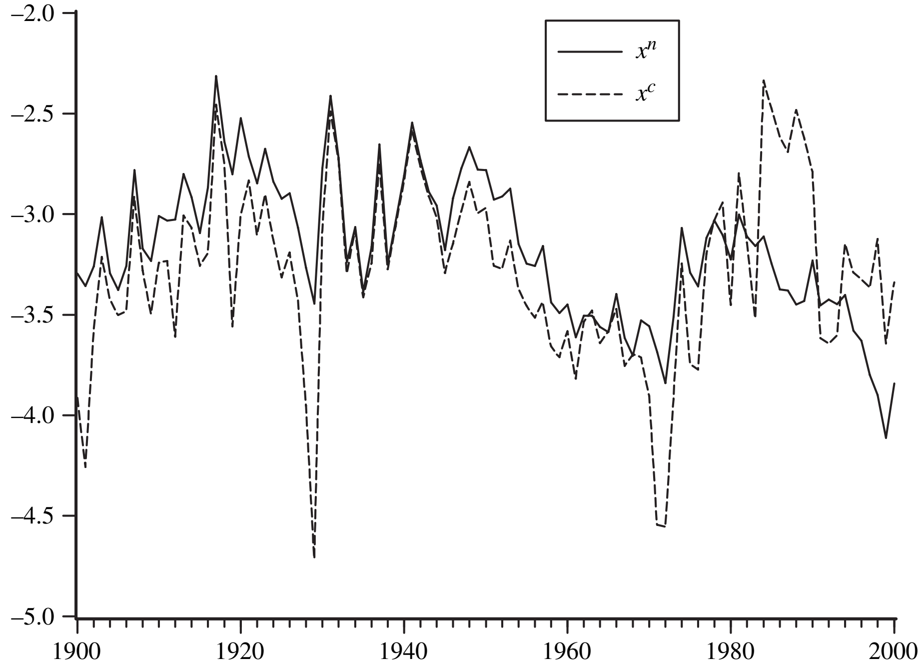

Figure 2. Narrow (n) and cashflow (c) log dividend yields

data.[6] These have been largely ignored in firm-level studies; however, at the aggregate level, as noted in the previous section, they have an identical role in returning cash to shareholders (as noted by, e.g. Bagwell and Shoven, 1989; Ackert and Smith, 1993; Mehra, 1998; Franklin and Michaely, 2003). The chart also shows that, while the cashflow yield is distinctly more volatile than the narrow yield, it appears to have a stronger tendency to mean reversion. The narrow yield shows a distinct downward drift in the latter part of the sample, whereas the cashflow yield ended the century more or less at its mean.

IV. The econometric framework

The cointegrating vector autoregressive representation

The discussion of alternative dividend yields in section II showed that all analysis could be expressed in terms of the three variables v_t, d_t and n_t. Our starting point is therefore a VAR representation of the vector $\mathbf{y}_t = [v_t\ d_t\ n_t]'$. Standard lag order selection criteria show that a VAR(2) representation is easily accepted. The residuals are serially uncorrelated, but they are strongly kurtotic and heteroscedastic. Hence, to guarantee valid inference, for all the tests of sections V and VI we report bootstrapped as well as asymptotic P-values. The bootstrap methodology is described in

[6]Wright (2004, section 3.3) provides a detailed comparison.

detail in the 'Bootstrap methodology' section and Appendix C. On the presumption that all three series may be unit-root processes (we present tests in the next section), we estimate a reduced rank system following Johansen (1991) and many others, of the form:

$$\Delta \mathbf{y}_t = \boldsymbol{\Psi} + \boldsymbol{\Phi}\Delta \mathbf{y}_{t-1} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{13}$$

$\boldsymbol{\Phi}$ is a full rank (3×3) matrix; $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both $3 \times r$, where r is the number of cointegrating relations (the rank of the matrix $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$);[7] and $\boldsymbol{\varepsilon}_t$ is a vector of residuals with covariance matrix $\boldsymbol{\Sigma}$.

Time-series properties of the data

Table 2 summarizes the time-series properties of the data, both in terms of the underlying three series, v_t, d_t and n_t, and in terms of the two alternative log dividend yields and in terms of returns.

Univariate test statistics of the null of a unit root given in panel A show that there is fairly strong reason to conclude that v_t and d_t (real market value and real dividends) are both trending unit-root processes, but are stationary in first differences. The tests also clearly reject a unit root in the aggregate returns process.

The evidence is more ambigous in the case of n_t, our measure of the log impact of net new issues (and hence of non-dividend cashflows to shareholders). Our fairly strong prior would be that a representation in which n_t may be subject to permanent shifts is more appropriate than to assume mean reversion. A firm's payout policy is affected by regulatory and fiscal factors that can change in a persistent fashion. For instance, a fiscal reform that makes capital gains more desirable than dividends may induce a persistent (or indeed permanent) change in the mix of dividends and share repurchases.

The presumption that n_t is subject to permanent shifts, and hence is more appropriately represented as a unit-root process, is supported by multivariate tests on the joint properties of the three series, as represented by the system in equation (13). Panel B of Table 2 reports Johansen (1988) rank tests for the existence of r stationary linear combinations of the three series. The test statistics show very strong evidence of at least one such stationary combination, with a more marginal indication that there may be a second. Typically, when such tests are carried out within a system in which all series can be safely assumed to be unit-root processes, r indicates the number of cointegrating relations: linear combinations of the individually non-stationary series that are stationary. In principle, however, it is possible that one of the combinations may contain only a single, independently stationary series. Thus, rather than testing the null of a unit root in n_t, as in standard univariate tests, the multivariate framework allows us to test the null that n_t is stationary, for given values of r. Table 2 panel C

[7]This terminology whereby r is the rank of the system is so standard that we stick to it, trusting that the context in which it is used will avoid any potential confusion with r_t, the log return.

TABLE 2

Univariate and multivariate time-series properties

Panel A: univariate tests for unit roots

Series	*ADF*	*Lags*	*Trend*
Underlying series			
v_t	−2.05	2	Yes
d_t	−2.55	2	Yes
n_t	−3.44*	2	Yes
n_t	−2.62*	2	No
Δv_t	−8.96***	1	No
Δd_t	−8.42***	1	No
Δn_t	−9.32***	1	No
r_t	−8.83***	1	No
Alternative log dividend yields			
$x_t^c \equiv d_t - v_t - n_t$	−3.36**	2	No
$x_t^n \equiv d_t - v_t + \delta n_t$	−1.37	2	No

Panel B: Johansen (1988) trace tests

Estimated eigenvalues: 0.184, 0.126, 0.00008

H_0	*LR*	*5% CV*	*1% CV*
$r = 0$	33.54	29.68	35.65
$r \leq 1$	13.37	15.41	20.04
$r \leq 2$	0.008	3.76	6.65

Panel C: multivariate stationarity tests on n_t

r	*d.o.f*	*LR*	*P-value*	*ALR*	*P-value*
1	2	5.70	0.058	5.027	0.081
2	1	5.30	0.021	4.745	0.029

Notes: *, **, * * * indicate rejection at 10%, 5%, 1% significance level. The number of lagged difference terms included in ADF regressions was chosen on the basis of the Akaike information criterion.

ALR, likelihood ratio including small sample correction (Johansen, 2000).

presents likelihood ratio tests of this null. These either reject the restriction at standard significance levels or imply that there is a very low probability of it holding.[8]

If n_t is a unit-root process, then r will give the number of cointegrating relationships between the three series. Furthermore, if $n_t \sim I(1)$, it is impossible for *both* log yield measures to be cointegrating relations. In due course, we shall test, in a multivariate context, the (mutually exclusive) null hypotheses that each of the alternative log dividend yield measures is a cointegrating relation. As a preliminary to these tests, in panel A of Table 2, we provide univariate unit-root tests of the null

[8]We present likelihood ratio statistics both in their asymptotic form and with a small sample correction for restrictions on cointegrating relations proposed by Johansen (2000).

that each yield is *not* a cointegrating relation. A unit root is strongly rejected for the cashflow dividend yield, but cannot be rejected for the narrow dividend yield. This finding is in line with that of other recent papers (cf., for example Goyal and Welch, 2003). It is also, of course, consistent with the log difference between the two yields ($=(1-\delta)n_t$) being a unit-root process.[9]

Crucially, however, we shall see that none of our results depends on the assumption that n_t contains a unit root. The proposition that n_t is non-stationary (and that x_t^c embodies the correct cointegrating relationship) is one of the conclusions, rather than an assumption, of our analysis.

Reparameterizing the cointegrating VAR

Subject to appropriate restrictions on cointegrating relations, and our two log-linearizations, we can reparameterize the system in $\mathbf{y}_t$ into a stationary VAR in which the first equation is a predictive regression for returns. To show this, we first note that any given cointegrating system of n variables of the form in equation (13) can always be reparameterized in terms of

$$\mathbf{z}_t = \begin{bmatrix} \Delta\tilde{\mathbf{y}}_t \\ \boldsymbol{\beta}'\mathbf{y}_t \end{bmatrix} = \mathbf{K}\Delta\mathbf{y}_t + \begin{bmatrix} \mathbf{0}_{n-r} \\ \mathbf{I}_r \end{bmatrix} \boldsymbol{\beta}'\mathbf{y}_{t-1} \tag{14}$$

with

$$\mathbf{K} = \begin{bmatrix} \mathbf{I}_{n-r} & \mathbf{0} \\ & \boldsymbol{\beta}' \end{bmatrix} \tag{15}$$

where $\mathbf{K}$ is a full-rank matrix,[10] and the sub-vector $\tilde{\mathbf{y}}_t$ picks out the first $n-r$ elements in $\mathbf{y}_t$. Substituting from equation (13), a sequence of manipulations implies a stationary representation in terms of $\mathbf{z}_t$ of the form:

$$\mathbf{z}_t = \mathbf{K}\boldsymbol{\Psi} + \mathbf{A}_1\mathbf{z}_{t-1} + \mathbf{A}_2\Delta\mathbf{z}_{t-1} + \mathbf{K}\boldsymbol{\varepsilon}_t. \tag{16}$$

Derivations and definitions of $\mathbf{A}_1$ and $\mathbf{A}_2$ are provided in Appendix B. The vector $\mathbf{z}_t$ will be stationary as long as $\boldsymbol{\beta}$ does indeed pick out stationary combinations of the elements of $\mathbf{y}_t$.

In our framework, one or other of the dividend yield measures will, under the appropriate null, be contained within $\boldsymbol{\beta}'\mathbf{y}_t$, and we can form the log return as a linear combination of the elements of $\mathbf{z}_t$ using either equation (9) or (10). In all cases, there will, thus, be a straightforward transformation that maps $\mathbf{z}_t$ to $\mathbf{w}_t = [r_t\, x_t\, s_t]'$ of the form

[9]In the case where, under the null, n_t is the *only* stationary relationship in the system, the test statistic is marginal at classical significance levels. But under this null, there can be no cointegration between d_t and v_t, implying that *neither* of the potential dividend yields we wish to examine would be stationary. By implication, from either equations (9) or (10), returns would also be non-stationary. Given the arguments of section II, we have *a priori* grounds for regarding this as unlikely, which is supported by the low P-value for the associated likelihood ratio statistic.

[10]There will always be an ordering (and usually more than one) that ensures this.

$$\mathbf{w}_t = \mathbf{M}\mathbf{z}_t \tag{17}$$

where r_t is (approximated) log returns, x_t is either the (approximated) narrow dividend yield or the (exact) cashflow yield, and s_t is an additional model-specific variable that absorbs the remaining features of the VAR. The matrix $\mathbf{M}$ is full rank, and takes alternative forms for alternative models (see Appendix B). Hence, we have

$$\begin{aligned}\mathbf{w}_t &= \mathbf{MK\Psi} + \mathbf{MA}_1\mathbf{M}^{-1}\mathbf{w}_{t-1} + \mathbf{MA}_2\mathbf{M}^{-1}\Delta\mathbf{w}_{t-1} + \mathbf{MK}\varepsilon_t \\ &= \mathbf{B}_0 + \mathbf{B}_1\mathbf{w}_{t-1} + \mathbf{B}_2\Delta\mathbf{w}_{t-1} + \mathbf{v}_t. \end{aligned} \tag{18}$$

As long as $\boldsymbol{\beta}$ is normalized appropriately, both $\mathbf{K}$ and all versions of $\mathbf{M}$ have unit determinants, implying that the system log likelihood will be invariant to the re-parameterizations. We can thus test restrictions both on the form of the predictive regression for returns, and on the time-series process for alternative definitions of x_t, in the system specified in terms of $\mathbf{w}_t$, against the unrestricted model (of appropriate rank) specified in terms of $\mathbf{y}_t$.

Bootstrap methodology

In testing restrictions in the framework set out in the previous section, we face two complications. The first is directly linked to the data: the residuals from model (13) are heteroscedastic and strongly kurtotic. The second complication is an implication of our modelling strategy. Our null hypotheses are formulated in terms of approximated returns and yields, because these (unlike their exact counterparts) can be obtained as linear combinations of the elements of $\mathbf{y}_t$. The approximation parameters must be estimated from the data, but they are treated as deterministic constants when equation (13) is reparameterized into equation (18). Hence, the hypotheses are tested disregarding any variability in δ, ρ and φ.

The bootstrap is designed to allow us to bypass both problems simultaneously. We first simulate, using resampled residuals, systems of the same form as equation (18) but using *exact* returns and yields, as all the nulls we examine are specified in exact terms. We then generate, for each replication, data sets in terms of $\mathbf{y}_t = [v_t\ d_t\ n_t]$, using exact (nonlinear) identities; estimate the linearization parameters from the simulated data; and finally derive the likelihood ratio statistic using the implied approximated variables.[11] Bootstrapped P-values, shown as 'P-value$^{\mathrm{b}}$' in Tables 3 and 4, are derived from the distribution of the simulated test statistic over 10,000 replications.

V. Testing predictive return regressions in terms of the cashflow dividend yield

We first test specifications in terms of our Miller–Modigliani-consistent cashflow dividend yield, as the restrictions are relatively straightforward.

[11]A fuller description of the procedure, which also uses a 'wild' bootstrap to deal with heteroscedasticity problems, is given in Appendix C.

TABLE 3

Tests of predictive return regressions in terms of the cashflow dividend yield

Model	*r*	*d.o.f.*	*LR*	*P-value*	*P-value*[b]
Panel A: testing the cashflow yield as a cointegrating relation					
$C_1 : x_t^c \sim I(0)$	1	2	1.566	0.456	0.618
$C_2 : x_t^c \sim I(0);\, d_t - \gamma v_t \sim I(0)$	2	1	1.448	0.229	0.297
Panel B: testing exclusion restrictions on return regressions					
Model $C_1 : x_t^c \sim I(0) \rightarrow s_t^c = \Delta d_t\, (H_0 \mid r=1)$					
$s_t^c \not\Rightarrow r_t$		3	6.438	0.092	0.187
$s_t^c \not\Rightarrow r_t, x_t^c$		4	6.796	0.147	0.251
r_t, x_t^c depend only on x_{t-1}^c		8	13.908	0.084	0.152
Panel C: testing exclusion restrictions on return regressions					
Model $C_2 : x_t^c \sim I(0); d_t - \gamma v_t \sim I(0) \rightarrow s_t^c = d_t - \gamma v_t\, (H_0 \mid r=2)$					
$s_t^c \not\Rightarrow r_t$		3	6.345	0.096	0.146
$s_t^c \not\Rightarrow r_t, x_t^c$		5	7.708	0.173	0.229
r_t, x_t^c depend only on x_{t-1}^c		9	14.820	0.096	0.141

Note: $\Rightarrow$ represents Granger causality, *P*-value[b] denotes bootstrapped *P*-values.

TABLE 4

Tests of predictive return regressions in terms of the narrow dividend yield

Model	*r*	*d.o.f.*	*LR*	*P-value*	*P-value*[b]
Panel A: testing the narrow yield as a cointegrating relation					
$N_1 : x_t^n \sim I(0)$	1	2	16.755	0.000	0.057
$N_2 : x_t^n \sim I(0), d_t - \gamma v_t \sim I(0)$	2	1	11.289	0.003	0.035
$N_3 : x_t^n \sim I(0), \Delta v_t - \delta n_t \sim I(0)$	2	5	32.874	0.000	0.012
$N_4 : x_t^n \sim I(0), n_t \sim I(0)$	2	2	11.595	0.003	0.044
Panel B: testing exclusion restrictions on return regressions					
Model $N_3 : x_t^n \sim I(0); \Delta v_t - \delta n_t \sim I(0); s_t^n = \Delta n_t\, (H_0 \mid r=2)$					
$s_t^n \not\Rightarrow r_t$	6		34.296	0.000	0.013
$s_t^n \not\Rightarrow r_t, x_t^n$	7		34.297	0.000	0.017
r_t, x_t^n depend only on x_{t-1}^n	11		73.875	0.000	0.033
Panel C: testing exclusion restrictions on return regressions					
Model $N_4 : x_t^n \sim I(0); n_t \sim I(0); s_t^n = n_t\, (H_0 \mid r=2)$					
$s_t^n \not\Rightarrow r_t$	4		24.769	0.000	0.002
$s_t^n \not\Rightarrow r_t, x_t^n$	6		25.034	0.000	0.005
r_t, x_t^n depend only on x_{t-1}^n	10		64.612	0.000	0.000

Note: $\Rightarrow$ represents Granger causality, *P*-value[b] denotes bootstrapped *P*-values.

Table 3 panel A presents tests of the null that the cashflow dividend yield is a cointegrating relation. Given the definition of the log cashflow dividend yield in equation (12), the null is that one of the columns of $\boldsymbol{\beta}$ is $[-1\ 1\ -1]'$. The only potential complication arises from the ambiguity as to the cointegrating rank of the system, as discussed in the 'Time-series properities of the data' section. It turns out, however, that whether we assume $r = 1$ or 2 is of trivial importance for our results. The table shows that both versions of the test are very easily accepted by the data.[12]

Panels B and C of Table 3 summarize a sequence of restrictions on the predictive regression for returns, that constrain predictive power to arise solely from the cashflow dividend yield.

Given that the restriction on the cointegrating relations is accepted, we can reparameterize in terms of the stationary vector $\mathbf{w}_t^{\mathrm{c}} = [r_t \, x_t^{\mathrm{c}} \, s_t^{\mathrm{c}}]'$ where, in both $r = 1$ and $r = 2$ cases, the first and second equations are for the (approximated) log return and the cashflow dividend yield respectively; with the systems differing only in terms of the third variable, s_t^{c}. Details of the required definitions of $\mathbf{M}$ and other elements of the reparameterization are provided in Appendix B. We can then proceed by relatively straightforward exclusion restrictions on the system in $\mathbf{w}_t^{\mathrm{c}}$, but still testing such restricted versions against the underlying unrestricted system for $\mathbf{y}_t$ of appropriate rank.

We proceed in three stages. We first test the null that s_t^{c} (which is model specific) has no predictive power for r_t. We then test the additional restriction that the second equation (x_t^{c}) is also unaffected by the third variable (hence, there is a block recursive subsystem in $[r_t \, x_t^{\mathrm{c}}]'$). Finally, we test the most restrictive but commonly used form, as in equations (1) and (2), in which both r_t and x_t^{c} can be predicted solely in terms of x_{t-1}^{c}.

The table shows that, in both $r = 1$ and $r = 2$ specifications, all restrictions are statistically acceptable using both conventional and bootstrapped P-values, and that the results differ trivially between the two specifications. In both cases s_t^{c}, the third variable in $\mathbf{w}_t^{\mathrm{c}}$, is redundant in predictive terms for the first two elements in the system. We therefore conclude that the split between dividend and non-dividend cashflows is immaterial in terms of predicting aggregate stock returns, and that only total cashflow matters.

We examine the resulting return regressions, and compare them with those specified in terms of the narrow yield, in more detail in section VII.

VI. Testing predictive return regressions in terms of the narrow dividend yield

The tests of the narrow dividend yield provide a marked contrast.

Table 4 panel A provides tests of the restriction that the narrow dividend yield, x_t^{n} is a cointegrating relation. Given the approximation in equation (11), this requires that $\boldsymbol{\beta}$ contains the vector $[-1 \; 1 \; \delta]'$, where δ is as defined after equation (8). The

[12] In the $r = 2$ case, the remaining cointegrating relation is unrestricted, and can be parameterized in terms of any two of the three variables in the system. We parameterize as $d_t - \gamma v_t$. The estimated value of γ is around 0.75, capturing the historic tendency (discussed in relation to Figure 2) for dividends to grow less rapidly than market value. We are distinctly unconvinced that this second combination represents a truly stationary process; but, as Table 3 shows, in no case does this affect our results.

first two lines of Table 4 panel A show tests of this restriction under $r=1$ and 2 (in the latter case, the remaining cointegrating relation is again unrestricted). Both are strongly rejected using conventional P-values. However, these are tests of nulls under which all three series in the system are assumed to be $I(1)$. From the Campbell–Shiller approximation in equation (9), returns *cannot* be stationary under these nulls, as this 'per-share' version of the linear approximation also contains a term in the level of n_t, our proxy for the impact of non-dividend cashflows, which under the null is $I(1)$. These specifications would, thus, rule out any reparameterization of the system including a stationary return regression.

If we wish to test a null that is consistent with both stationary returns and $n_t \sim I(1)$, we, thus, need to impose additional restrictions alongside the cointegrating relations. The third line of panel A of Table 4 provides a test of restrictions that are consistent both with stationary returns, and with n_t being a unit-root process under the null. From the Campbell–Shiller approximation in equation (9), stationarity of both returns and the narrow yield requires that $\Delta p_t \approx \Delta v_t - \delta n_t$ is a stationary process. By implication, the growth rate of dividends per share, $\Delta d_t - \Delta e_{t-1} \approx \Delta d_t - \delta n_{t-1}$ must also be stationary. Both of these assumptions are quite standard in past research on return predictability. However, if n_t is a unit-root process, a clear (and hitherto neglected) implication is that Δd_t, and hence Δv_t will *not* be stationary processes: i.e. that, in level terms, $v_t, d_t \sim I(2)$. In Appendix B, we show that, while this complicates matters considerably, this null can be represented as a polynomially cointegrated $I(2)$ system, that is nested in the unrestricted $r=2$ case. This specification requires additional restrictions (including zero restrictions) on the structure of $\boldsymbol{\alpha}$, the matrix of error correction parameters, such that the assumed $I(2)$ nature of the processes for v_t and d_t does not 'contaminate' the $I(1)$ process for n_t. Whilst this establishes that there is indeed a logically possible framework (however convoluted it may be) in which both the narrow yield and returns are stationary, despite n_t being $I(1)$, Table 4 shows that the required restrictions are again strongly rejected.

The rejection of the narrow yield does not hinge in any way on our assumption that n_t is an $I(1)$ process. In the fourth line of Table 4 panel A we present a test of the joint null that the narrow yield and n_t are *both* stationary processes.[13] Note that, while this is presented as a test of the narrow yield, this null model could equally be well reparameterized in terms of the cashflow yield, as under the null the log difference between the two yields would be a stationary process. Under this null, therefore, the cointegrating restrictions implied by the two yield representations would be identical. However, as Table 4 shows, this version of the test does nothing to salvage the narrow yield.

The rejection of the various nulls using conventional P-values is generally confirmed by bootstrapped P-values. Under all null hypotheses relating to the narrow

[13]As in the tests discussed in the 'Time-series properties of the data' section, under the null the second cointegrating vector contains zeroes except in the third column; however, in this case *all* elements of $\boldsymbol{\beta}$ are imposed.

yield, there is considerably greater dispersion in the distribution of the likelihood ratio than in the conventional chi-squared distribution. The key factor increasing dispersion is that in some data sets (in contrast to the historic data) the approximated narrow yield can be a poor proxy for the true yield, particularly for those nulls where n_t is an $I(1)$ process. However, even allowing for the greater dispersion of the statistic, the likelihood ratio observed in the data is well out in the tail of the distribution, implying strong rejections of the null. The only case where the rejection is more marginal is the first null, N_1, that we have argued above is mis-specified.

We, thus, conclude that, however we parameterize the time-series properties of the underlying system, the data strongly reject the restrictions on the impact of non-dividend cashflows implied by the assumption that the narrow dividend yield is a stationary process. This in turn implies that any reparameterization in terms of $\mathbf{w}_t^{\mathrm{n}} = [r_t \, x_t^{\mathrm{n}} \, s_t^{\mathrm{n}}]'$, as set out in the 'Reparametrizing the cointegrating VAR' section, is invalid, as the narrow dividend yield is not a stationary process. Nonetheless, because such reparameterizations are implicit in the existing literature, we also examine the impact of imposing further restrictions on the two systems described above that are consistent with stationary returns. Appendix B derives the matrices $\mathbf{K}$ and $\mathbf{M}$ for the two reparameterizations in terms of $\mathbf{w}_t^{\mathrm{n}}$. Table 4 panels B and C show tests of a sequence of restrictions on the joint process for returns and the narrow yield that parallel those we described in section V. All are strongly rejected.

VII. Predictive regressions for returns

A comparison of predictive regressions

The tests in the preceding two sections point very strongly to the rejection of the restrictions implied by predictive return regressions using the narrow dividend yield. By contrast, those involving the cashflow yield are easily accepted by the data, which implies that payments via dividends or non-dividends are equivalent in terms of predicting returns. This result would, of course, only be of limited interest if the 'predictive equivalence' of d_t and n_t arose simply from a combined *failure* to predict returns. Thus, we also need to examine the predictive power of the resulting regressions. As a basis for comparison, we also examine the properties of the regressions specified in terms of the narrow yield, despite the fact that the restrictions required to arrive at these specifications are so strongly rejected by the data.

Table 5 panel A provides parameter estimates for the predictive regressions for returns from our various systems specified in terms of the cashflow yield (i.e. the first row of the system in $\mathbf{w}_t^{\mathrm{c}}$). Panel B of the table provides equivalent estimates from our system in $\mathbf{w}_t^{\mathrm{n}}$, using the narrow yield. In both cases, we provide maximum-likelihood estimates from the system and single equation ordinary least square (OLS) estimates; these are computed using both the approximate and exact definitions of the log return

TABLE 5

Comparing predictive return regressions

Panel A: return regressions using cashflow yield							
	r_{t-1}	x^{c}_{t-1}	Δx^{c}_{t-1}	s^{c}_{t-1}	Δs^{c}_{t-1}	$\bar{R}^2$	*SE*
Model $C_1: \rightarrow s^{c}_{t} = \Delta d_t$							
ML_{SYS}	0.200 (1.07)	0.168 (3.33)	0.002 (0.02)	−0.437 (−2.30)	—	0.157	0.192
OLS_E	0.187 (1.20)	0.168 (3.35)	−0.002 (−0.03)	−0.427 (−2.13)	—	0.155	0.191
OLS_A	0.200 (1.29)	0.168 (3.34)	0.002 (0.03)	−0.436 (−2.18)	—	0.157	0.192
Model $C_2: \rightarrow s^{c}_{t} = d_t - \gamma v_t$							
ML_{SYS}	−0.116 (−0.70)	0.171 (2.48)	0.006 (0.06)	0.025 (0.13)	−0.445 (−2.30)	0.148	0.194
OLS_E	−0.126 (−0.79)	0.170 (2.32)	0.004 (0.04)	0.027 (0.16)	−0.446 (−2.15)	0.147	0.192
OLS_A	−0.116 (−0.73)	0.171 (2.34)	0.006 (0.07)	0.025 (0.15)	−0.445 (−2.13)	0.148	0.192
Further restrictions on models C_1 and C_2							
ML_{SYS}	0.029 (0.19)	0.199 (3.81)	−0.107 (−1.30)	—	—	0.124	0.196
OLS_E	0.020 (0.15)	0.199 (4.07)	−0.109 (−1.45)	—	—	0.122	0.195
OLS_A	0.029 (0.21)	0.199 (4.08)	−0.107 (−1.41)	—	—	0.123	0.196
ML_{SYS}	—	0.155 (3.44)	—	—	—	0.100	0.198
OLS_E	—	0.154 (3.45)	—	—	—	0.100	0.198
OLS_A	—	0.155 (3.45)	—	—	—	0.100	0.198

Panel B: return regressions using narrow yield							
Model	r_{t-1}	x^{n}_{t-1}	Δx^{n}_{t-1}	s^{n}_{t-1}	Δs^{n}_{t-1}	$\bar{R}^2$	*SE*
Model: $N_3: \Delta v_t - \delta n_t \sim I(0) \rightarrow s^{n}_{t} = \Delta n_t$							
ML_{SYS}	−0.226 (−1.07)	0.113 (1.45)	−0.433 (−2.14)	−0.101 (−1.20)	—	0.058	0.202
OLS_E	−0.240 (−1.21)	0.110 (1.66)	−0.439 (−2.47)	−0.099 (−1.12)	—	0.058	0.202
OLS_A	−0.226 (−1.14)	0.113 (1.71)	−0.433 (−2.44)	−0.101 (−1.15)	—	0.058	0.203
Model: $N_4: n_t \sim I(0) \rightarrow s^{n}_{t} = n_t$							
ML_{SYS}	−0.262 (−1.33)	0.160 (2.18)	−0.436 (−2.46)	−0.211 (−3.29)	0.027 (0.26)	0.150	0.193
OLS_E	−0.277 (−1.47)	0.157 (2.45)	−0.442 (−2.62)	−0.212 (−3.37)	0.030 (0.32)	0.151	0.192
OLS_A	−0.262 (−1.40)	0.160 (2.49)	−0.436 (−2.59)	−0.211 (−3.35)	0.027 (0.29)	0.150	0.193

TABLE 5

(continued)

Panel B: return regressions using narrow yield							
Model	r_{t-1}	x^{n}_{t-1}	Δx^{n}_{t-1}	s^{n}_{t-1}	Δs^{n}_{t-1}	$\bar{R}^2$	*SE*
Further restrictions on models N_3 and N_4							
ML_{SYS}	−0.257 (−1.31)	0.113 (1.46)	−0.418 (−2.07)	—	—	0.055	0.203
OLS_{E}	−0.271 (−1.38)	0.110 (1.66)	−0.425 (−2.40)	—	—	0.055	0.203
OLS_{A}	−0.257 (−1.31)	0.113 (1.70)	−0.418 (−2.36)	—	—	0.055	0.203
$OLS_{S\&P}$	−0.228 (−1.09)	0.087 (1.45)	−0.334 (−1.8)	—	—	0.024	0.199
ML_{SYS}	—	0.081 (1.24)	—	—	—	0.001	0.208
OLS_{E}	—	0.080 (1.32)	—	—	—	0.001	0.208
OLS_{A}	—	0.081 (1.33)	—	—	—	0.001	0.208
$OLS_{S\&P}$	—	0.059 (1.09)	—	—	—	0.002	0.201

Notes: OLS_{E}, OLS_{A}, $OLS_{S\&P}$ are OLS single-equation estimates for exact, approximated and S&P log returns. ML_{SYS} are FIML system estimates for the approximated log-return. *t*-statistics are given in parentheses.

and narrow dividend yield, which provides a cross-check on our approximations. In addition, in the case of equations specified solely in terms of the return and the narrow dividend yield (i.e. in the bottom part of panel B) we also provide estimates based on returns and yields for the S&P 500 index.

A number of points in these tables are worth noting. In section V, we showed that, with the cashflow yield x_t^{c}, the exclusion of the third variable s_t^{c} (dividend growth or the lagged cointegrating relation involving dividends) could not be rejected when imposed as a joint test against the unrestricted model of equivalent rank. We would argue that this is the appropriate test, as we have a theory-based argument for imposing all these restrictions at the same time. Nonetheless, it is of interest to note that the first two blocks of return regressions in Table 5 show that s_{t-1}^{c} (in the case of C_1) or $\Delta s_{t-1}^{\mathrm{c}}$ (in the case of C_2) is individually significant. The significance is likely to stem from the dividend growth Δd_{t-1}, that enters both terms, allowing differential short-term predictive power of dividends compared with non-dividend payments. However, given the well-known dangers of being guided solely by within-sample fit, we would not wish to place very much weight on this feature. Once any impact of s_t^{c} is excluded, the return regression specified in terms of the cashflow yield can be reduced with only modest reduction in predictive power to the simple form of equation (1), where the only predictor of returns is the lagged cashflow yield. There seems little reason to dislike this very simple predictive regression for returns on the basis of in-sample fit.

The prediction equations in terms of the narrow yield x_t^{n}, shown in panel B of Table 5, provide a marked contrast. At the outset, it should of course be borne in mind that, on the basis of the results in section VI, x_t^{n} is non-stationary and all such equations are mis-specified. In the case of model N_3, which is specified under the null that $n_t \sim I(1)$, this translates to very weak predictive power even in the most general specification (which includes a predictive role for lagged returns and the lagged 'absorbing' variable, Δn_t). Once the additional regressors are excluded, and the only predictor variable is the lagged narrow yield, predictive power for returns essentially disappears. The table shows that this feature is in no sense an artefact of our data set: very similar results are found if S&P series are used.

On the face of it, the alternative narrow yield specification in model N_4 displays distinctly more predictive power for returns when only cointegrating restrictions are imposed. However, this reveals the predictive power of the cashflow yield, rather than the conventional narrow yield. These predictive regressions include terms in n_{t-1} (because they are specified under the null that $n_t \sim I(0)$, in which case $s_t^{\mathrm{n}} = n_t$). On closer examination, it becomes evident that most of the predictive power does not stem from x_{t-1}^{n} but from n_{t-1} itself, and that the coefficients on the two terms are similar and of opposite sign. Given that $x_t^{\mathrm{n}} \approx d_t - v_t + \delta n_t$ and $x_t^{\mathrm{c}} = d_t - v_t - n_t$, this means that the equation can be reparameterized in terms of x_{t-1}^{c}. In that case, the coefficient on n_{t-1} drops to -0.04, with a t-statistic of only -0.5, thus again confirming the cashflow yield as a sufficient predictor variable for returns.

Biases and coefficient instability

A number of econometric issues that have been the subject of debate in recent research on dividend yields and return predictability appear to be closely related to the consequences of estimating equations that we have shown to be mis-specified. Two problem areas have been highlighted in past research: (1) upward bias in coefficients on the lagged dividend yield; and (2) parameter instability.

The first problem arises because, if the innovations to the predictor variable are correlated with those to returns, the coefficient on the predictor variable in the return regression is biased away from zero (Stambaugh, 1999; Lewellen, 2004; Campbell and Yogo, 2006). Campbell and Yogo note that the size of the conventional t-test for ϕ in equations of the same form as equations (1) and (2) depends on the estimated $\hat{\rho}$ and on the correlation between the innovations. They produce critical values for a pretest of the extent of size distortion. In the case of the cashflow yield, using their Table 1, we are comfortably outside the range of substantial size distortion: $\hat{\rho} = 0.59$ is sufficiently far from unity that the estimated bias in $\hat{\rho}$ itself is small, and the relatively weak link between the innovations in x_t^{c} and r_t means that only a part of this bias is transmitted to $\hat{\phi}$. Hence, the t-statistics on x_t^{c} shown in Table 5, panel A, are reliable. By contrast, for the narrow yield, we have $\hat{\rho} = 0.81$, and innovations in both r_t and x_t^{n} are dominated by innovations to Δv_t. Thus, bias correction reinforces the already very strong statistical case against x_t^{n}.

Goyal and Welch (2003) attribute the apparent significance of the narrow dividend yield in predicting returns in earlier research to a fortuitous choice of data sample. Robertson and Wright (2006) perform an extensive recursive analysis of OLS return regressions using equations (1) and (2) and the two alternative yield measures. In line with the results of Goyal and Welch,[14] the coefficient on x_t^{n} displays considerable instability, and the associated t-statistic achieves at best only marginal significance levels. By contrast, the coefficient on x_t^{c} is remarkably stable over time; its t-statistic rises steadily throughout the sample period, consistent with stability of the true coefficient, given the increase in sample size. Hence, bias and instability problems that have beset previous specifications in terms of the narrow dividend yield do not apply in the case of the cashflow yield.

VIII. Conclusions

In the introduction to this paper, we noted the divide between the very large literature finding evidence of predictability of returns, and the more recent revisionist literature that casts doubt on this evidence. Our results offer something to both sides of the debate. We share with the revisionists the conclusion that standard measures of the dividend yield have little if any predictive power for returns. On the other hand, we

[14]The dependent variable in Goyal and Welch (2003) is the excess return on stocks over bills, rather than the real return; they also use a shorter sample period, starting in 1926; however, in comparable sample periods, our results using real returns are very similar.

do not conclude that this result sounds the death knell for the predictability literature. Our proposed alternative measure, the 'cashflow' dividend yield, that treats dividends and non-dividend cashflows to shareholders as equivalent, in line with Miller and Modigliani (1961), is theoretically sound, appears stable, and has had strong and consistent predictive power for returns over a long sample.

Final Manuscript Received: November 2007

References

Ackert, L. and Smith, B. (1993). 'Stock price volatility, ordinary dividends, and other cash flows to shareholders', *Journal of Finance*, Vol. 48, pp. 1147–1160.

Bagwell, L. and Shoven, J. (1989). 'Cash distributions to shareholders', *Journal of Economic Perspectives*, Vol. 3, pp. 129–140.

Bossaerts, P. and Hillion, P. (1999). 'Implementing statistical criteria to select return forecasting models: what do we learn', *Review of Financial Studies*, Vol .12, pp. 405–428.

Boudoukh, J., Michaely, R., Richardson, M. and Roberts, M. R. (2007). 'On the importance of measuring payout yield: implications for empirical asset pricing', *Journal of Finance*, Vol. 62, pp. 877–916.

Bureau of Economic Analysis (1977). *Historical Statistics of the United States, Colonial Times to 1970*, US Census Bureau.

Campbell, J. and Shiller, R. (1988). 'The dividend–price ratio and expectations of future dividends and discount factors', *Review of Financial Studies*, Vol. 1, pp. 195–227.

Campbell, J. Y. and Shiller, R. J. (1998). 'Valuation ratios and the long-run stock market outlook', *Journal of Portfolio Management*, Vol. 28, pp. 11–26.

Campbell, J. and Yogo, M. (2006). 'Efficient tests of stock return predictability', *Journal of Financial Economics*, Vol. 81, pp. 27–60.

Cochrane, J. H. (2001). *Asset Pricing*, Princeton University Press, Princeton, NJ.

Davidson, R. and Flachaire, E. (2001). *The Wild Bootstrap: Tamed at Last*, DARP Discussion Paper no. 58, February.

Fama, E. and French, K. (1988). 'Dividend yields and expected stock returns', *Journal of Financial Economics*, Vol. 22, pp. 3–25.

Fama, E. and French, K. (2001). 'Disappearing dividends: changing firm characteristics or lower propensity to pay?', *Journal of Financial Economics*, Vol. 60, pp. 3–43.

Foster, F. D., Smith, T. and Whaley, R. (1997). 'Assessing goodness-of-fit of asset pricing models: the distribution of the maximal R^2', *Journal of Finance*, Vol. 52, pp. 591–607.

Franklin, A. and Michaely, R. (2003). 'Payout policy', Handbook of the economics of finance, *Corporate Finance*, Vol. 1A, pp. 337–429.

Goetzmann, W. N. and Jorion, P. (1993). 'Testing the predictive power of dividend yields', *Journal of Finance*, Vol. 48, pp. 663–679.

Goyal, A. and Welch, I. (2003). 'Predicting the equity premium with dividend ratios', *Management Science*, Vol. 49, pp. 639–654.

Grullon, G. and Michaely, R. (2002). 'Dividends, share repurchases and the substitution hypothesis', *Journal of Finance*, Vol. 57, pp. 1649–1684.

Jegadeesh, N. (1990). 'Evidence of predictable behavior of security returns', *Journal of Finance*, Vol. 45, pp. 881–898.

Jiang, X. and Lee, B. (2005). 'An empirical test of the accounting-based residual income model and the traditional dividend discount model', *Journal of Business*, Vol. 78, pp. 1465–1504.

Johansen, S. (1988). 'Statistical analysis of cointegration vectors', *Journal of Economic Dynamics and Control*, Vol. 12, pp. 231–254.

Johansen, S. (1991). 'Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models', *Econometrica*, Vol. 59, pp. 1551–1580.
Johansen, S. (2000). 'A Bartlett correction factor for tests on the cointegrating relations', *Econometric Theory*, Vol. 16, pp. 740–778.
Kirby, C. (1997). 'Measuring the predictable variations in stock and bond returns', *Review of Financial Studies*, Vol. 10, pp. 579–630.
Lee, B. S. and Rui, O. (2007). 'Time-series behavior of share repurchases and dividends', *Journal of Financial and Quantitative Analysis*, Vol. 42, pp. 119–142.
Lewellen, J. (2004). 'Predicting returns with financial ratios', *Journal of Financial Economics*, Vol. 74, pp. 209–235.
Liang, J. N. and Sharpe, S. A. (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.
Mehra, R. (1998). 'On the volatility of stock prices: an exercise in quantitative theory', *International Journal of Systems Science*, Vol. 29, pp. 1203–1211.
Miller, M. H. (1963). 'Stabilisation policies', edited by the Commission: *Commission on Money and Credit, Stabilization Policies*, Prentice-Hall, Englewood Cliffs, NJ.
Miller, M. H. and Modigliani, F. (1961). 'Dividend policy, growth, and the valuation of shares', *Journal of Business*, Vol. 34, pp. 411–433.
Nelson, C. R. and Kim, M. J. (1993). 'Predictable stock returns: the role of small sample bias', *Journal of Finance*, Vol. 48, pp. 641–661.
Pesaran, M. H. and Timmermann, A. F. (1995). 'Predictability of stock returns: robustness and economic significance', *Journal of Finance*, Vol. 50, pp. 1201–1228.
Robertson, D. and Wright, S. (2006). 'Dividends, total cash flow to shareholders, and predictive return regressions', *Review of Economics and Statistics*, Vol. 88, pp. 91–99.
Stambaugh, R. F. (1999). 'Predictive regressions', *Journal of Financial Economics*, Vol. 54, pp. 375–421.
Wright, S. (2004). 'Measures of stock market value and returns for the US nonfinancial corporate sector, 1900–2000', *Review of Income and Wealth*, Vol. 50, pp. 561–584.

Appendix A: Derivation of approximations for Δe_t and x_t^{n}

Using $E_t \equiv V_t/P_t$ and $I_t \equiv P_t(E_t - E_t)$, we can write

$$\frac{E_{t-1}}{E_t} = 1 - \frac{I_t}{V_t} \tag{A.1}$$

hence, using the definitions of I_t, N_t and X_t^{c},

$$\begin{aligned}\Delta e_t &= -\log\left\{1 - \frac{D_t - C_t}{V_t}\right\} = -\log\{1 - N_t X_t^{\mathrm{c}} + X_t^{\mathrm{c}}\} \\ &= -\log\{1 - [\exp(n_t) - 1]\exp(x_t^{\mathrm{c}})\} \equiv f(n_t, x_t^{\mathrm{c}}).\end{aligned}$$

Assume that $E(n_t) = 0$ and let $E(x_t^{\mathrm{c}}) = \overline{x^{\mathrm{c}}}$. Note that the existence of an unconditional mean for these variables does not imply any mean reversion. In particular, n_t could be a stationary AR(1) process without intercept or a driftless random walk with initial value $n_0 = 0$. We can then log-linearize $f(n_t, x_t^{\mathrm{c}})$ around $(0, \overline{x^{\mathrm{c}}})$, using $f(0, \overline{x^{\mathrm{c}}}) = 0$; $\partial f(0, \overline{x^{\mathrm{c}}})/\partial n_t = \exp(\overline{x^{\mathrm{c}}})$; $\partial f(0, \overline{x^{\mathrm{c}}})/\partial(x_t^{\mathrm{c}}) = 0$, giving

$$\Delta e_t \approx [\exp(\overline{x^{\mathrm{c}}})] n_t \equiv \delta n_t \tag{A.2}$$

it follows that $x_t^{\mathrm{n}} \approx d_t - v_t + \delta n_t$. Under $E(n_t)=0$, we obtain $E(x_t^{\mathrm{c}})=E(x_t^{\mathrm{n}})=\overline{x^{\mathrm{c}}}$: the yields have the same unconditional mean. If $n_t \sim I(0)$, they will also have the same order of integration. If instead, n_t, d_t and v_t are all $I(1)$, then only one of the yields can be stationary: $x^{\mathrm{c}} \sim I(0)$ and $x_t^{\mathrm{n}} \sim I(0)$ are mutually exclusive possibilities.

In the data, the sample mean of n_t is, as noted in the text, very close to zero; so, we exploit this simplified version of the approximation in our estimation and in the discussion of the main paper. However, our results in no sense hinge on this feature. More generally, for $E(n_t) \neq 0$, we can still derive a linearization around sample means of n and the cashflow yield, of the form

$$\Delta e_t \approx \overline{\Delta e} + \delta_1 (n_t - \bar{n}) + \delta_2 (x_t^{\mathrm{c}} - \overline{x^{\mathrm{c}}}) \tag{A.3}$$

where $\delta_1 = \partial f(\bar{n}, \overline{x^{\mathrm{c}}})/\partial n_t$; $\delta_2 = \partial f(\bar{n}, \overline{x^{\mathrm{c}}})/\partial x_t^{\mathrm{c}}$. Using this alternative linearization, it is straightforward to show that under the null that x_t^{n} is stationary the linear combination $d_t - v_t + \tilde{\delta} n_t$ will (up to a linear approximation) be stationary, for some alternative parameter, $\tilde{\delta}$, and an approximation for returns of the same form as equation (9) remains valid. Thus, even with the more general form of the approximation, we can still apply exactly the same techniques as in the main paper. We exploit these more general forms of the approximations in some of our bootstrapped simulations described in Appendix C.

Appendix B: Reparameterizations of the cointegrating VAR

Reparameterization to system in $\mathbf{z}_t$ for the general case

Given the definition of $\mathbf{z}_t$ in equation (14) in the main text, we can substitute for $\mathbf{y}_t$ from equation (13), and write

$$\begin{aligned}
\mathbf{z}_t &= \mathbf{K\Psi} + \left[\begin{matrix} \mathbf{0}_{n-r} \\ \mathbf{I}_r \end{matrix} + \mathbf{K}\boldsymbol{\alpha} \right] \boldsymbol{\beta}' \mathbf{y}_{t-1} + \mathbf{K}\boldsymbol{\phi}\,\Delta \mathbf{y}_{t-1} + \mathbf{K}\boldsymbol{\varepsilon}_t \\
&= \mathbf{K\Psi} + \left[\begin{matrix} \mathbf{0}_{n-r} \\ \mathbf{I}_r \end{matrix} + \mathbf{K}\boldsymbol{\alpha} \right] \boldsymbol{\beta}' \mathbf{y}_{t-1} + \gamma \left[\begin{matrix} \Delta \tilde{\mathbf{y}}_{t-1} \\ \Delta \boldsymbol{\beta}' \mathbf{y}_{t-1} \end{matrix} \right] + \mathbf{K}\boldsymbol{\varepsilon}_t \\
&= \mathbf{K\Psi} + \left[\gamma_1 \quad \left[\begin{matrix} \mathbf{0}_{n-r} \\ \mathbf{I}_r \end{matrix} + \mathbf{K}\boldsymbol{\alpha} \right] \right] \mathbf{z}_{t-1} + [\,\mathbf{0} \quad \gamma_2\,]\, \Delta z_{t-1} + \mathbf{K}\boldsymbol{\varepsilon}_t
\end{aligned} \tag{B.1}$$

where $\boldsymbol{\gamma} = \mathbf{K}\boldsymbol{\phi}\,\mathbf{K}^{-1}$ can be partitioned as $[\gamma_1 \;\; \gamma_2]$. This is of the same form as equation (16) in the main text for appropriate definitions of $\mathbf{A}_1$ and $\mathbf{A}_2$. Note that the reduced rank of $\boldsymbol{\alpha\beta}'$ in the underlying cointegrating VAR translates to a reduced rank of $\mathbf{A}_2$ in this representation. The AR(2) element in $\mathbf{y}_t$ is captured both by the combination of the terms in $\Delta \tilde{\mathbf{y}}_{t-1}$ that appear in $\mathbf{z}_{t-1}$ and by the lagged changes in the cointegrating relations that appear in $\Delta \mathbf{z}_{t-1}$.

Reparameterizations to systems in $\mathbf{w}_t^{\mathrm{c}}$ (cashflow yield as cointegrating relation)

There are two alternative specifications, depending only on the assumed rank of the cointegrating system:

$$r=1: \quad \boldsymbol{\beta}'=[-1 \quad 1 \quad -1]; \quad \mathbf{M}=\begin{bmatrix}1 & 0 & 1-\rho\\ 0 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}; \quad \mathbf{w}_t^{\mathrm{c}}=\begin{bmatrix}r_t\\ x_t^{\mathrm{c}}\\ s_t^{\mathrm{c}}\end{bmatrix}; \quad s_t^{\mathrm{c}}=\Delta d_t$$

$$r=2: \quad \boldsymbol{\beta}'=\begin{bmatrix}-1 & 1 & -1\\ -\gamma & 1 & 0\end{bmatrix}; \quad \mathbf{M}=\begin{bmatrix}1 & 1-\rho & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}; \quad \mathbf{w}_t^{\mathrm{c}}=\begin{bmatrix}r_t\\ x_t^{\mathrm{c}}\\ s_t^{\mathrm{c}}\end{bmatrix}; \quad s_t^{\mathrm{c}}=d_t-\gamma v_t.$$

Reparameterizations to systems in $\mathbf{w}_t^{\mathrm{n}}$ (narrow yield as cointegrating relation)

Reparameterization under null that $r_t \sim I(0); n_t \sim I(1)$

Under the joint null that the narrow yield is a cointegrating relation, and that returns are stationary but that n_t is a unit-root process, we need to work in terms of an underlying system of the same form as equation (13), but in which the first two elements are in per-share terms, i.e. using the approximation in equation (8) we can define

$$\mathbf{y}_t^{\mathrm{ps}} \equiv \begin{bmatrix} v_t-\delta\sum_{i=0}^{t} n_{t-i}\\ d_t-\delta\sum_{i=0}^{t} n_{t-i-1}\\ n_t\end{bmatrix} \approx \begin{bmatrix} p_t\\ d_t-e_{t-1}\\ n_t\end{bmatrix}$$

as in this representation all series are $I(1)$ under the null. Under the null, the narrow dividend yield must be stationary, and can be written either in terms of the per-share system, or in terms of the underlying VAR variables, because

$$x_t^{\mathrm{n}} \approx d_t-v_t+\delta n_t=[-1 \quad 1 \quad \delta]\,\mathbf{y}_t=[-1 \quad 1 \quad 0]\,\mathbf{y}_t^{\mathrm{ps}}. \tag{B.2}$$

Thus, under the null, exploiting equation (B.2), and ignoring irrelevant constants, the per-share system can then be written as

$$\Delta\mathbf{y}_t^{\mathrm{ps}}=\boldsymbol{\Phi}_I\Delta\mathbf{y}_{t-1}^{\mathrm{ps}}+\boldsymbol{\alpha}_I\boldsymbol{\beta}_I'\,\mathbf{y}_{t-1}+\boldsymbol{\varepsilon}_t \tag{B.3}$$

which after some algebra can also be written as

$$\mathbf{D}\Delta\mathbf{y}_t=\boldsymbol{\Phi}_I[\mathbf{I}+\delta\mathbf{J}_2\mathbf{J}_3']\Delta\mathbf{y}_{t-1}+[\,\boldsymbol{\alpha}_I \quad \mathbf{d}\,][\,\boldsymbol{\beta}_I \quad \mathbf{J}_3\,]'\,\mathbf{y}_{t-1} \tag{B.4}$$

where

$$\mathbf{D}=\begin{bmatrix}1 & 0 & -\delta\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}; \quad \mathbf{J}_2=\begin{bmatrix}0\\ 1\\ 0\end{bmatrix}; \quad \mathbf{J}_3=\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}; \quad \mathbf{d}=\delta(\mathbf{I}-\boldsymbol{\Phi}_I)\begin{bmatrix}1\\ 1\\ 0\end{bmatrix}$$

and $\mathbf{d}$ has the form $[d_1\ d_2\ 0]'$. Hence, the per-share system is nested within the structure of equation (13) with $r=2$, and

$$\begin{aligned}\boldsymbol{\alpha\beta}' &= \mathbf{D}^{-1}[\,\boldsymbol{\alpha}_I \quad \mathbf{D}\,][\,\boldsymbol{\beta}_I \quad \mathbf{J}_3\,]' \\ &= \begin{bmatrix} \alpha_{I1}+\delta\alpha_{I3} & d_1 \\ \alpha_{I2} & d_2 \\ \alpha_{I3} & 0 \end{bmatrix}\begin{bmatrix} -1 & 1 & \delta \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{B.5}$$

Note that d_1 and d_2 are not free parameters, as they depend on the estimated parameters in $\boldsymbol{\Phi}_I$. As there are no free parameters in the resulting $\boldsymbol{\beta}$ matrix (which has two free parameters in the unrestricted $r=2$ specification), there are a total of five implied coefficient restrictions. Under this null, $d_t, v_t \sim I(2)$, but are polynomially cointegrated with n_t.

To reparameterize the system in terms of $\mathbf{w}_t^{\mathrm{n}}$, we define $\mathbf{K}$ consistent with the definition of $\boldsymbol{\beta}$ in equation (B.5), and set

$$\mathbf{M} = \begin{bmatrix} 1 & 1-\rho & -\delta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{w}_t^{\mathrm{n}} = \begin{bmatrix} r_t \\ x_t^{\mathrm{n}} \\ s_t^{\mathrm{n}} \end{bmatrix}; \quad s_t^{\mathrm{n}} = \Delta n_t$$

where the third variable equals Δn_t, rather than n_t, given the zero restriction on the bottom-right element of α, such that n_t is still represented by an $I(1)$ process.

Note that when we use the more general form of the approximations for the return and the narrow yield based on equation (A.3) in Appendix A, the $\mathbf{K}$ and $\mathbf{M}$ matrices of the same general form can be derived if the cointegrating relation is defined as the linear combination $d_t - v_t + \tilde{\delta} n_t$ (a scaling of the approximated yield). These also have unit determinants, leaving the log likelihood of reparameterized systems unchanged.

Reparameterization under null that $n_t \sim I(0)$

This case is straightforward:

$$r=2: \quad \boldsymbol{\beta}' = \begin{bmatrix} -1 & 1 & \delta \\ 0 & 0 & 1 \end{bmatrix}; \ \mathbf{M} = \begin{bmatrix} 1 & 1-\rho & -\delta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \ \mathbf{w}_t^{\mathrm{c}} = \begin{bmatrix} r_t \\ x_t^{\mathrm{c}} \\ s_t^{\mathrm{c}} \end{bmatrix}; \ s_t^{\mathrm{c}} = n_t.$$

Appendix C: Bootstrapped simulations

As shown in the previous appendix, under any given null hypothesis H_0, the VAR (13) can be reparameterized into a system as in equation (18) with $\tilde{\mathbf{w}}_t = (\tilde{r}_t, \tilde{x}_t, s_t)$, where tildes denote approximated variables, $\tilde{x}_t$ is either x_t^{c} or $\tilde{x}_t^{\mathrm{n}}$, and s_t is a third null-specific variable. Our bootstrap methodology is designed to produce simulated P-values that take into account both the non-standard distribution of the residuals and the variability of the linearization parameters. We proceed as follows.

We first estimate the equivalent restricted system (18) using the exact variables $\mathbf{w}_t = (r_t, x_t, n_t)$, to which the relevant null relates. In the data, our approximations are, as noted in section III, extremely good; so, the implied systems are virtually identical to those in terms of $\tilde{\mathbf{w}}_t$, and hence have similarly non-Gaussian residuals.

A single bootstrap simulation then consists of the following steps:

- Resample the residuals, and multiply the resampled residuals by a random variable ξ_t such that $\text{Prob}(\xi_t = 1) = \text{Prob}(\xi_t = -1) = 0.5$.[15]
- Generate a bootstrapped sample of the exact $\mathbf{w}_t$ vector: $\mathbf{w}'_t = (r'_t, x'_t, s'_t)$.
- Compute the underlying $\mathbf{y}_t$ series by inverting the identities of 'The aggregate stock market return' section: $(r'_t, x'_t, s'_t) \Rightarrow (v'_t, d'_t, n'_t)$.
- Calculate approximated variables $\tilde{\mathbf{w}}_t = (\tilde{r}_t, \tilde{x}_t, s_t)$ using linearization parameters derived from (r'_t, x'_t, s'_t).
- Obtain the likelihood ratio statistic by comparing the VECM (13) in $\mathbf{y}'_t = (v'_t, d'_t, s'_t)$ to a system (18) in $\mathbf{w}'_t = (\tilde{r}'_t, \tilde{x}'_t, s'_t)$.

In the case of nulls relating to the cashflow yield, the only approximation used is the return approximation (10). For nulls relating to the conventional 'narrow' yield, we also require an approximation for the yield itself, which in turn derives from the approximation for Δe_t derived in Appendix A. The simple form of the approximation in equation (A.2), and hence in equation (11) is derived on the assumption that $\bar{n} = 0$. While this is (conveniently) consistent with the sample mean in the historic data set, in a significant number of replications the sample mean of n_t is a long way from zero (this is unsurprisingly especially the case when $n_t \sim I(1)$ under the null); so, in bootstrapping P-values relating to these nulls, we use the more general forms of the approximations for both the 'narrow' yield and the return, as described in Appendix A, in both the historic and simulated data. In the historic data set, the resulting approximations are virtually identical to those using the simpler linearization.

[15]Davidson and Flachaire (2001) demonstrate the optimal properties of this version of the 'wild bootstrap'.