Stock Return Predictability: Is it There?

Andrew Ang
Columbia University and NBER

Geert Bekaert
Columbia University, NBER and CPER

We examine the predictive power of the dividend yields for forecasting excess returns, cash flows, and interest rates. Dividend yields predict excess returns only at short horizons together with the short rate and do not have any long-horizon predictive power. At short horizons, the short rate strongly negatively predicts returns. These results are robust in international data and are not due to lack of power. A present value model that matches the data shows that discount rate and short rate movements play a large role in explaining the variation in dividend yields. Finally, we find that earnings yields significantly predict future cash flows. (*JEL* C12, C51, C52, E49, F30, G12)

In a rational no-bubble model, the price-dividend ratio is the expected value of future cash flows discounted with time-varying discount rates. Because price-dividend ratios, or dividend yields, vary over time, dividend yield variability can be attributed to the variation of expected cash flow growth, expected future risk-free rates, or risk premia. The "conventional wisdom" in the literature (see, among others, Campbell, 1991; Cochrane, 1992) is that aggregate dividend yields strongly predict excess returns, and the predictability is stronger at longer horizons.[1] Since dividend yields only weakly predict dividend growth, conventional wisdom attributes most of the variation of dividend yields to changing forecasts of expected returns. We critically and comprehensively re-examine this conventional wisdom regarding return predictability on the aggregate market.

We thank Xiaoyan Zhang for help with data. We thank Kobi Boudoukh, Michael Brandt, John Campbell, John Cochrane, George Constantinides, Cam Harvey, Bob Hodrick, Charles Jones, Owen Lamont, Martin Lettau, Jun Liu, Sydney Ludvigson, Chris Neely, Jim Poterba, Tano Santos, Bob Shiller, Tim Simin, Ken Singleton, Rob Stambaugh, Meir Statman, Jim Stock, Ane Tamayo, Samuel Thomson, Ivo Welch, Jeff Wurgler, and seminar participants at Columbia University, Duke University, INSEAD, London Business School, the NY Federal Reserve, NYU, Stanford University, UC San Diego, Wharton, Yale, the American Finance Association, an NBER Asset Pricing meeting, the European Finance Association, and the Western Finance Association for helpful comments. We also thank three anonymous referees and Yacine Aït-Sahalia (the editor) for valuable, extensive comments that greatly improved the article. Andrew Ang thanks the Chazen Institute at Columbia Business School for financial support. Geert Bekaert thanks the NSF for financial support. Address correspondence to Geert Bekaert, Columbia Business school, Room 802, Uris Hall, 3022 Broadway, New York, NY 10027, or e-mail: gb241@columbia.edu.

[1] Among those examining the predictive power of the dividend yield on excess stock returns are Fama and French (1988), Campbell and Shiller (1988a,b), Goetzmann and Jorion (1993, 1995), Hodrick (1992), Stambaugh (1999), Wolf (2000), Goyal and Welch (2003, 2004), Engstrom (2003), Valkanov (2003), Lewellen (2004), and Campbell and Yogo (2006).

doi:10.1093/rfs/hhl021 Advance Access publication July 6, 2006

Our main findings can be summarized as follows. First, the statistical inference at long horizons critically depends on the choice of standard errors. With the standard Hansen–Hodrick (1980) or Newey–West (1987) standard errors, there is some evidence for long horizon predictability, but it disappears when we correct for heteroskedasticity and remove the moving average structure in the error terms induced by summing returns over long horizons (Richardson and Smith 1991, Hodrick 1992, Boudoukh and Richardson, 1993).

Second, we find that the most robust predictive variable for future excess returns is the short rate, but it is significant only at short horizons.[2] Whereas, the dividend yield does not univariately predict excess returns, the predictive ability of the dividend yield is considerably enhanced, at short horizons, in a bivariate regression with the short rate. To mitigate data snooping concerns (Lo and MacKinlay 1990, Bossaerts and Hillion 1999, Ferson, Sarkissian, and Simin 2003, Goyal and Welch 2004), we confirm and strengthen this evidence using three other countries: the United Kingdom, France, and Germany.

Third, the dividend yield's predictive power to forecast future dividend growth is not robust across sample periods or countries. We find that high dividend yields are associated with high future interest rates. While the statistical evidence for interest rate predictability is weak, the same positive relationship is implied by an economic model, and we observe the same patterns across countries.

To help interpret our findings and to deepen our understanding of the data, we provide additional economic analysis. First, we build a nonlinear present value model with stochastic discount rates, short rates, and dividend growth, that matches our evidence on excess return predictability. As is true in the data, the model implies that the dividend yield only weakly predicts future cash flows but is positively related to future movements in interest rates. While excess discount rates still dominate the variation in price-dividend ratios, accounting for 61%, of the variation short rate movements account for up to 22% of the variation. In comparison, dividend growth accounts for around 7% of the variance of price-dividend ratios. The rest of the variation is accounted for by covariance terms.

Because many studies, particularly in the portfolio choice literature, use univariate dividend yield regressions to compute expected returns (Campbell and Viceira, 1999), we use the nonlinear present value model to examine the fit of regression-based expected returns with true expected returns. Consistent with the data, we find that a univariate dividend yield

[2] Authors examining the predictability of excess stock returns by the nominal interest rate include Fama and Schwert (1977), Campbell (1987), Breen, Glosten, and Jagannathan (1989), Shiller and Beltratti (1992), and Lee (1992).

regression provides a rather poor proxy to true expected returns. However, using both the short rate and dividend yield considerably improves the fit, especially at short horizons.

Second, using the present value model, we show that long-horizon statistical inference with the standard Hansen–Hodrick (1980) or Newey–West (1987) standard errors is treacherous. We find that both Hansen–Hodrick and Newey–West standard errors lead to severe over-rejections of the null hypothesis of no predictability at long horizons but that standard errors developed by Hodrick (1992) retain the correct size in small samples. The power of Hodrick *t*-statistics exceeds 0.60 for a 5% test for our longest sample. Moreover, when we pool data across different countries, the power for our shortest sample increases to 74%. Hence, lack of power is unlikely to explain our results.[3]

Finally, we focus on expanding the information set to obtain a potentially better estimate of true value-relevant cash flows in the future. Dividends may be potentially poor instruments because dividends are often manipulated or smoothed. Bansal and Lundblad (2002) and Bansal and Yaron (2004) argue that dividend growth itself follows an intricate ARMA process. Consequently, it is conceivable that more than one factor drives the dynamics of cash flows. One obvious way to increase the information set is to use earnings. Lamont (1998) argues that the earnings yield has independent forecasting power for excess stock returns in addition to the dividend yield. When we examine the predictive power of the earnings yield for both returns and cash flows, we find only weak evidence for Lamont's excess return predictability results. However, we detect significant predictability of future cash flows by earnings yields.

This article is organized as follows. Section 1 describes the data. Section 2 contains the main predictability results for returns, while Section 3 discusses cash flow and interest rate predictability by the dividend yield. Section 4 develops a present value model under the null and various alternative models to interpret the empirical results. In Section 5, we conduct a size and power analysis of Hodrick (1992) standard errors. Section 6 investigates the predictive power of the earnings yield for excess returns and cash flows. Section 7 concludes and briefly discusses a number of contemporaneous papers on stock return predictability. It appears that the literature is converging to a new consensus, substantially different from the old view.

[3] Given the excellent performance of Hodrick (1992) standard errors, we do not rely on the alternative inference techniques that use unit root, or local-to-unity, data generating processes (see, among others, Richardson and Stock 1989, Richardson and Smith 1991, Elliot and Stock 1994, Lewellen 2004, Torous, Valkanov, and Yan 2004, Campbell and Yogo 2006, Polk, Thompson, and Vuolteenaho 2006, Jansson and Moreira 2006). One major advantage of Hodrick standard errors is that the set up can handle multiple regressors, whereas the inference with unit root type processes relies almost exclusively on univariate regressors. The tests for multivariate predictive regressions using local-to-unity data generating processes developed by Polk et al. (2006) involve computationally intensive bootstrapping procedures. This test also has very poor size properties under the nonlinear present value we present in Section 4. These results are available upon request.

1. Data

We work with two data sets, a long data set for the United States, United Kingdom and Germany and a shorter data set for a sample of four countries (United States, United Kingdom, France, and Germany). In the data, dividend and earnings yields are constructed using dividends and earnings summed up over the past year. Monthly or quarterly frequency dividends and earnings are impossible to use because they are dominated by seasonal components.

We construct dividend growth and earnings growth from these ratios, producing rates of annual dividend or earnings growth over the course of a month or a quarter. To illustrate this construction, suppose we take the frequency of our data to be quarterly. We denote log dividend growth at a quarterly frequency as $g_t^{d,4}$, with the superscript 4 to denote that it is constructed using dividends summed up over the past year (four quarters). We compute $g_t^{d,4}$ from dividend yields D_t^4/P_t, where the dividends are summed over the past year, using the relation

$$g_t^{d,4} = \log\left(\frac{D_t^4/P_t}{D_{t-1}^4/P_{t-1}} \times \frac{P_t}{P_{t-1}}\right), \tag{1}$$

where $D_t^4 = D_t + D_{t-1} + D_{t-2} + D_{t-3}$ represents dividends summed over the past year and P_t/P_{t-1} is the price return over the past quarter.

In our data, the long sample is at a quarterly frequency and the short sample is at a monthly frequency. In the case of the monthly frequency, we append dividend yields, earnings yields, dividend growth and earnings growth with a superscript of 12 to indicate that dividends and earnings have been summed over the past 12 months. We also denote log dividend yields by lower case letters. Hence $dy_t^4 = \log(D_t^4/P_t)$, in the case of quarterly data and $dy_t^{12} = \log(D_t^{12}/P_t)$ in the case of monthly data. We also use similar definitions for log earnings yields: ey_t^4 and ey_t^{12}.

1.1 Long sample data

Our US data consists of price return (capital gain only), total return (capital gain plus dividend), and dividend and earnings yields on the Standard & Poor's Composite Index from June 1935 to December 2001. This data is obtained from the *Security Price Index Record*, published by Standard & Poor's Statistical Service. Lamont (1998) uses the same data set over a shorter period. The long-sample UK data comprises price returns and total returns on the Financial Times (FT) Actuaries Index, and we construct implied dividend yields from these series. For our German data, we take price returns, total returns, and dividend yields on the composite DAX (CDAX) index from the Deutsche Borsche. The long-sample UK and Ge-rman data span June 1953 to December 2001 and were purchased from Global Financial Data. All the long sample data for the United States, United Kingdom, and Germany are at the quarterly frequency, and we consequently use three-month T-bills as quarterly short rates.

Panel A of Table 1 lists summary statistics. US earnings growth is almost as variable as returns, whereas the volatility of dividend growth is less than half the return volatility. The variability of UK and German dividend growth rates is of the same order of magnitude as that of returns. The instruments (short rates, dividend and earnings yields) are all highly persistent. Because the persistence of these instruments plays a crucial role in the finite sample performance of predictability test statistics, we report test statistics under the null of a unit root and a stationary process in Panel A. Investigating both null hypotheses is important because unit root tests have very low power to reject the null of a stationary, but persistent, process.

In the United Kingdom and Germany, dividend yields are unambiguously stationary, as we reject the null of a unit root and fail to reject the null of stationarity at the 5% level. For the US dividend yield, the evidence for non-stationarity is weak as we fail to reject either hypothesis. This is surprising because the trend toward low dividend yields in the 1990s has received much attention. Figure 1 plots dividend yields for the

Table 1
Sample moments, unit root, and stationarity tests

	Excess return	Short rate	Dividend yield	Earnings yield	Dividend growth	Earnings growth
Panel A: Long-sample data						
US S&P Data, June 1935–December 2001						
Mean	0.0749	0.0409	0.0403	0.0768	0.0532	0.0548
Stdev	0.1684	0.0317	0.0150	0.0297	0.0658	0.1572
Auto	0.1173	0.9548	0.9504	0.9517	0.4071	0.3832
Test statistics						
H_0: unit root	−14.50**	−2.194	−1.187	−1.183	−10.83**	−10.55**
H_0: stationary	0.073	0.635*	0.372	0.336	0.035	0.026
UK FT Data, June 1953–December 2001						
Mean	0.0563	0.0751	0.0478		0.0670	
Stdev	0.1938	0.0331	0.0131		0.1866	
Auto	0.0907	0.9400	0.8290		−0.0486	
Test statistics						
H_0: unit root	−12.66**	−2.559	−4.125**		−14.64**	
H_0: stationary	0.037	0.637*	0.199		0.068	
Germany DAX Data, June 1953–December 2001						
Mean	0.0577	0.0467	0.0287		0.0788	
Stdev	0.1921	0.0198	0.0090		0.2086	
Auto	0.0851	0.9376	0.9087		0.1136	
Test statistics						
H_0: unit root	−12.89**	−3.036*	−3.336*		−12.34**	
H_0: stationary	0.091	0.313	0.328		0.156	

Table 1
(continued)

	Excess return	Short rate	Dividend yield	Earnings yield	Dividend growth	Earnings growth
Correlations of excess returns, June 1953–December 2001						
	United States	United Kingdom				
United Kingdom	0.6281					
Germany	0.5118	0.4598				
Panel B: MSCI data						
February 1975–December 2001						
United States						
Mean	0.0576	0.0745	0.0353	0.0744	0.0529	0.0501
Stdev	0.1513	0.0341	0.0143	0.0305	0.0589	0.0883
Auto	−0.0044	0.9675	0.9892	0.9867	−0.3187	0.1490
United Kingdom						
Mean	0.0604	0.0989	0.0456	0.0874	0.0812	0.0651
Stdev	0.1824	0.0355	0.0123	0.0346	0.0707	0.0973
Auto	−0.0184	0.9615	0.9701	0.9758	−0.0524	0.2346
France						
Mean	0.0542	0.0906	0.0415	0.0675	0.0782	0.0690
Stdev	0.2079	0.0476	0.0188	0.0397	0.0849	0.5868
Auto	0.0745	0.8741	0.9849	0.9627	−0.0068	−0.0765
Germany						
Mean	0.0498	0.0563	0.0359	0.0688	0.0643	0.0564
Stdev	0.1925	0.0241	0.0125	0.0287	0.0876	0.2217
Auto	0.0665	0.9764	0.9860	0.9836	0.0936	0.1781
Correlations of excess returns						
	United States	United Kingdom	France			
United Kingdom	0.5960					
France	0.5237	0.5184				
Germany	0.4951	0.4742	0.6178			

Panel A reports summary statistics of long-sample data for the United States, United Kingdom, and Germany, all at a quarterly frequency. Panel B reports statistics for monthly frequency Morgan Stanley Capital International MSCI data. Excess returns and short rates are continuously compounded. Sample means and standard deviations (Stdev) for excess returns, dividend, and earnings growth have been annualized by multiplying by 4 (12) and $\sqrt{4}$ ($\sqrt{12}$), respectively, for the case of quarterly (monthly) frequency data. Short rates for the long-sample (MSCI) data are three-month T-bill returns (one month EURO rates). Dividend and earnings yields, and the corresponding dividend and earnings growth are computed using dividends or earnings summed up over the past year. In Panel A, the unit root test is the Phillips and Perron (1988) test for the estimated regression $x_t = \alpha + \rho x_{t-1} + u_t$ under the null $x_t = x_{t-1} + u_t$. The critical values corresponding to p-values of 0.01, 0.025, 0.05, and 0.10 are −3.46, −3.14, −2.88, and −2.57, respectively. The test for stationarity is the Kwiatkowski *et al.* (1992) test. The critical values corresponding to p-values of 0.01, 0.025, 0.05, and 0.10 are 0.739, 0.574, 0.463, 0.347, respectively.
*p<0.05.
**p<0.01.

United States, United Kingdom, and Germany. For the United Kingdom, the dividend yield also declined during the late 1990s, but the United Kingdom experienced similar low level dividend yields during the late 1960s and early 1970s. For Germany, there is absolutely no trend in the dividend yield. If a time trend in dividend yields is a concern for

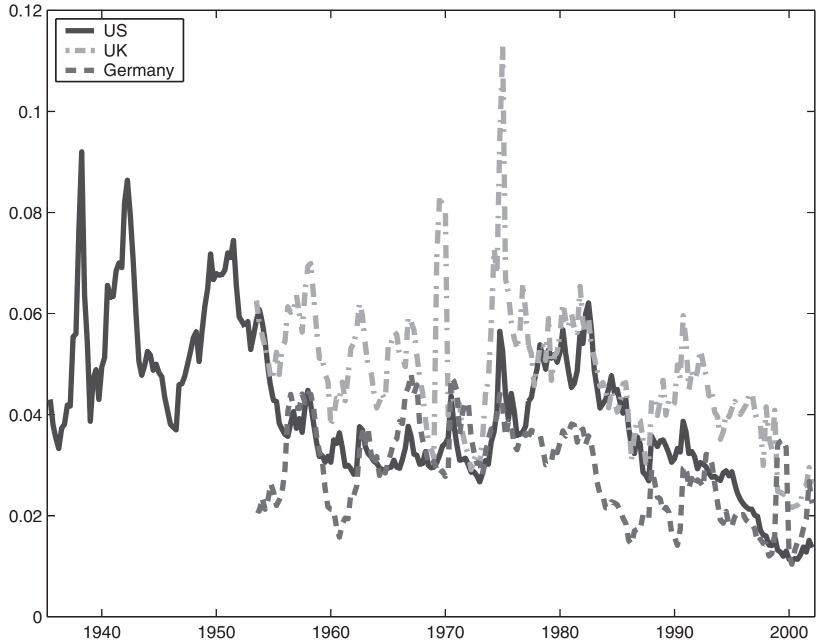

Figure 1
Dividend yields over the long sample
We plot dividend yields from June 1935 to December 2001 for the United States and from March 1953 to December 2001 for the United Kingdom and Germany.

interpreting the evidence on excess return predictability using the dividend yield, international data are clearly helpful. Present value models which impose transversality also imply that dividend yields must be stationary.

Interest rates are also highly persistent variables. While German interest rates appear to be stationary, there is some evidence of borderline non-stationary behavior for both US and UK interest rates. Most economic models also imply that interest rates are stationary (Clarida, Galí, and Gertler, 1999). Our present value model incorporates realistic persistence in short rates, but because of the high persistence of the short rate, we check the robustness of interest rate predictability by using a detrended short rate.

1.2 Short sample MSCI data

The data for the United States, United Kingdom, France, and Germany consist of monthly frequency price indices (capital appreciation only), total return indices (including income), and valuation ratios from Morgan Stanley Capital International (MSCI) in local currency, from February

1975 to December 2001. We use the one-month EURO rate from Datastream as the short rate.

Panel B of Table 1 shows that the United States has the least variable stock returns with the least variable cash flow growth rates. The extreme variability of French earnings growth rates is primarily due to a few outliers between May 1983 and May 1984, when there are very large movements in price-earnings ratios. Without these outliers, the French earnings growth variability drops to 33%. The variability of short rates, dividend, and earnings yields is similar across countries. The equity premium for the United States, France, Germany, and the United Kingdom roughly lies between 4 and 6% during this sample period. Dividend yields and short rates are again very persistent over the post-1975 sample. We also report excess return correlations showing that correlations range between 0.47 and 0.60. The correlations for the United States, United Kingdom, and Germany are similar to the correlations over the post-1953 period reported in Panel A.

2. The Predictability of Equity Returns

2.1 Predictability regressions

Denote the gross return on equity by $Y_{t+1} = (P_{t+1} + D_{t+1})/P_t$ and the continuously compounded return by $y_{t+1} = \log(Y_{t+1})$. The main regression we consider is

$$\tilde{y}_{t+k} = \alpha_k + \beta_k' z_t + \varepsilon_{t+k,k}, \tag{2}$$

where $\tilde{y}_{t+k} = (\tau/k)[(y_{t+1} - r_t) + ... + (y_{t+k} - r_{t+k-1})]$ is the annualized k-period excess return for the aggregate stock market, r_t is the risk-free rate from t to $t+1$, and $y_{t+1} - r_t$ is the excess one period return from time t to $t+1$. A period is either a month ($\tau = 12$) or a quarter ($\tau = 4$). All returns are continuously compounded. The error term $\varepsilon_{t+k,k}$ follows a $MA(k-1)$ process under the null of no predictability ($\beta_k = 0$) because of overlapping observations. We use log dividend yields and annualized continuously compounder risk-free rates as instruments in z_t.

We estimate the regression (2) by OLS and compute standard errors of the parameters $\theta = (\alpha\ \beta_k')'$ following Hodrick (1992). Using generalized method of moments, (GMM) θ has an asymptotic distribution $\sqrt{T}(\hat{\theta} - \theta) \overset{a}{\sim} N(0,\Omega)$ where $\Omega = Z_0^{-1} S_0 Z_0^{-1}$, $Z_0 = \mathrm{E}(x_t x_t')$, and $x_t = (1\ z_t')'$. Hodrick exploits covariance stationarity to remove the overlapping nature of the error terms in the standard error computation. Instead of summing $\varepsilon_{t+k,k}$ into the future to obtain an estimate of S_0, Hodrick sums $x_t x_{t-j}'$ into the past and estimates S_0 by

$$\hat{S}_0 = \frac{1}{T}\sum_{t=k}^{T} wk_t wk_t', \tag{3}$$

where

$$wk_t = \varepsilon_{t+1,1}\left(\sum_{i=0}^{k-1} x_{t-i}\right).$$

We show in section 5 that the performance of Hodrick (1992) standard errors is far superior to the Newey–West (1987) standard errors or the robust GMM generalization of Hansen and Hodrick (1980) standard errors (see Appendix A) typically run in the literature. Hence, our predictability evidence exclusively focuses on Hodrick t-statistics. Mindful of Richardson's (1993) critique of focusing predictability tests on only one particular horizon k, we also compute joint tests across horizons. For the quarterly (monthly) frequency data, we test for predictability jointly across horizons of 1, 4, and 20 quarters (1, 12, and 60 months). Appendix B details the construction of joint tests across horizons accommodating Hodrick standard errors. Finally, when considering predictability in multiple countries, we estimate pooled coefficients across countries and provide joint tests of the null of no predictability. Pooled estimations mitigate the data-mining problem plaguing US data and, under the null of no predictability, enhance efficiency because the correlations of returns across countries are not very high (Table 1). Appendix C details the econometrics underlying the pooled estimations.

2.2 Return predictability in the United States

We report results for several sample periods, in addition to the full sample 1935–2001. Interest rate data are hard to interpret before the 1951 Treasury Accord, as the Federal Reserve pegged interest rates during the 1930s and the 1940s. Hence, we examine the post-Accord period, starting in 1952. Second, the majority of studies establishing strong evidence of predictability use data before or up to the early 1990s. Studies by Lettau and Ludvigson (2001) and Goyal and Welch (2003) point out that predictability by the dividend yield is not robust to the addition of the 1990s decade. Hence, we separately consider the effect of adding the 1990s to the sample.

We start by focusing on a univariate regression with the dividend yield as the regressor. Figure 2 shows the slope coefficients for three different sample periods, using the quarterly US S&P data. The left-hand column reports the dividend yield coefficients, whereas the right-hand column reports t-statistics computed using Newey–West (1987), robust Hansen–Hodrick (1980), and Hodrick (1992) standard errors. For the

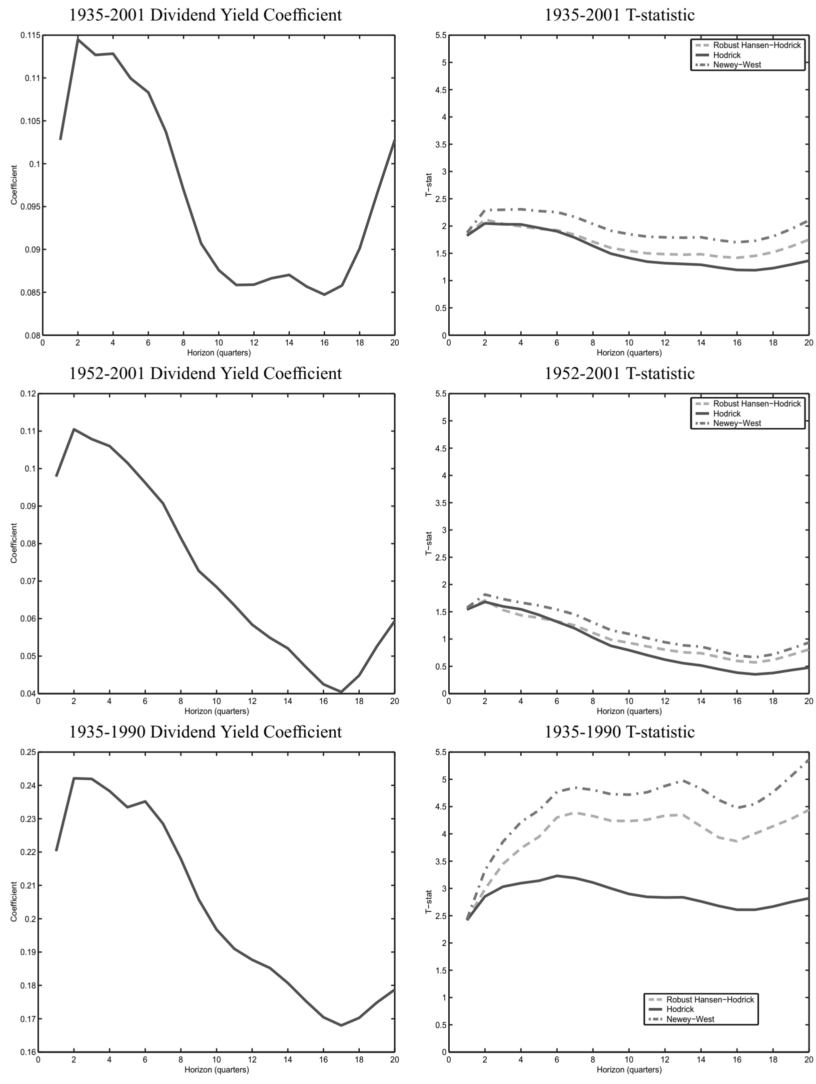

Figure 2
Dividend yield coefficients and *t*-statistics from US regressions
The left (right) column shows the dividend yield coefficients β_k (t-statistics) in the regression $\tilde{y}_{t+k} = \alpha + \beta_k dy_t^4 + \varepsilon_{t+k,k}$, where $\tilde{y}_{t+k}$ is the cumulated and annualized k-quarter ahead excess return and dy_t^4 is the log dividend yield. T-statistics are computed using Robust Hansen-Hodrick (1980), Hodrick (1992) or Newey-West (1987) standard errors. The quarterly data is from Standard and Poors.

Newey–West errors, we use $k + 1$ lags. The coefficient pattern is similar across the three periods, but the coefficients are twice as large for the period omitting the 1990s from the sample. For the other two periods (1935–2001 and 1952–2001), the one-period coefficient is about 0.110, rises until the one-year horizon, and then decreases, before increasing again near 20 quarters.

Over 1935–2001, the Hodrick t-statistic is above 2 only for horizons 2–4 quarters. However, there is no evidence of short-run predictability (at the one-quarter horizon) or long-horizon predictability. We draw a very different picture of predictability if we use Newey–West or robust Hansen–Hodrick t-statistics, which are almost uniformly higher than Hodrick t-statistics. Using Newey–West standard errors, the evidence in favor of predictability would extend to eight quarters for the full sample. Over the 1952–2001 sample, there is no evidence of predictability, whereas for the 1935–1990 period, the evidence for predictability is very strong, whatever the horizon, with all three t-statistics being above 2.4.

Table 2 summarizes the excess return predictability results for horizons of one month (quarter), one year, and five years. We only report t-statistics using Hodrick standard errors. In addition to the sample periods shown in Figure 2, we also show the 1952–1990 period, which is close to the 1947–1994 sample period in Lamont (1998). When we omit the 1990s, we confirm the standard results found by Campbell and Shiller (1988a,b) and others: the dividend yield is a significant predictor of excess

Table 2
Predictability of US excess returns

		Univariate regression	Bivariate regression		
	k-mths	dy^{12}	r	dy^{12}	χ^2 test
Panel A: quarterly S&P data					
1935–2001	1	0.1028	−1.0888	0.0857	0.070
		(1.824)	(−1.608)	(1.530)	
	4	0.1128	−0.5596	0.1032	0.104
		(2.030)*	(−0.827)	(1.865)	
	20	0.1028	−0.3187	0.0952	0.323
		(1.364)	(−0.598)	(1.255)	
1952–2001	1	0.0979	−2.1623	0.1362	0.003**
		(1.541)	(−2.912)**	(2.152)*	
	4	0.1060	−1.4433	0.1313	0.041*
		(1.546)	(−1.930)	(1.921)	
	20	0.0594	−0.4829	0.0774	0.714
		(0.477)	(−0.745)	(0.600)	
1935–1990	1	0.2203	−1.0380	0.1917	0.027*
		(2.416)*	(−1.543)	(2.126)*	
	4	0.2383	−0.4865	0.2254	0.008**
		(3.097)**	(−0.714)	(3.006)**	
	20	0.1787	−0.3229	0.1719	0.017*
		(2.819)**	(−0.569)	(2.832)**	
1952–1990	1	0.2962	−2.7329	0.4125	0.002**
		(2.783)**	(−3.504)**	(3.672)**	
	4	0.3070	−1.9840	0.3935	0.000**
		(3.000)**	(−2.508)*	(3.700)**	
	20	0.1689	−0.7120	0.2057	0.052
		(1.916)	(−1.087)	(2.387)*	

Table 2
(continued)

	k-mths	Univariate regression	Bivariate regression		
		dy^{12}	r	dy^{12}	χ^2 test
Panel B: monthly MSCI data					
1975–2001	1	0.0274	−2.4358	0.1364	0.057
		(0.405)	(−2.388)*	(1.626)	
	12	0.0106	−1.2470	0.0669	0.395
		(0.141)	(−1.361)	(0.744)	
	60	−0.0884	0.3451	−0.1207	0.857
		(−0.475)	(0.238)	(−0.397)	

We estimate regressions of the form $\tilde{y}_{t+k} = \alpha_k + z_t^{'}\beta + \epsilon_{t+k,k}$ where $\tilde{y}_{t+k}$ is the cumulated and annualized k-period ahead excess return, with instruments z_t being log dividend yields or risk-free rates and log dividend yields together. T-statistics in parentheses are computed using Hodrick (1992) standard errors. For Panel A (B), horizons k are quarterly (monthly). The χ^2 test column reports a p-value for a test that both the risk-free rate and log dividend yield coefficients are jointly equal to zero.
*p<0.05.
**p<0.01.

returns at all horizons. However, when we use all the data, we only find 5% significance at the one-year horizon for the longest sample. A test for predictability by the dividend yield jointly across horizons rejects with a p-value of 0.014 for the 1935–1990 period, but fails to reject with a p-value of 0.587 over the whole sample, even though the shorter horizon t-statistics all exceed 1.5 in absolute value. While it is tempting to blame the bull market of the 1990s for the results, our data extend until the end of 2001 and hence incorporate a part of the bear market that followed. For the 1975–2001 sample, reported in Panel B, the dividend yield also fails to predict excess returns.

Table 2 also reports bivariate regression results with the short rate as an additional regressor. For the post-Treasury Accord 1952–2001 sample, a 1% increase in the annualized short rate decreases the equity premium by about 2.16%. The effect is significant at the 1% level. A joint test on the interest rate coefficients across horizons rejects strongly for both the 1952–2001 period (p-value = 0.004) and the 1952–1990 period (p-value = 0.000). The predictive power of the short rate dissipates quickly for longer horizons but remains borderline significant at the 5% level at the one-year horizon.[4] If expected excess returns are related only to short rates and short rates follow a univariate autoregressive process, the persistence of the interest rate (0.955 in Table 1) implies that the coefficient on the short rate should tend to zero slowly for long horizons. In fact, the decay rate should be $1/k \times (1-\rho^k)/(1-\rho)$ for horizon k. The

[4] The results do not change when a detrended short rate is used instead of the level of the short rate or when we use a dummy variable over the period from October 1979 to October 1982 to account for the monetary targeting period.

decay rate in data is clearly more rapid, indicating that either expected excess returns or risk-free rates, or both, are multifactor processes.

In the bivariate regression, the dividend yield coefficient is only significant at the 5% level for the one-quarter horizon. Joint tests reject at the 1% (5%) level for the one-quarter (four-quarter) horizon but fail to reject at long horizons. When we omit the 1990s, the predictive power of the short rate becomes even stronger. A joint predictability test still fails to reject the null of no predictability at long horizons, but the *p*-value is borderline significant (0.052). Over the 1975–2001 sample, the coefficient on the short rate remains remarkably robust and is significant at the 5% level. While the coefficient on the dividend yield is no longer significantly different from zero, it is similar in magnitude to the full sample coefficient and a joint test is borderline significant (p-value $= 0.057$). The Richardson (1993) joint predictability test over all horizons and both predictors rejects at the 1% level in the samples excluding the 1990s and the full sample, rejects at the 5% level for 1952–2001, and rejects at the 10% level for 1975–2001.

Looking at the 1951–2001 and 1975–2001 samples, the evidence for the bivariate regression at short horizons is remarkably consistent. Moreover, the coefficient on the dividend yield is larger in the bivariate regression than in the univariate regression. This suggests that the univariate regression suffers from an omitted variable bias that lowers the marginal impact of dividend yields on expected excess returns. Engstrom (2003), Menzly, Santos, and Veronesi (2004), and Lettau and Ludvigson (2005) also note that a univariate dividend yield regression may understate the dividend yield's ability to forecast returns.

2.3 Predictability of excess returns in four countries

The weak predictive power of the univariate dividend yield in the full sample may simply be a small sample phenomenon due to the very special nature of the last decade for the US stock market. Alternatively, the conventional wisdom of strong long-horizon excess return predictability by dividend yields before 1990 may be a statistical fluke. International evidence can help us to sort out these two interpretations of the data and check the robustness of predictability patterns observed in US data.

Figure 3 displays the univariate dividend yield coefficients and their *t*-statistics using Hodrick standard errors in the 1975–2001 sample. First, none of the patterns in other countries resembles the US pattern. For France and Germany, and to a lesser degree for the United Kingdom, the coefficients first increase with horizon, then decrease, and finally increase again. This is roughly the pattern we see in US data for the longer samples. However, for France and Germany, the coefficients are small at short horizons and are negative for many horizons. They are also never statistically significant. The UK coefficient is larger and remains positive across horizons: it is also significantly different from zero at the

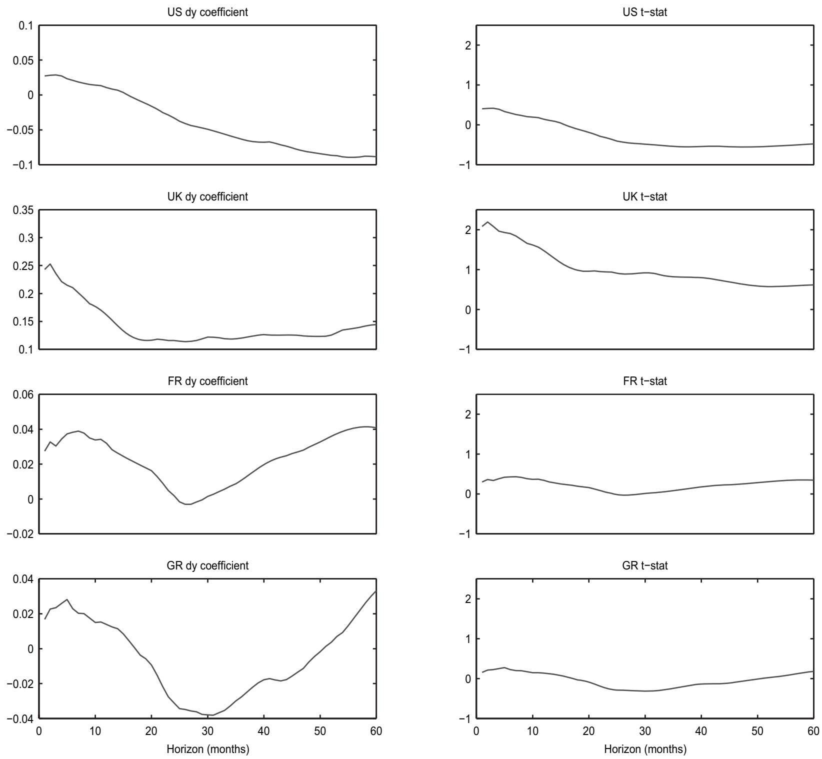

Figure 3
Dividend yield coefficients and *t*-statistics in four countries
The left (right) column shows the dividend yield coefficients β_k (t-statistics) in the regression $\tilde{y}_{t+k} = \alpha + \beta_k dy_t^{12} + \varepsilon_{t+k,k}$, where $\tilde{y}_{t+k}$ is the cumulated and annualized k-month ahead excess return and dy_t^{12} is the log dividend yield. T-statistics are computed using Hodrick (1992) standard errors. The monthly data is from MSCI and the sample period is from 1975 to 2001.

very shortest horizons. These results are opposite to the results in a recent study by Campbell (2003), who reports strong long-horizon predictability for France, Germany, and the United Kingdom over similar sample periods. We find that Campbell's conclusions derive from the use of Newey–West (1987) standard errors, and the predictability disappears when Hodrick (1992) standard errors are employed.

For the United Kingdom and Germany, we also investigate the longer 1953–2001 sample in Panel A of Table 3. The first column reports the univariate dividend yield coefficients. We only find significance at the one-year horizon for the UK, but the coefficients are all positive and more than twice as large as the US coefficients.[5] Germany's dividend yield

[5] For the United Kingdom, we also looked at a sample spanning 1935–2001, where we find a significant univariate dividend yield coefficient at the five-year horizon but not at the one-quarter horizon.

Table 3
Excess return regressions across countries

		Univariate regression		Bivariate regression			
	k-qtrs	dy^4 only	J-test	r	dy^4	χ^2 Test	J-test
Panel A: quarterly long-sample data							
United Kingdom, 1953–2001	1	0.2280		−1.2148	0.2596	0.069	
		(1.449)		(−1.500)	(1.675)		
	4	0.2600		−0.7138	0.2777	0.041*	
		(2.075)*		(−0.896)	(2.273)*		
	20	0.1290		0.1075	0.1264	0.326	
		(1.403)		(0.182)	(1.338)		
Germany, 1953–2001	1	0.0740		−3.4079	0.1083	0.030*	
		(0.827)		(−2.545)*	(1.204)		
	4	0.1235		−2.1027	0.1443	0.094	
		(1.366)		(−1.679)	(1.610)		
	20	0.0415		0.4758	0.0372	0.814	
		(0.397)		(0.489)	(0.360)		
Pooled United States,	1	0.1230	0.496	−1.958	0.1600	0.001**	0.133
United Kingdom,		(1.964)*		(−2.927)**	(2.626)**		
Germany, 1953–2001	4	0.1523	0.344	−1.2561	0.1754	0.008**	0.307
		(2.268)*		(−1.909)	(2.709)**		
	20	0.0657	0.495	0.0157	0.0653	0.637	0.658
		(0.763)		(0.0319)	(0.818)		

		Univariate regression		Bivariate regression			
	k-mths	dy^{12} only	J-test	r	dy^{12}	chi χ^2 test	J-test
Panel B: monthly MSCI data							
Pooled United States,	1	0.0560	0.096	−1.8161	0.1640	0.031*	0.016*
United Kingdom, Germany,		(0.866)		(−2.718)**	(2.222)*		
France, 1975–2001	12	0.0386	0.327	−1.1392	0.1060	0.229	0.113
		(0.533)		(−2.045)*	(1.337)		
	60	0.0169	0.663	0.1799	0.0035	0.699	0.887
		(0.130)		(0.405)	(0.033)		

We estimate regressions of the form $\tilde{y}_{t+k} = \alpha_k + z_t'\beta + \varepsilon_{t+k,k}$ where $\tilde{y}_{t+k}$ is the cumulated and annualized k-period ahead excess return, with instruments z_t being log dividend yields or risk-free rates and log dividend yields together. T-statistics in parentheses are computed using Hodrick (1992) standard errors. Panel A estimates the regression pooling data across the United States, United Kingdom, and Germany on data from 1953–2001. The estimates listed in the United Kingdom and Germany panels allow each country to have its own predictive coefficients and intercepts, but we compute Seemingly Unrelated Regression (SUR) standard errors following the method outlined in the Appendix. The coefficients listed in the pooled panel are produced by constraining the predictive coefficients to be the same across countries. In Panel B, monthly frequency MSCI data is used from 1975–2001. The column labeled "χ^2 test" reports a p-value for a test that both the risk-free rate and log dividend yield coefficients are jointly equal to zero. The "J-test" columns report p-values for a χ^2 test of the overidentifying restrictions.
*p<0.05.
**p<0.01.

coefficients are the same order of magnitude than those of the United States, but are all insignificant.

Figure 4 displays the coefficient patterns for the annualized short rate and its associated t-statistics in the bivariate regression for the 1975–2001 sample. Strikingly, this coefficient pattern is very robust across countries. For all countries, the one-month coefficient is negative, below −3 for

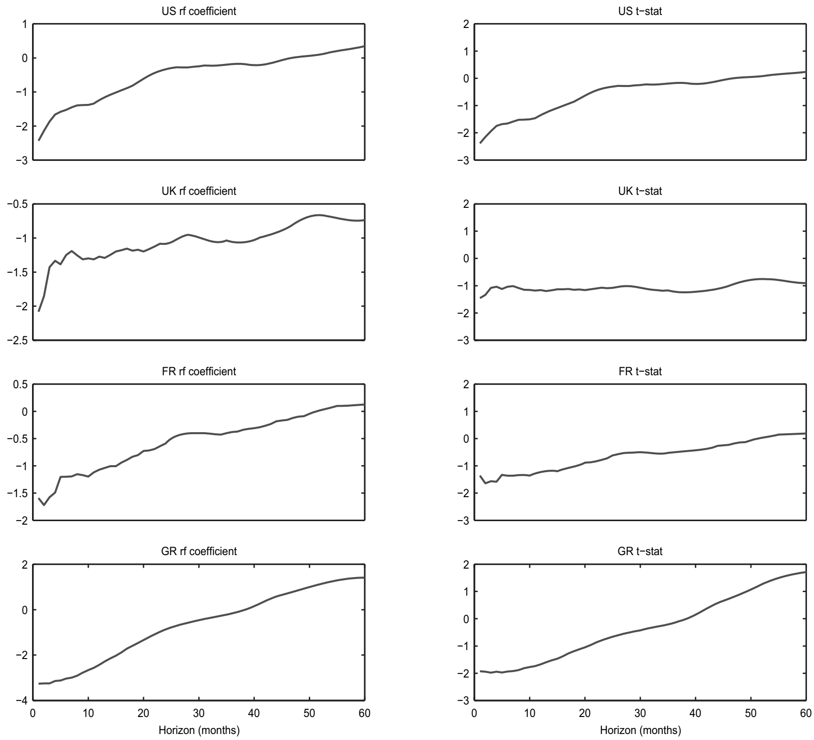

Figure 4
Short rate coefficients from bivariate regressions in four countries
The left (right) column shows the risk-free rate coefficients β_k (t-statistics) from the bivariate regression $\tilde{y}_{t+k} = \alpha + z_t\beta_k + \varepsilon_{t+k,k}$, where $\tilde{y}_{t+k}$ is the cumulated and annualized k-month ahead excess return and $z_t = (r_t \ dy_t^{12})$ contains the annualized risk-free rate and the log dividend yield. We report only the short rate coefficient. T-statistics are computed using Hodrick (1992) standard errors. The monthly data is from MSCI and the sample period is from 1975 to 2001.

Germany and around −1.5 for France. The coefficient monotonically increases with the horizon, leveling off around 0.35 for the United States, 0.13 for France, 1.41 for Germany, and −0.74 for the United Kingdom. The t-statistics are larger in absolute magnitude for short horizons. In particular, at the one-month horizon, the short rate coefficients are statistically different from zero for the United States and Germany and the t-statistics are near 1.5 (in absolute value) for the United Kingdom and France.

Panel A of Table 3 reports the bivariate coefficients for the long sample for the United Kingdom and Germany. Both countries have negative coefficients on the short rate. For Germany, the short rate coefficient is highly significant, while the UK t-statistic is only −1.5. For both countries, the short rate coefficients increase with horizon and turn positive at the five-year horizon. Similar to the United States, the dividend yield coefficients are larger in the bivariate regressions than in the univariate regression. However, the dividend yield coefficient is still only significantly different from zero at the 5% level in the United Kingdom at the one-year horizon.

To obtain more clear-cut conclusions, Table 3 reports pooled predictability coefficients and tests. We pool across the United States, United Kingdom, and Germany for the long sample in Panel A, and pool across all four countries in Panel B. The univariate dividend yield regression delivers mixed evidence across the two samples. For the long sample, the dividend yield coefficients are larger than 0.10 at the one-quarter and one-year horizons and statistically significant at the 5% level. The joint predictability tests for the shorter sample reveal a pattern of small dividend yield coefficients that decrease with horizon and are never significantly different from zero. We also report a J-test of the overidentifying restrictions for the joint estimation (see Appendix C). This test fails to reject for all horizons in both the long and short samples, which suggests that pooling is appropriate.

What is most striking about the bivariate regression results across the long and short samples is the consistency of the results. At the one-period forecasting horizon, the short rate coefficient is -1.96 in the long sample and -1.82 in the short sample, both significant at the 1% level. The bivariate regression also produces a dividend yield coefficient around 0.16 that is significant at the 1% (5%) level in the long (short) sample. Not surprisingly, the joint test rejects at the 5% level. However, for the short sample, the test of the overidentifying restrictions rejects at the 5% level, suggesting that pooling may not be appropriate for this horizon. For longer horizons, this test does not reject, and the evidence for predictability weakens. Nevertheless, for the long sample, we still reject the null of no predictability at the 1% level for the one-year horizon.

We conclude that whereas the dividend yield is a poor predictor of future returns in univariate regressions, there is strong evidence of predictability at short horizons using both dividend yields and short rates as instruments. The short rate is the stronger predictor and predicts excess returns with a coefficient that is negative in all four countries that we consider.

3. Do Dividend Yields Predict Cash flows or Interest Rates?

Our predictability results overturn some conventional, well-accepted results regarding the predictive power of dividend yields for stock returns. The dividend yield is nonetheless a natural predictor for stock returns. Define the discount rate δ_t as the log conditional expected total return, $\ln(E_t[Y_{t+1}])$:

$$\exp(\delta_t) = \mathrm{E}_t[(P_{t+1} + D_{t+1})/P_t] \equiv \mathrm{E}_t[Y_{t+1}]. \tag{4}$$

Denoting g^d_{t+1} as log dividend growth, $g^d_{t+1} = \log(D_{t+1}/D_t)$, we can rearrange (4) and iterate forward to obtain the present value relation:

$$\frac{P_t}{D_t} = \mathrm{E}_t\left[\sum_{i=1}^{\infty} \exp\left(-\sum_{j=0}^{i-1} \delta_{t+j} + \sum_{j=1}^{i} g^d_{t+j}\right)\right], \tag{5}$$

assuming a transversality condition holds. Note that Equation (5) is different from the Campbell and Shiller (1988a,b) log linear approximation for the log price-dividend ratio $p_t - d_t = \log(P_t/D_t)$:

$$p_t - d_t \approx c + \mathrm{E}_t\left[\sum_{j=1}^{\infty} \rho^{j-1}(y_{t+j} - g^d_{t+j})\right], \tag{6}$$

where c and ρ are linearization constants. Equation (5) is an exact expression and involves true expected returns. In contrast, the approximation in Equation (6) involves actual total log returns y_t.

Since the price-dividend ratio varies through time, so must some component on the RHS of Equation (5). As the discount rate is the sum of the risk-free rate and a risk premium, time-varying price-dividend ratios or dividend yields consequently imply that either risk-free rates, risk premiums, or cash flows must be predictable by the dividend yield. Although we find predictable components in excess returns, the dividend yield appears to be a strong predictive instrument at short horizons only when augmented with the short rate. Of course, the nonlinearity in Equation (5) may make it difficult for linear predictive regressions to capture these predictable components. In this section, we examine whether the dividend yield predicts cash flow growth rates or future interest rates.[6]

3.1 Dividend growth predictability

Panel A of Table 4 investigates US dividend growth over two samples, 1935–2001 and 1952–2001. Over the longer sample, we find no evidence of dividend growth predictability. For the shorter sample, high dividend yields predict future high dividend growth at the one- and four-quarter

[6] Goyal and Welch (2003) show that in a Campbell and Shiller (1988a,b) log-linear framework, the predictive coefficient on the log dividend yield in a regression of the one-period total return on a constant and the log dividend yield can be decomposed into an autocorrelation coefficient of the dividend yield and a coefficient reflecting the predictive power of dividend yields for future cash-flows. In Section 4, we attribute the time variation of the dividend yield into its three possible components—risk-free rates, excess returns, and cash flows—using a nonlinear present value model.

Table 4
Predictability of dividend growth

		Univariate regression	Bivariate regression		
	k-qtrs	dy^4 only	r	dy^4	χ^2 test
Panel A: US S&P data					
1935–2001	1	−0.0036	0.0586	−0.0027	0.974
		(−0.175)	(0.227)	(−0.147)	
	4	−0.0126	−0.0389	−0.0133	0.738
		(−0.504)	(−0.157)	(−0.588)	
	20	−0.0100	0.0384	−0.0091	0.886
		(−0.489)	(0.299)	(−0.483)	
1952–2001	1	0.0251	0.3541	0.0188	0.000**
		(2.476)*	(2.191)*	(1.599)	
	4	0.0259	0.1791	0.0228	0.005**
		(2.503)*	(1.131)	(1.903)	
	20	0.0165	0.0863	0.0132	0.325
		(0.927)	(0.808)	(0.649)	
Panel B: pooled across the United States, United Kingdom, and Germany					
1953–2001	1	−0.1545	0.6991	−0.1677	0.000**
		(−15.54)**	(4.296)**	(−14.16)**	
	4	−0.1552	0.4897	−0.1642	0.000**
		(−14.64)**	(3.005)**	(−13.19)**	
	20	−0.0489	0.2678	−0.0555	0.030*
		(−2.544)*	(2.276)	(−2.452)*	
		Univariate regression	Bivariate regression		
	k-mths	dy^{12} only	r	dy^{12}	χ^2 test
Panel C: pooled across United States, United Kingdom, Germany, and France					
1975–2001	1	−0.0179	0.3248	−0.0371	0.616
		(−0.757)	(1.194)	(−1.286)	
	12	−0.0116	−0.3348	0.0082	0.377
		(−0.440)	(−1.511)	(0.281)	
	60	−0.0077	−0.0702	−0.0025	0.266
		(−0.138)	(−0.359)	(−0.059)	

We estimate regressions of cumulated and annualized k-period ahead dividend growth, on log dividend yields alone or risk-free rates and log dividend yields together. Panels B and C pool data jointly across countries, constraining the predictive coefficients to be the same across countries. The χ^2 test column reports a p-value for a test that both the risk-free rate and log dividend yield coefficients are jointly equal to zero. T-statistics in parentheses are computed using Hodrick (1992) standard errors.
*$p<0.05$.
**$p<0.01$.

horizons. The magnitude of the coefficients is preserved in the bivariate regression, but the coefficient is no longer significantly different from zero at the one-quarter horizon and borderline significant at the one-year horizon. However, the short rate coefficient is positive and strongly significant at the one-quarter horizon. The coefficient becomes smaller and insignificant at longer horizons. The joint tests (across the two coefficients) reject at the 1% level for both the one- and four-quarter horizons.

Campbell and Shiller (1988a,b) note that the approximate linear relation (6) implies a link between high dividend yields today and either high future returns, or low future cash flows, or both. Hence, the positive sign of the dividend yield coefficient in the short sample is surprising. However, the Campbell–Shiller intuition is incomplete because it relies on a linear approximation to the true present value relation (5). Positive dividend yield coefficients in predictive cash-flow regressions can arise in rational models. For example, Ang and Liu (2006) show how the nonlinearity of the present value model can induce a positive dividend yield coefficient. Menzly, Santos, and Veronesi (2004) show that the dividend yield coefficient is a function of a variable capturing shocks to aggregate preferences. Consequently, it changes over time and can take positive values.

In Panels B and C of Table 4, we investigate the relation between dividend yields and cash-flows for other countries. Panel B pools data across the United States, United Kingdom, and Germany for the 1953–2001 sample. Unlike the US post-1952 sample, the dividend yield coefficients are strongly negative. Because the United Kingdom and German coefficients are so different from the United States (data not shown), a pooled result is hard to interpret, and the GMM over-identifying restrictions are strongly rejected with a *p*-value of less than 0.001. Nevertheless, pooling yields negative, not positive, dividend yield coefficients. The short rate coefficients are strongly positive and are about twice the magnitude of the US coefficients (a 1% increase in the short rate approximately forecasts an annualized 70 basis point increase in expected dividend growth over the next quarter).

Panel C reports coefficients for the MSCI sample. The dividend yield coefficients are small, mostly negative and never statistically significantly different from zero. The short rate coefficients are also insignificant, although they are similar in magnitude to the coefficient found in long-term US data. There is no general pattern in the individual country dividend yield coefficients (data not shown): the dividend yield coefficient in the univariate regression is positive (negative) in the United States and United Kingdom (France and Germany), with the dividend yield coefficients retaining the same signs in the bivariate regression in each country. All in all, we conclude that the evidence for linear cash-flow predictability by the dividend yield is weak and not robust across countries or sample periods.[7]

[7] It is conceivable that dividend yields exhibit stronger predictive power for *real* dividend growth. However, we find the results for real and nominal growth to be quite similar. In the long US sample, the dividend yield fails to forecast future ex-post real dividend growth and the coefficients are positive. For the shorter sample, pooled results across the four countries produce negative coefficients that are actually significant at short horizons. These results are available upon request. Campbell (2003) also finds analogous results.

3.2 Interest Rate Predictability

We examine the possibility that the dividend yield predicts risk-free rates in Table 5, which reports coefficients of a regression of future annualized cumulated interest rates on log dividend yields. The persistence of the risk-free rate causes two econometric problems in linear regressions. First, because the interest rate and dividend yield are both persistent variables, the regression is potentially subject to spurious regression bias. To address this issue, we also report results using the detrended dividend yield, which is the dividend yield relative to a 12-month moving average (Campbell 1991, Hodrick 1992). Second, the residuals in the predictive regression are highly autocorrelated, which means that the use of Hodrick (1992) standard errors is inappropriate. For a one-period horizon, we use Cochrane–Orcutt standard errors and generalize the use of this procedure to panel data in the Appendix. We do not report standard errors for horizons greater than one period, because the residuals contain both autocorrelation and moving average effects that cannot be accommodated in a simple procedure.

Table 5 reports that the long sample for the United States shows a positive effect of the dividend yield on future interest rates. The effect is economically small (a 1% increase in the log dividend yield predicts an increase in the one-quarter short rate by 1.7 basis points on an annualized

Table 5
Predictability of risk-free rates

	k-qtrs	dy^4	Detrended dy^4
Panel A: US Data			
1952–2001	1	0.0171	0.0458
		(0.202)	(0.416)
	4	0.0161	0.0429
	20	0.0267	0.0413
Panel B: pooled United States, United Kingdom and Germany			
1953–2001	1	0.0170	0.0211
		(0.923)	(0.765)
	4	0.0138	0.0170
	20	0.0099	0.0074
	k-mths	dy^{12}	Detrended dy^{12}
Panel C: pooled United States, United Kingdom, Germany, and France			
1975–2001	1	0.0594	0.0407
		(1.525)	(0.201)
	12	0.0588	0.0676
	60	0.0619	0.0495

We estimate regressions of cumulated and annualized k-period ahead average risk-free rates by log dividend yields. The detrended log dividend yield refers to the difference between the log dividend yield and a moving average of log dividend yields over the past year. Panel c pools data across countries. We compute Cochrane-Orcutt t-statistics (in parentheses) for a one-quarter horizon.

basis next quarter). Using a detrended dividend yield also leads to the same positive sign. While both effects are statistically insignificant, we view the relationship between dividend yields and future interest rates as economically important because interest rates are a crucial component of a present value relation. From the present value relation (5), we expect a positive relation between dividend yields and future discount rates. The interest rate enters the discount rate in two ways. The discount rate is the sum of the risk-free rate and the risk premium and enters these two components with opposite signs. It is the first component that gives rise to the positive relation.

Although not statistically significant, the positive sign of interest rate predictability by dividend yields is robust. First, omitting the 1990s does not change the inference, but actually increases the *t*-statistics. Second, we also find a positive sign for Germany and the United Kingdom in the long sample and for all countries in the short sample. In particular, for MSCI data, the individual coefficients range from 0.036 in the United States to 0.085 in France at the one-month horizon.[8]

4. A Present Value Model for Stock Returns

In this section, we present a present value model to shed light on what kind of discount rate processes are most consistent with the predictability evidence.

4.1 The model

We start with the basic present value relation in (5) and parameterize the dynamics of the discount rates and cash flows. We assume that the continuously compounded risk-free rate r_t and log dividend growth g_t^d follow the VAR:

$$X_t = \mu + \Phi X_{t-1} + \varepsilon_t, \tag{7}$$

where $X_t = (r_t, g_t^d)'$ and $\varepsilon_t \sim \text{IID}\, N(0,\Sigma)$. Let the discount rate, δ_t, in Equation (4) follow the process:

$$\delta_t = \alpha + \xi' X_t + \phi\delta_{t-1} + u_t, \tag{8}$$

with $u_t \sim \text{IID}\, N(0,\sigma^2)$ and ε_t and u_t are independent. We denote the individual components of ξ as $\xi = (\xi_r, \xi_{g^d})'$.

[8] We also examine the predictive power of the dividend yield for ex-post real interest rates, similar to Campbell (2003). Although the individual coefficients across countries fail to have a consistent sign, pooled results produce positive coefficients at all horizons, as in the nominal case. However, the coefficients are not statistically significant.

Proposition 4.1. *Assuming that* $X_t = (r_t, g_t^d)'$ *follows Equation (7) and that the log conditional total expected return* δ_t *follows (8), the price-dividend ratio* P_t/D_t *is given by*

$$\frac{P_t}{D_t} = \sum_{i=1}^{\infty} \exp(a_i + b_i' X_t + c_i \delta_t), \tag{9}$$

where

$$\begin{aligned} a_{i+1} =& \ a_i + c_i\alpha + \frac{1}{2}c_i^2\sigma^2 + (e_2 + b_i + c_i\xi)' \\ & \mu + \frac{1}{2}(e_2 + b_i + c_i\xi)'\Sigma(e_2 + b_i + c_i\xi) \\ b_{i+1}' =& \ (e_2 + b_i + c_i\xi)'\Phi \\ c_{i+1} =& \ \phi c_i - 1, \end{aligned} \tag{10}$$

where $e_2 = (0, 1)'$, a_i *and* c_i *are scalars, and* b_i *is a* 2×1 *vector. The initial conditions are given by*

$$\begin{aligned} a_1 &= e_2'\mu + \frac{1}{2}e_2'\Sigma e_2 \\ b_1' &= e_2'\Phi \\ c_1 &= -1. \end{aligned} \tag{11}$$

Proof: See Appendix D.

Proposition 4.1 implies that the dividend yield is a highly nonlinear function of interest rates, excess returns, and cash flows. Not surprisingly, if discount rates are persistent, the c_i coefficients are negative and higher discount rates decrease the price-dividend ratio. Analogously, when dividend growth is positively autocorrelated, a positive shock to dividend growth likely increases the price-dividend ratio, unless it entails an opposite discount rate effect ($\xi_g^d > 0$).

The present value model endogenously generates heteroskedasticity. While previous studies model returns and dividend yields in finite-order VAR systems (see, among many others, Hodrick 1992, Campbell and Shiller 1988a,b, Stambaugh 1999), a VAR cannot fully capture the nonlinear dynamics of dividend yields implied by the present value model. We can also contrast our present value model with Goetzmann and Jorion (1993, 1995) and Bollerslev and Hodrick (1996), who either ignore the cointegrating relation between dividends and price levels that characterizes rational pricing or only develop approximate solutions. In contrast, we impose cointegration between dividends and prices and our solution is exact.

Under the special case of constant total returns, that is $\delta_t = \alpha$, with $\xi = \phi = \sigma^2 = 0$, and IID dividend growth ($g_t^d = \mu_d + \sigma_d \varepsilon_t^g$), Equation (9) simplifies to a version of the Gordon model:

$$\frac{P_t}{D_t} = \frac{\exp(\mu_d + \frac{1}{2}\sigma_d^2)}{1 - \exp(\mu_d + \frac{1}{2}\sigma_d^2 - \alpha)}.$$

Another important special case of the model is constant expected excess returns, where $\delta_t = \alpha + r_t$, so $\xi = (1, 0)'$, and $\phi = \sigma^2 = 0$. In this case, the time variation in total expected returns is all due to the time variation in interest rates. This is the relevant null model for our excess return regressions where the expected excess return is constant but the total expected return varies with the interest rate.

Under the null of constant expected excess returns, the gross total return $Y_t = (P_{t+1} + D_{t+1})/P_t)$ less the gross interest rate $\exp(r_t)$ is

$$\mathrm{E}_t[Y_{t+1} - \exp(r_t)] = \exp(r_t)[\exp(\alpha) - 1],$$

so regressing the simple net excess return on the interest rate actually yields a nonzero coefficient on r_t. The scaled expected return, $\mathrm{E}_t[Y_{t+1}/\exp(r_t)]$ is constant and equal to α. The predictability regressions typically run in the literature do not correspond to any of these two concepts, since they use log returns $\tilde{y}_{t+1} \equiv \log(Y_{t+1}) - r_t$. In our economy, regressing log returns onto state variables does not yield zero coefficients because the log excess return is heteroskedastic. However, we would expect these coefficients to be small, relative to the null of time-varying expected excess returns (where δ_t takes the full specification in Equation (8)).

Under the alternative of time-varying discount rates in Equation (8), total expected returns can depend on both fundamentals (short rates and dividend growth) and exogenous shocks. The case of $\xi = 0$ represents fully exogenous time-varying expected returns. By specifying $\delta_t = \alpha + \xi' X_t$, Equation (8) also nests the case of state-dependent expected returns.

4.2 Estimation

The estimation of the present value model is complicated by the fact that in the data, we observe dividends summed up over the past year, but we specify a quarterly frequency in the model. We estimate the present value model with simulated method of moments (SMM) (Duffie and Singleton 1993) on US data from January 1952 to December 2001. We provide full details of the estimation in Appendix E.

Panel A of Table 6 reports the VAR parameters. Dividend growth displays significant positive persistence, and the interest rate has a small and insignificant effect on dividend growth. The implied unconditional standard deviation of g_t^d from the estimation is 0.0173 per quarter. If we estimate a VAR on $(r_t, g_t^{d,4})$, the implied unconditional standard deviation of $g_t^{d,4}$ is 0.0156 per quarter. Hence, summing up dividends over the past four quarters effectively creates a smoother series of dividend growth compared to the true, but unobservable, cash-flow process.

Panel B presents the parameter estimates for five different discount rate processes in Equation (8). The first model we estimate (Null 1) is a simple constant total expected stock return benchmark model. The second model (Null 2) is our main null model because it imposes constant expected

Table 6
Calibration of the present value model

		Φ		$\Sigma^{1/2}$	
	μ	r_t	g_t^d	r_t	g_t^d
Panel A: estimates of the VAR for $(r_t g_t^d)'$					
r_t	0.0010	0.9263	0.0000	0.0026	0.0000
	(0.0005)	(0.0396)		(0.0006)	
g_t^d	0.0053	−0.0078	0.5489	0.0036	0.0171
	(0.0013)	(0.0899)	(0.1208)	(0.0001)	(0.0048)
	Null 1	Null 2	Alt 1	Alt 2	Alt 3
Panel B: estimates of the discount rate process					
$\alpha \times 100$	2.0168	0.8508	0.0407	0.0476	−0.0988
	(0.0358)	(0.0349)	(0.0123)	(0.0034)	(0.2525)
ϕ			0.9816		0.9298
			(0.0052)		(0.0616)
ξ_r		1.0000		−2.0454	0.0007
				(0.4038)	(0.0011)
ξ_g^d				0.0041	0.0014
				(0.5704)	(0.0024)
$\sigma \times 100$			0.1751		0.5024
			(0.0475)		(0.1616)
χ^2 test p-value	0.0000	0.0000	0.0101	0.0000	0.1332

The table reports parameter estimates and standard errors in parentheses of the present value model. Panel A reports estimates of the VAR of short rates and dividend growth in Equation (7). The short rate r_t equation is an AR(1), with standard errors produced by GMM with four Newey–West (1987) lags. The parameters for g_t^d are estimated using SMM by matching first and second moments of $g_t^{d,4}$, along with the moments $E(r_t g_t^{d,4})$, $E(g_{t-4}^{d,4} g_t^{d,4})$, and $E(r_{t-4} g_t^{d,4})$. Panel B reports parameter estimates for the discount rate process $\delta_t = \alpha + \xi' X_t + \phi\delta_{t-1} + u_t$ [see Equation (8)], with $\xi = (\xi_r, \xi_{g^d})'$. The Null Models 1 and 2 impose the restriction $\sigma = \xi_{g^d} = \phi = 0$, with $\xi_r = 0$ for Null Model 1 or $\xi_r = 1$ for Null Model 2. These represent the null hypotheses of constant expected total returns (Null Model 1) or constant expected excess returns (Null Model 2). Alternative Model 1 sets $\xi_r = \xi_{g^d} = 0$, so the discount rate process is entirely exogenous, whereas Alternative Model 2 imposes $\phi = \xi_{g^d} = 0$, so the discount rate process is entirely endogenous. In Alternative Model 3, all parameters of the discount rate process are nonzero. The estimation is done by holding the VAR parameters fixed and matching the first and second moments of excess returns and dividend yields, along with lagged short rates and lagged dividend growth as instruments. The last row reports the p-value from a χ^2 overidentification test.

excess returns by setting $\xi_r = 1$ and $\xi_{g_d} = \phi = \sigma = 0$. The first model under the alternative of time-varying discount rates that we consider (Alternative 1) features completely exogenous discount rates ($\xi = 0$). The estimation shows that the log discount rate is very persistent ($\phi = 0.98$) and its unconditional variance is about 1% at the quarterly level. Alternative 2 sets $\phi = 0$ but allows the discount rate process to depend on the two state variables. We find a slightly positive but insignificant effect of dividend growth rates on the discount rate, but a strong and significantly negative interest rate effect. Alternative 3 combines the two models. The negative interest rate effect disappears, but the zero coefficient means that excess returns are negatively related to interest rates. The persistence of the discount rate process now drops to 0.93. The last line of Panel B reports the p-value of the χ^2-test of the overidentifying restrictions for the SMM estimation. Only Alternative 3 passes this test.

4.3 Economic implications

In this section, we investigate how well the present value models match the data moments and decompose the variability of the price-dividend ratio into its components. We also examine how well linear predictive regressions capture true expected returns implied by the models.

4.3.1 Moments and price-dividend ratio variance decomposition. Table 7 reports a number of implied moments for the various models. Panel A reports the variance and mean of the dividend yield and excess returns. We start by examining the null models. Because Null 1 has no time variation in expected returns, it underestimates the volatility of excess returns and generates only one-tenth of the dividend yield variability present in the data. The annualized mean equity premium is only 2.6% instead of 6.12% in the data, but this is still comfortably within two standard errors of the data moment. The model also matches the mean dividend yield. The Null 2 model has similar mean implications, but the variation of excess returns increases substantially and endogenous dividend yield volatility triples. By definition, all of the variation in price-dividend ratios should come from either short rates or cash flows, which we confirm in a variance decomposition of the price-dividend ratio reported in Panel B.

The variance decompositions represent the computation:

$$1 - \frac{\text{var}_z(P/D^4)}{\text{var}(P/D^4)}, \tag{12}$$

where $\text{var}(P/D^4)$ is the variance of price-dividend ratios implied by the model, and $\text{var}_z(P/D^4)$ is the variance of the price-dividend ratio produced by the model when the variable z is nonstochastic and set at its

Table 7
Economic implications of the present value model

		Null 1	Null 2	Alt 1	Alt 2	Alt 3	US data Estm	US data SE
Panel A: implied selected moments								
Mean excess return		0.0065	0.0066	0.0066	0.0054	0.0068	0.0153	0.0057
Mean dividend yield		0.0335	0.0339	0.0343	0.0348	0.0352	0.0349	0.0008
Volatility excess return		0.0393*	0.0472*	0.0785	0.0824	0.0811	0.0779	0.0061
Volatility dividend yield		0.0010*	0.0031*	0.0120	0.0072*	0.0115	0.0114	0.0011
Panel B: decompositions of the variance of the price-dividend ratio								
Percent of short rate		−0.0009	0.8745	−0.0003	0.9757	0.2221		
Percent of dividend growth		1.0000	0.0545	0.0042	0.0328	0.0689		
Percent of total discount rate		−0.0009	0.1418	0.9955	0.1593	0.9295		
Percent of excess return		0.0008	0.0000	0.0214	0.2911	0.6132		
Panel C: correlations between true expected excess returns and forecasts from predictive regressions								
$k=1$, univariate regression	dy^4			0.7364	0.7783	0.7859		
$k=1$, bivariate regression	dy^4,r			0.9325	0.9785	0.8654		
$k=4$, univariate regression	dy^4			0.7521	0.7684	0.7710		
$k=4$, bivariate regression	dy^4,r			0.9328	0.9916	0.8297		
$k=20$, univariate regression	dy^4			0.8506	0.7539	0.7976		
$k=20$, bivariate regression	dy^4,r			0.9177	0.9971	0.8165		

The table reports various economic and statistical implications from the present value models. Panel A reports various moments and summary statistics implied from each model. The quarterly moments reported are the mean and volatility of excess returns and dividend yields in levels. The data standard errors of the moments are computed by GMM with four Newey–West (1987) lags. In Panel B, the variance decompositions report the computation $1 - \text{var}_z(P/D^4)/\text{var}(P/D^4)$, where $\text{var}(P/D^4)$ is the variance of price-dividend ratios implied by the model, and $\text{var}_z(P/D^4)$ is the variance of the price-dividend ratios where all realizations of $z = r_t$, g_t, δ_t or the risk premium are set at their unconditional means. Panel C reports the correlation between fitted values from the excess return predictive regressions and true conditional expected excess returns $E_t(\tilde{y}_{t+k})$ implied by the model. The population moments implied by the model are computed using 1,00,000 simulations from the estimates in Table 6. In Panel A, the standard errors for the sample moments are computed using GMM.
*Indicates that the population moments lie outside a two standard deviation bound around the point estimate of the sample moment.

long-run mean. We take z to be short rates, dividend growth, total discount rates, and excess discount rates, respectively. For Null 2, the price-dividend ratio variance accounted for by short rate movements is 87.4%. The total discount rate accounts for 14.2% of the variation of the price-dividend ratio, but this is all due to time-varying interest rates. By construction, the excess discount rate does not account for any of the volatility of the price-dividend ratio. Dividend growth also accounts for only a small part (5.5%) of total price-dividend ratio volatility.

The alternative models 1–3 match the data much better than the null models, showing that some variation in (excess) discount rates is essential. In particular, all three alternative models match the variability of excess returns and, in addition, Alternatives 1 and 3, both featuring exogenous discount rate variation, match the variability of dividend yields.

Alternative 3, the best fitting model, also perfectly matches the first-order autocorrelation of the dividend yield (0.9596 compared to 0.9548 in US data spanning 1952–2001). The variance decomposition of the various models are as expected, given how the discount rates are modeled in each specification: in Alternative 1, most variation comes from the exogenous discount rate; in Alternative 2, almost all variation comes from the short rate (which in turn drives variation in the discount rate), and in Alternative 3, the exogenous discount rate dominates but interest rates are still important. Since Alternative 3 fits the data the best, the variance decomposition is of considerable interest. It suggests that 61% of the price-dividend ratio variation is driven by risk premiums, 22% by the short rate and 7% by dividend growth. The remainder is accounted for by covariance terms.

4.3.2 How well do predictive regressions capture true expected returns?

In Panel C of Table 7, we examine how well linear forecasting models can capture true time variation in expected returns in the alternative models. We compute the correlation of expected log excess returns, $E_t(\tilde{y}_{t+k})$, implied by the model with the fitted value on the RHS of the regression (2).[9] There are two striking results in Panel C. First, a univariate dividend regression captures a much smaller proportion of movements in log expected excess returns than a bivariate regression including both risk-free rates and dividend yields. For example, for Alternative 1 (3), the correlation rises from 74% (75%) in a univariate regression to 93% (84%) in a bivariate regression at a one-quarter horizon.

Second, long-horizon regressions do not necessarily capture the dynamics of true expected excess returns better than short-horizon regressions. Cochrane (2001) and others, show that in linear VAR models, long-horizon regressions more successfully capture predictable components in expected returns. However, in our nonlinear present value model, long-horizon regressions may fare worse than short-horizon regressions in capturing true expected returns. For example, for Alternative 3, the forecasts from a one-quarter regression with the short rate and dividend yield have a correlation of 87% with true expected excess returns, whereas the correlation is 82% at a 20-quarter horizon. In Alternative 2, where there are no exogenous components in discount rates, the correlation between the bivariate regression and true expected returns is around

[9] In a present value model, the economically relevant quantity for discount rates is the log expected return $\delta_t = \log((P_{t+1} + D_{t+1})/P_t) = \log(E_t(Y_{t+1}))$. However, the predictive regressions produce a forecast of expected log returns $E_t(\log(Y_{t+1}))$. The two quantities $\log(E_t(Y_{t+1}))$ and $E_t(\log(Y_{t+1}))$ are not equivalent because of time-varying Jensen's inequality terms. For example, the correlation of $\sum \delta_{t+j} - r_{t+j}$ in Null Model 2 is zero with any variable, whereas the correlation of expected excess log returns $E_t(\tilde{y}_{t+k})$ is not. We compute $E_t(\tilde{y}_{t+k})$ following the method described in Appendix E.

98% for all horizons. However, this alternative has the worst fit with the data.

4.4 Implications for predictive regressions

We now investigate how well the present value models fit the linear patterns of predictability that we observe in the data. In Table 8, we compare regression coefficients implied by the present value models to their values in data.

4.4.1 Expected excess return regressions. In the Null 1 model, total expected returns are constant. Hence excess returns are, by construction,

Table 8
Predictive regressions implied by the present value model

		Null 1	Null 2	Alt 1	Alt 2	Alt 3	US data Estm	US data SE
Excess return regressions								
$k = 1$, univariate regression	dy^4	−0.0416*	−0.0404*	0.0886	0.3564*	0.1864	0.0979	0.0635
$k = 1$, bivariate regression	r	−1.0500	−0.0824*	−0.9310	−2.6369	−1.0737	−2.1623	0.7426
	dy^4	−0.0049*	−0.0214*	0.0878	0.0574	0.2269	0.1362	0.0633
$k = 4$, univariate regression	dy^4	−0.0269	−0.0177	0.0870	0.3122*	0.1723	0.1060	0.0686
$k = 4$, bivariate regression	r	−0.9419	−0.1191	−0.8541	−2.5694	−0.8562	−1.4433	0.7478
	dy^4	0.0062	0.0097	0.0862	0.0208	0.2046	0.1313	0.0683
$k = 20$, univariate regression	dy^4	−0.0213	0.0086	0.0787	0.1772	0.1232	0.0594	0.1245
$k = 20$, bivariate regression	r	−0.4983	0.0050	−0.4170	−1.5665	−0.3456	−0.4829	3.2404
	dy^4	−0.0040	0.0075	0.0783	−0.0004	0.1363	0.0774	0.6445
Dividend growth regressions								
$k = 1$, univariate regression	dy^4	0.8543*	0.0918*	0.0045*	0.0199	0.0412	0.0251	0.0101
$k = 1$, bivariate regression	r	0.2611	0.1548	0.2942	0.8837*	0.1154	0.3541	0.1616
	dy^4	0.8452*	0.0561*	0.0047	0.1201*	0.0369	0.0188	0.0118
$k = 4$, univariate regression	dy^4	0.9392*	0.0920*	0.0050	0.0259	0.0212	0.0259	0.0103
$k = 4$, bivariate regression	r	0.1478	−0.1019	0.1846	0.7031*	0.0997	0.1791	0.1584
	dy^4	0.9340*	0.1155*	0.0052	0.1057*	0.0174	0.0228	0.0120
$k = 20$, univariate regression	dy^4	0.2274*	0.0375	−0.0017	−0.0016	0.0059	0.0165	0.0178
$k = 20$, bivariate regression	r	0.0778	−0.0146	0.0861	0.1776	0.0707	0.0863	0.5339
	dy^4	0.2247*	0.0408	−0.0016	0.0186	0.0032	0.0132	0.1021
Risk-free rate regressions								
$k = 1$, univariate regression	dy^4	0.0342*	0.2142*	−0.0008	−0.1052*	0.0354	0.0171	0.0390
$k = 4$, univariate regression	dy^4	0.0334	0.1928	−0.0008	−0.0946	0.0321	0.0161	–
$k = 20$, univariate regression	dy^4	0.0230	0.1163	−0.0016	−0.0568	0.0194	0.0267	–

The table reports implied predictive regression coefficients from the present value models. The LHS variables are cumulated log excess returns, dividend growth, and risk-free rates. The population moments are computed using 1,00,000 simulations from the estimates in Table 6. The standard errors for the regressions in US data are computed using Hodrick (1992) standard errors for the excess return and dividend growth regressions and using a Cochrane-Orcutt procedure for the $k = 1$ risk-free rate regression. All horizons k are in quarters.
*Indicates that the coefficients lie outside a two standard deviation bound around the point estimate of the sample coefficient.

negatively related to interest rates, which explains the large negative short rate coefficients in the bivariate regressions. Because dividend yields are correlated with interest rates, the univariate regression also picks up some predictability. In the Null 2 model, expected excess returns are constant, so that the only predictability we should observe comes from nonlinearities. As in the Null 1 model, the coefficients on the dividend yield are invariably small and of the wrong sign compared to the predictions in Campbell and Shiller (1988a,b).

The alternative models imply different patterns of predictability. Despite substantial variation in the exogenous discount rate, Alternative 1 generates a population slope coefficient of only 0.089 in a dividend yield regression, which is slightly below what is observed in the data. Because the discount rate is not linked to the interest rate, an excess return projection on dividend yields and the interest rate leads to coefficients on interest rates in the neighborhood of -1 and a dividend yield coefficient of 0.088. Both coefficients are somewhat lower in absolute magnitude than what is observed in the data.

Alternative 2 implies a dividend yield predictability regression coefficient of 0.356. This is very close to the dividend yield coefficient for the United States if we ignore the 1990s (0.296 in Table 2). Hence, in a world where discount rates only depend on short rates and dividend growth, the dividend yield would indeed be a strong predictor of excess returns. However, because the variation in discount rates is mostly driven by short rates, the dividend yield coefficient drops to 0.057 in a bivariate regression and the predictable component is now mostly absorbed by the short rate. This is inconsistent with the data, where bivariate regressions yield larger dividend yield coefficients, not smaller ones.

Alternative 3 combines features of both Alternatives 1 and 2 and the coefficients are nicely in between the two alternatives. This model also yields a negative omitted variable bias in the univariate dividend yield regression as is true in the data. Note that, with the exception of the univariate slope coefficient in the dividend yield regression under Alternative 2, all predictability coefficients for all alternative models are within two standard errors of the observed coefficients in the data.

4.4.2 Dividend growth regressions Table 8 also reports how the different models fare with respect to the predictability of dividend growth. The main feature in the data is that both dividend yields and short rates receive positive coefficients in a bivariate regression. Since practically all of the variation in dividend yields is due to dividend growth in the Null 1 model, dividend growth is much too predictable in this model. Dividend growth predictability in the Null 2 model is also inconsistent with the data.

In the alternative models, positive dividend yield coefficients in dividend growth regressions occur primarily in two ways. First, the discount rate can depend positively on dividend growth [$\xi_g^d > 0$ in Equation (8)]. Positive persistence in dividend growth normally leads to a negative association between dividend yields and future cash flows, but in this case, high dividend growth also increases the discount rate and increases dividend yields. Second, the short rate may enter negatively in the dividend growth Equation (7). This means that low discount rates (which tend to raise the dividend yield) are directly associated with relatively higher expected dividend growth rates. Our estimates for Alternative 3 share both these features (Table 6), and the implied moments of this model are fully consistent with what is observed in the data.

4.4.3 Risk-free rate regressions A rational present value model with stochastic interest rates also implies that dividend yields predict interest rates. Under Null 2, the coefficient is 0.214, much larger than what is observed in the data. Since interest rates do not affect the risk premium in this model, we observe a very strong positive relation between the current dividend yield and future interest rates. The smaller coefficient in the data indicates that the risk premium component is likely to negatively depend on interest rates. Hence, the current dividend yield should show a negative relation with the interest rate component of future risk premiums. In Alternative 1, the total discount rate does not depend on the interest rate, yielding a coefficient very close to zero. Alternative 2 makes the interest rate dependence of the discount rate too strong, resulting in a negative coefficient. In contrast, Alternative 3 most closely (but not perfectly) matches the coefficient in the data. Hence, we conclude that the present value model represented by Alternative 3 is remarkably consistent with the data.

5. Bias, Size and Power

The inference regarding predictability may critically depend on the finite sample properties of the estimator. In this section, we investigate the finite sample bias, size, and power of the estimators in the linear predictive regressions.

5.1 Small sample bias

In a linear return-dividend yield system (Stambaugh 1999, Amihud and Hurvich 2004, Lewellen 2004), there is an upward bias in the predictive coefficient on the dividend yield, deriving from the negative correlation between return and dividend yield innovations and the persistence of dividend yields. This analysis does not apply in our framework for two reasons. First, both the return and the dividend yield are nonlinear

Table 9
Small sample bias under null model 2

	k-qtrs		1975–2001, $T = 104$	1952–2001, $T = 200$	1935–2001, $T = 267$	Population, $T = \infty$
Scaled returns						
Univariate	1	dy^4	0.0235	0.0045	0.0012	0.0000
Bivariate	1	r	0.3469	0.1799	0.1435	0.0000
		dy^4	−0.0339	−0.0286	−0.0264	0.0000
Excess log returns						
Univariate	1	dy^4	0.0245	0.0065	−0.0027	−0.0404
	4	dy^4	0.0408	0.0204	0.0105	−0.0177
	20	dy^4	0.0441	0.0273	0.0185	0.0086
Bivariate	1	r	0.3521	0.1831	0.1388	−0.0824
		dy^4	−0.0343	−0.0275	−0.0295	−0.0214
	4	r	0.2632	0.1345	0.0989	−0.1191
		dy^4	0.0038	−0.0005	−0.0057	0.0097
	20	r	0.1941	0.1003	0.0696	0.0050
		dy^4	0.0137	0.0105	0.0067	0.0075

The table reports small sample and population parameter coefficient values for regressions of scaled or cumulated log excess returns onto log dividend yields (univariate regression) or annualized short rates and log dividend yields (bivariate regression) from the Null 2 Model. Scaled returns are defined as $Y_{t+1}/\exp(r_t)$, where Y_{t+1} is the gross total equity return. The population moments from the model are computed using 1,00,000 simulations. The small sample moments are computed using 10,000 simulations of samples of varying length. All horizons k are in quarters.

processes in our model. Second, once we consider multivariate regressions, the bias can no longer be signed in all cases, as already noted by Stambaugh (1999).

Table 9 reports the small sample bias in several regressions based on data generated from the Null 2 model (constant expected excess returns). We consider the small sample distributions for samples of length 104, 200, and 267 quarters, which correspond to the 1975–2001, 1952–2001, and 1935–2001 sample periods, respectively. We start with regressions using scaled returns, $Y_{t+1}/\exp(r_t)$, as the dependent variable. In these regressions, the population coefficients are zero as the null imposes constant expected scaled returns. In small samples, the univariate regression yields a Stambaugh bias that is negligible for our longest sample (0.0012). Interestingly, in the bivariate regression, the bias on the dividend yield coefficient is now negative, but the bias on the short rate coefficient is positive and rather large, ranging from 0.35 for the smallest sample to 0.14 for the longest sample. In this model, the short rate plays the role of the dividend yield in the linear systems. It is the prime determinant of the variation in the price-dividend ratio, it is contemporaneously negatively correlated with returns, and it is very persistent. Hence, a Stambaugh-like bias for short rates, but not dividend yields, results.

In the second panel of Table 9, we report the bias results for the predictive regressions using log excess returns. Note that the population coefficients are not zero because of heteroskedasticity. For the univariate dividend yield regressions, the population coefficients are slightly negative at short horizons and positive at the five-year horizon. For our small sample, they interact with the Stambaugh bias to produce a small upward bias for our smallest sample but a downward, negligible bias for our longest sample. For the bivariate regressions, the population biases are negative for both the short rate and dividend yield coefficients. However, in small samples, the scaled return biases clearly dominate and we find the short rate (dividend yield) coefficient to be biased upward (downward). The small sample biases become significantly smaller at long horizons and turn slightly positive for the dividend yield coefficient. Clearly, recent inference techniques that focus on univariate regressions in a linear framework [Valkanov (2003), Torous, Valkanov, and Yan (2004), Polk, Thompson, and Vuolteenaho (2006)] are of little use in our setting.

In summary, the small sample biases under the null strengthen our empirical evidence. That is, the univariate dividend yield predictability coefficients are slightly overestimated, and the small sample biases for the bivariate coefficients go the wrong way at short horizons: the estimated short rate coefficient is negative, but the small sample bias is positive; and the estimated dividend yield coefficient is positive, but the small sample bias is negative.

5.2 Size

Table 10 reports empirical sizes for tests of a 5% nominal (asymptotic) size. In the shortest sample for the one-quarter horizon, the univariate dividend yield regression displays negligible size distortions, but for the bivariate regressions, all tests slightly over-reject at asymptotic critical values. For longer horizons, the performance of the Newey–West and robust Hansen–Hodrick estimators deteriorates, with the empirical size exceeding 33% for a 5% test in the univariate dividend regression.[10] For our longer samples, these distortions become smaller but do not disappear. For example, at the five-year horizon, the empirical size of the dividend yield regression still exceeds 18.5% for both Newey–West and robust Hansen–Hodrick estimators in the 267 quarter sample.

The Newey–West and robust Hansen–Hodrick standard errors are too small because they underestimate the serial correlation in the error terms as the autocorrelation estimates are downwardly biased. The Newey–West standard error also uses a Bartlett kernel of declining, tent-shaped weights. Under the null, the kernel is rectangular, so the Newey–West standard errors underweight the effect of autocorrelations at long lags unless a higher

[10] Hodrick standard errors also have relatively few size distortions, especially compared to Newey-West standard errors, for standard linear VARs, like the Stambaugh (1999) system. Hodrick (1992) also demonstrates that Hodrick standard errors are correctly sized for multivariate VARs.

Table 10
Empirical size

	Univariate regression			Bivariate regression							
	dy^4			r			dy^4			χ^2 tests joint across horizon	
k-qtrs	N–W	Robust H–H	Hodrick	N–W	Robust H–H	Hodrick	N–W	Robust H–H	Hodrick	Univariate	Bivariate
Sample length=104 quarters (1975–2001)											
1	0.053	0.047	0.047	0.076	0.070	0.070	0.065	0.059	0.059	0.051	0.010
4	0.132	0.104	0.044	0.151	0.138	0.062	0.113	0.109	0.050		
20	0.338	0.373	0.047	0.335	0.349	0.060	0.232	0.244	0.037		
Sample length=200 quarters (1952–2001)											
1	0.045	0.042	0.042	0.057	0.054	0.054	0.059	0.054	0.054	0.056	0.012
4	0.112	0.077	0.044	0.109	0.092	0.051	0.095	0.084	0.046		
20	0.226	0.230	0.043	0.205	0.231	0.051	0.129	0.156	0.039		
Sample length=267 quarters (1935–2001)											
1	0.044	0.041	0.041	0.062	0.058	0.058	0.057	0.055	0.055	0.052	0.013
4	0.100	0.069	0.040	0.096	0.079	0.046	0.087	0.075	0.044		
20	0.188	0.186	0.039	0.173	0.194	0.048	0.110	0.138	0.038		

The table lists empirical size properties corresponding to a nominal size of 5% of Newey–West (1987) (N–W) with $k+1$ lags, Robust Hansen–Hodrick (1980) (Robust H–H) and Hodrick (1992) t-statistics. We examine a univariate regression of excess returns on dy^4 and a bivariate regression of excess returns on r and dy^4. We simulate 10,000 samples of various lengths from the Null 2 Model (constant expected excess returns) and record the percentage of observations greater than the nominal critical values under the null hypothesis of no predictability. The χ^2 tests report the proportion of rejections (using a 5% nominal size) for testing predictability jointly across horizons using Hodrick standard errors. In the case of the bivariate regression, joint predictability of both short rates and log dividend yields is considered across horizons. All horizons k are in quarters.

lag length is chosen. While the robust Hansen–Hodrick standard errors employ the correct rectangular filter, the estimate of the covariance matrix is not guaranteed to be invertible in small samples. In contrast, Hodrick uses covariance stationarity to remove the MA structure in the residuals. Hodrick's covariance estimator in Equation (3) is invertible and avoids the biased estimation of autocorrelations at long lags.

For Hodrick standard errors, the worst size distortion occurs at the one-quarter horizon with a 7% empirical size for a 5% nominal test on the short rate coefficient. For longer horizons, the Hodrick tests become slightly conservative, with the worst size distortion occurring for the dividend yield coefficient in the bivariate regression (a 3.7% empirical size for a 5% nominal test). The joint tests across horizons show few size distortions for the univariate regression, but are somewhat too conservative for the bivariate regression. In summary, the Hodrick standard errors display very satisfactory small sample properties that are far superior to those of Newey–West and Hansen–Hodrick standard errors.

5.3 Power

5.3.1 Power in one country. Table 11 summarizes the size-adjusted power for the Hodrick standard errors under Alternatives 1–3. Alternative 1 generates about the same predictive coefficient on the dividend yield as observed in the data but the predictive power of the short rate is unrealistically weak. Panel A shows that the power of the univariate dividend yield regression is very small for the shortest sample, only 25.5% (12.1%) at the one-quarter (20-quarter) horizon. For the longer samples, power rises to around 60% (40%) at the one-quarter (20-quarter) horizon. The results for the bivariate regressions are quite similar, with the power for the short rate coefficient slightly weaker than for the dividend yield. Hence, power is only satisfactory for the long samples, and even here, we might fail to reject the null of no predictability even though predictability is truly present.

Alternative 2, reported in Panel B, generates as much short rate predictability as in the data but slightly overpredicts the univariate predictive power of the dividend yield. For the 104-quarter length sample at short horizons, the power of the univariate dividend yield regression is satisfactory (55.2%). The power deteriorates with the horizon reaching 12.6% at five years. Power increases substantially with the sample size; for our longest sample, the Hodrick test has a power of over 97% at the one-quarter horizon. If predictability in the data is as strong as under Alternative 2, it is unlikely that we failed to detect univariate predictability. The power to detect the predictive ability of the short rate in the bivariate regression is very high for all samples and at all horizons. In the bivariate regression, there is little power to detect the true relation of future excess returns with the dividend yield, because the dividend yield coefficient is

Table 11
Size-adjusted power for Hodrick standard errors for one country

		Univariate regression	Bivariate regression		χ^2 tests joint across horizon	
Sample length (Qtrs)	Horizon k-qtrs,	dy^4	r	dy^4	Univariate	Bivariate
Panel A: power under alternative 1: $\delta_t = \mu + \phi\delta_{t-1} + u_t$						
104	1	0.255	0.201	0.283	0.325	0.335
	4	0.232	0.189	0.268		
	20	0.121	0.096	0.156		
200	1	0.625	0.415	0.650	0.520	0.557
	4	0.556	0.410	0.608		
	20	0.395	0.248	0.436		
267	1	0.596	0.344	0.596	0.632	0.691
	4	0.548	0.338	0.589		
	20	0.391	0.161	0.431		
Panel B: power under alternative 2: $\delta_t = \mu + \beta' X_t$						
104	1	0.552	0.636	0.075	0.559	0.036
	4	0.463	0.795	0.041		
	20	0.126	0.837	0.007		
200	1	0.892	0.795	0.091	0.891	0.042
	4	0.808	0.942	0.048		
	20	0.362	0.984	0.004		
267	1	0.971	0.853	0.094	0.967	0.048
	4	0.940	0.975	0.054		
	20	0.577	0.998	0.004		
Panel C: power under alternative 3: $\delta_t = \mu + \phi\delta_{t-1} + \beta' X_t + u_t$						
104	1	0.566	0.138	0.628	0.587	0.617
	4	0.484	0.090	0.562		
	20	0.159	0.016	0.198		
200	1	0.874	0.243	0.902	0.866	0.905
	4	0.775	0.181	0.859		
	20	0.417	0.021	0.483		
267	1	0.959	0.292	0.971	0.952	0.976
	4	0.923	0.225	0.951		
	20	0.620	0.024	0.682		

The table lists empirical power properties corresponding to a size-adjusted level of 5% of Hodrick (1992) t-statistics. We examine a univariate regression of excess returns on dy^4 and a bivariate regression of excess returns on r and dy^4. We simulate 10,000 samples of various lengths from Alternatives 1–3 (Panels A–C) and record the percentage of observations greater than the 5% critical values recorded under the Null 2 Model (constant expected excess returns) using the simulations in Table 10. The χ^2 tests report the proportion of rejections (using a 5% nominal size) for testing predictability jointly across horizons of both short rates and log dividend yields. All horizons k are quarterly.

very small in the presence of the short rate as a predictor. Consistent with this, the univariate joint χ^2 tests across horizons is very powerful, but the power of the test in the bivariate regression is minimal.

Panel C reports the power under Alternative 3, which fits the data the best. The power of the univariate tests is slightly better than the power under Alternative 2. For the bivariate regression, the predictive coefficient of the short rate (dividend yield) is underestimated (overestimated) relative to the data coefficients under Alternative 3. Consequently, the power to detect the true predictive relation with the short rate is low at long horizons (because the predictability through the short rate is short-lived), whatever the sample size is. Power at a one-quarter horizon rises from 13.8% for the shortest sample to 29.2% for the longest sample. Fortunately, the power to detect predictability through the dividend yield in the bivariate system is quite high: power is 62.8% (97.1%) for the shortest (longest) sample at a one-quarter horizon. For a 20-quarter horizon, we need longer samples to obtain a powerful test (power is only 19.8% for the 1975–2001 sample but is 68.2% for the 1935–2001 sample). With such high power, it is unlikely that we would have failed to uncover predictability by the dividend yield. Overall, the joint tests are quite powerful (at least 58.7% power). We conclude that lack of power is unlikely to drive our failure to find predictability of excess returns by dividend yields, particularly for the longer US sample, unless the true predictability in the data is quite weak to begin with.

5.3.2 Power pooling cross-country data To examine the increase in power that pooling cross-country data allows, our data generating process must match the empirical correlations of excess returns across countries in the data. For the United States, United Kingdom, France, and Germany, the cross-country correlations for excess returns are all around 0.50 (Table 1). To account for the cross-sectional correlation, we modify the present value model in the following fashion. First, we consider each country to be a separate draw of (X_t, δ_t), using Alternative 3. Second, we specify the process for X_t in each country to be independent but allow the discount rates for different countries to be correlated. Specifically, we allow the shocks u_t^i in the discount rate (8) for country i to be correlated with the discount rate shocks u_t^j for country j. We set this correlation at 0.80, which produces an unconditional correlation coefficient of 0.535 between the implied excess equity returns for any two countries.

To compare the increase in power pooling international data, Table 12 reports power under Alternative 3 for a cross-sectional panel of N countries and compares it to increasing the short US sample by N times. Pooling data by adding one other country turns out to be only slightly worse than doubling the sample size of the United States. For four countries, the pooled test is better than using a sample of 312 quarters, which is larger than the longest US sample we have. Hence, pooling information across countries should increase the confidence in our results more than the US long sample does from a pure statistical power perspective.

Table 12
Power properties for pooling cross-country data

	One country				Multiple countries			
	Univariate regression	Bivariate regression			Univariate regression	Bivariate regression		
k-qtrs	dy^4	r	dy^4	χ^2 test	dy^4	r	dy^4	χ^2 test
Sample length=104 qtrs								
1	0.575	0.096	0.612	0.487				
4	0.509	0.058	0.536	0.412				
20	0.201	0.018	0.169	0.133				
Sample length = 208 qtrs					Two countries			
1	0.891	0.193	0.913	0.839	0.887	0.143	0.905	0.835
4	0.847	0.115	0.878	0.779	0.827	0.070	0.847	0.747
20	0.503	0.008	0.499	0.350	0.326	0.009	0.327	0.227
Sample length = 312 qtrs					Three countries			
1	0.980	0.289	0.986	0.967	0.974	0.218	0.978	0.963
4	0.962	0.179	0.977	0.941	0.952	0.111	0.963	0.922
20	0.745	0.010	0.762	0.603	0.556	0.009	0.590	0.446
Sample length = 416 qtrs					Four countries			
1	0.998	0.387	0.999	0.995	0.995	0.293	0.996	0.991
4	0.993	0.255	0.996	0.989	0.988	0.153	0.991	0.978
20	0.881	0.012	0.898	0.790	0.739	0.013	0.767	0.652

The table lists empirical power properties of Hodrick (1992) t-statistics, comparing a sample of one country of increasing length versus a cross-sectional panel of countries, each of length 104 quarters. (The number of observations 104 quarters corresponds to the sample period 1975–2001.) Power is taken corresponding to a nominal (asymptotic) size of 5% using Alternative 3. We examine a univariate regression of excess returns on log dividend yields and a bivariate regression of excess returns on short rates and log dividend yields. The χ^2 test reports a test for the joint predictability of the short rate and log dividend yield for a given horizon. The population correlation coefficient between the excess returns of any two countries is 0.535.

6. The Predictive Power of the Earnings Yield

Our findings so far suggest that dividend yields have marginal predictive power for returns within a bivariate regression, with the short rate dominating the dividend yield. Campbell and Shiller (1988a) and Lamont (1998) claim that the earnings yield has information over and above the dividend yield in capturing predictable components in returns.

6.1 The earnings yield and excess return predictability

The first column of Panel A, in Table 13 reports a univariate regression with the earnings yield as the regressor. The results are similar to what we found for the dividend yield regression in Table 2. This suggests that the weak univariate relation between returns and yield variables primarily comes from the price in the denominator of both variables.

To allow comparison with Lamont (1998), we report a bivariate regression of excess returns on log dividend and log earnings yields. Lamont

Table 13
Predictability of excess returns by earnings yields

		Univariate regression	Lamont regression		Trivariate regression			χ^2 tests	
	k-qtrs	ey^4	dy^4	ey^4	r	dy^4	ey^4	Lamont	Trivariate
Panel A: US quarterly S&P Data									
1935–2001	1	0.0665	0.1842	−0.1019	−1.0334	0.1000	−0.0168	0.183	0.119
		(1.258)	(1.330)	(−0.741)	(−1.296)	(0.605)	(−0.103)		
	4	0.1018	0.1119	0.0012	−0.7759	0.0472	0.0654	0.108	0.106
		(1.961)*	(0.981)	(0.011)	(−1.084)	(0.377)	(0.556)		
	20	0.0719	0.1111	−0.0092	−0.4190	0.0666	0.0291	0.347	0.260
		(1.415)	(0.946)	(−0.128)	(−0.682)	(0.486)	(0.361)		
1952–2001	1	0.0646	0.1948	−0.1113	−2.6243	0.0337	0.1272	0.192	0.006**
		(0.993)	(1.557)	(−0.873)	(−2.854)**	(0.239)	(0.815)		
	4	0.0798	0.1615	−0.0636	−1.8146	0.0479	0.1029	0.277	0.056
		(1.218)	(1.195)	(−0.450)	(−2.063)*	(0.687)	(0.687)		
	20	0.0223	0.1458	−0.0765	−0.3548	0.1199	−0.0418	0.854	0.872
		(0.320)	(0.561)	(−0.520)	(−0.489)	(0.432)	(−0.249)		
1935–1990	1	0.0896	0.3848	−0.1800	−0.6544	0.3051	−0.1125	0.049*	0.050*
		(1.304)	(1.849)	(1.062)	(−0.696)	(1.112)	(−0.499)		
	4	0.1473	0.2869	−0.0532	−0.4708	0.2300	−0.0046	0.007**	0.018*
		(2.309)*	(1.872)	(−0.411)	(−0.031)	(1.336)	(−0.031)		
	20	0.1178	0.1916	−0.0139	−0.4239	0.1423	0.0297	0.017*	0.041*
		(2.392)*	(1.874)	(−0.181)	(−0.714)	(1.597)	(0.431)		
1952–1990	1	0.1028	0.9349	−0.5240	−2.0892	0.6798	−0.2418	0.000**	0.000**
		(1.168)	(3.710)**	(−2.618)**	(−2.096)*	(2.338)*	(−0.949)		
	4	0.1287	0.8201	−0.4212	−1.3320	0.6565	−0.2392	0.000**	0.001**
		(1.521)	(3.234)**	(−2.080)*	(−1.376)	(2.303)*	(−0.960)		
	20	0.0798	0.4167	−0.1999	−0.3068	0.3793	−0.1570	0.010**	0.028*
		(1.133)	(2.685)**	(−1.570)	(−0.374)	(2.188)*	(−0.991)		

Table 13
(continued)

		Univariate regression	Lamont regression		Trivariate regression			χ^2 tests	
	k-qtrs	ey^4	dy^4	ey^4	r	dy^4	ey^4	Lamont	Trivariate
		Univariate regression	Lamont regression		Trivariate regression			χ^2 tests	
	k-mths	ey^{12}	dy^{12}	ey^{12}	r	dy^{12}	ey^{12}	Lamont	Trivariate
Panel B: pooled-country monthly MSCI data									
1975–2001	1	0.0251	0.0730	−0.0192	−1.82d23	0.1609	0.0035	0.612	0.000**
		(0.498)	(0.969)	(−0.366)	(−2.792)**	(2.001)*	(0.070)		
	12	−0.0098	0.0994	−0.0680	−1.0401	0.1493	−0.0548	0.385	0.000**
		(−0.195)	(1.178)	(−1.280)	(−1.956)	(1.697)	(−1.116)		
	60	−0.0369	0.0877	−0.0710	0.2956	0.0686	−0.0739	0.132	0.321
		(−0.746)	(0.882)	(−1.980)*	(0.731)	(0.774)	(−2.370)*		

We estimate regressions of the form $\tilde{y}_{t+k} = \alpha + z_t^{'}\beta + \epsilon_{t+k,k}$, where $\tilde{y}_{t+k}$ is the cumulated and annualized k-period ahead return, with the instruments z_t being log earnings yields (univariate regression), or log dividend yields and log earnings yields (Lamont bivariate regression), or risk-free rates, log dividend yields, and log earnings yields (trivariate regression). T-statistics in parentheses are computed using Hodrick (1992) standard errors. For Panel A (B), horizons k are quarterly (monthly). Panel B pools coefficients jointly across the United States, United Kingdom, France, and Germany, constraining the coefficients to be the same across countries. The χ^2 test columns report a p-value for a test that all the coefficients in each regression are jointly equal to zero.
*p<0.05.
*p<0.01.

finds a positive coefficient on the dividend yield and a negative coefficient on the earnings yield. He argues that the predictive power of the dividend yield stems from the role of dividends in capturing permanent components of prices, whereas the negative coefficient on the earnings yield is due to earnings being a good measure of business conditions. Table 13 summarizes that the coefficients over the long US sample, while having the same sign found by Lamont, are insignificant and not one joint test of the predictive power of dividend and earnings yields is significant at the 10% level. Only when the 1990s are excluded, as in the 1952–1990 sample similar to Lamont's paper, do we find significant coefficients for dividend and earnings yields. When we add the short rate as a predictor in a trivariate regression of excess returns on risk-free rates, dividend and earnings yields, the coefficients on dividend and earnings yields remain insignificantly different from zero, and the sign on the earnings yield is fragile. For the post-1952 samples, the short rate predictive power remains robust in the presence of the earnings yield.

Panel B of Table 13 pools all the specifications across the United States, United Kingdom, France, and Germany. There are no significant coefficients in the univariate regression. The bivariate specification preserves the Lamont coefficient pattern with the earnings yield coefficient now significantly negative at the five-year horizon. However, joint χ^2 predictability tests fail to reject the null of no predictability. In the trivariate specification, the short rate and dividend yield coefficients have similar signs and significance to the results from the bivariate regressions in Section 2. The earnings yield coefficient is significantly negative only at the five-year horizon, but a joint test across all predictors at the five-year horizon fails to reject the null of no predictability. For individual countries, the Lamont coefficient pattern is neither robust nor significant (data not shown). In conclusion, there is little evidence that earnings yields predict excess returns. The earnings yields coefficients are not robust across different sample periods or countries.

6.2 The earnings yield and cash-flow predictability

In this section, we use both dividend and earnings yields to predict cash flows, as measured by dividend growth and earnings growth. Panel A of Table 14 reports the results for dividend growth. In the univariate regression, the earnings yield is a better predictor of dividend growth than the dividend yield. The coefficients are larger in magnitude than the dividend yield coefficients, and, except for the coefficient in the 20-quarter regression, significantly different from zero. The earnings yield coefficients are all positive.

The bivariate dividend yield–earnings yield regression shows an intriguing result. The Lamont (1998) pattern of a positive dividend yield coefficient and a negative earnings yield coefficient is reversed, and highly significant, for the cash-flow regressions! At short horizons, both

Table 14
Predictability of cash-flow growth by earnings yields

			Univariate regression	Lamont regression		Trivariate regression			χ^2 tests	
		k	ey^4	dy^4	ey^4	r	dy^4	ey^4	Lamont	Trivariate
Panel A: predictability of dividend growth										
US quarterly S&P data										
	1935–2001	1	0.0594	−0.1897	0.2328	−0.9766	−0.2694	0.3132	0.002**	0.001**
			(2.936)**	(−2.724)**	(3.175)**	(−3.527)**	(−3.412)**	(3.785)**		
		4	0.0370	−0.1494	0.1713	−0.8333	−0.2188	0.2402	0.031*	0.013**
			(2.061)*	(−2.021)*	(2.377)*	(−2.877)**	(−2.559)*	(2.874)**		
		20	0.0062	−0.0456	0.0395	−0.1428	−0.0608	0.0526	0.106	0.194
			(0.493)	(−1.260)	(1.985)*	(−0.835)	(−1.310)	(1.821)		
Pooled-country monthly MSCI data										
	1975–2001	1	0.0214	−0.0792	0.0695	0.2059	−0.0891	0.0669	0.002**	0.004**
			(1.183)	(−2.741)**	(3.396)**	(0.775)	(−2.783)**	(3.365)**		
		12	0.0047	−0.0332	0.0241	−0.3872	−0.0146	0.0290	0.447	0.161
			(0.254)	(−1.063)	(1.216)	(−1.769)	(−0.440)	(1.527)		
		60	−0.0132	0.0090	−0.0167	−0.0447	0.0118	−0.0163	0.413	0.580
			(−0.570)	(0.277)	(−1.319)	(−0.262)	(0.308)	(−1.355)		
Panel B: predictability of earnings growth										
US quarterly S&P data										
	1935–2001	1	−0.0582	0.0970	−0.1468	−0.0355	0.0941	−0.1439	0.391	0.550
			(−1.120)	(0.681)	(−1.079)	(−0.049)	(0.582)	(−0.895)		
		4	−0.0822	0.1994	−0.2613	−0.1563	0.1864	−0.2484	0.026*	0.057
			(−1.791)	(1.467)	(−2.220)*	(−0.223)	(1.235)	(−1.824)		

	20	−0.0674	0.0876	−0.1313	0.1797	0.1066	−0.1478	0.007**	0.004**
		(−1.457)	(0.659)	(−1.901)	(0.357)	(0.671)	(−1.695)		
Pooled-country monthly MSCI data									
1975–2001	1	−0.2767	0.2723	−0.4419	−0.2867	0.2861	−0.4383	0.000**	0.000**
		(−8.381)**	(3.566)**	(−5.528)**	(−0.743)	(3.800)**	(−5.109)**		
	12	−0.3373	0.3452	−0.5392	−1.9369	0.4381	−0.5148	0.000**	0.000**
		(−9.787)**	(4.334)**	(−7.153)**	(−5.473)**	(5.627)**	(−6.883)**		
	60	−0.1534	0.2332	−0.2440	0.0974	0.2269	−0.2449	0.000**	0.000**
		(−3.321)**	(2.356)*	(−5.357)**	(0.3403)	(2.565)*	(−6.053)**		

We estimate regressions of the form $\tilde{y}_{t+k} = \alpha + z_t^{'}\beta + \epsilon_{t+k,k}$, where $\tilde{y}_{t+k}$ is the cumulated and annualized k-period dividend or earnings growth, with the instruments z_t being log earnings yields (univariate regression), or log dividend yields and log earnings yields (Lamont bivariate regression), or risk-free rates, log dividend yields, and log earnings yields (trivariate regression). T-statistics in parentheses are computed using Hodrick (1992) standard errors. For the quarterly US S&P data, horizons k are quarterly, whereas for the monthly MSCI data horizons k are monthly. The pooled-country panels pool coefficients jointly across the United States, United Kingdom, France, and Germany, constraining the coefficients to be the same across countries. The χ^2 test columns report a p-value for a test that all the coefficients in each regression are jointly equal to zero.
*$p<0.05$.
**$p<0.01$.

coefficients are different from zero at the 1% level and the joint test also rejects at the 1% level. The coefficients become smaller, and the statistical significance weakens, with longer horizons. We still find predictability at the 5% level for the one-year horizon, but only the earnings yield is significant (at the 5% level) for the five-year horizon. Adding the risk-free rate in the trivariate regression does not change this picture very much, but a high short rate is also a very strong signal of lower future dividend growth at the one- and four-quarter horizons. One possible interpretation is that since interest rates are high during recessions and recessions are persistent (Ang and Bekaert, 2002), high interest rates predict low future cash flows. The negative dividend yield coefficient can arise from prices reflecting future positive cash-flow prospects.

We find that the negative dividend yield and positive earnings yield coefficients for predicting dividend growth are very robust across different subsamples, but we do not report these results to conserve space. In particular, omitting the 1990s, the signs of the coefficients are the same, but the *t*-statistics are even larger in magnitude than the those reported for the full sample. The inverse Lamont pattern is also robust in the subsamples beginning in 1952.

When we pool data across countries using MSCI data, the cross-sectional variation in the coefficients makes the cash-flow Lamont pattern weaker, but it remains significant at the 1% level at the one-month horizon, in both the bivariate and trivariate regressions. At longer horizons, earnings and dividend yields do not predict future cash flows. Moreover, the pattern is also repeated in the coefficients for each individual country. Internationally, the short rate is not a robust predictor of future cash flows, perhaps because the cyclicality of interest rates is not consistent across countries.

In Panel B of Table 14, we repeat the same regressions for earnings growth. The univariate regression with the earnings yield delivers a negative coefficient. The sign of this point estimate could be potentially consistent with a standard price effect. However, it is statistically insignificant. Interestingly, the sign of the dividend and earnings yield coefficients are reversed for earnings growth, compared to dividend growth: for earnings growth, we see a positive (negative) coefficient on the dividend (earnings) yield. The effect is strongest for the earnings yield, which is significant at the 5% level for the one- and five-year horizons, although borderline in the latter case. Looking at the joint χ^2 tests, we conclude that there is some evidence of cash-flow growth predictability at longer horizons (the *p*-value is at most 0.057), primarily driven by the earnings yield. The 1952–2001 sample also preserves these coefficient patterns.

When we pool data across countries, the coefficients increase in magnitude and the statistical significance increases considerably. All the joint tests now reject at the 1% level, and even the short rate predicts cash-flow growth significantly at the one-year horizon. The strong international results are due to the fact that in each individual country, the dividend yield (earnings yield) coefficients are positive (negative).

We conclude that high dividend yields signal high future earnings growth and high earnings yields signal low future earnings growth rates. It is conceivable that the second effect is a genuine price effect (higher prices in response to predicted rises in future earnings), but the first effect is harder to interpret. The puzzling nature of our findings becomes more apparent when we look at the earnings and dividend growth simultaneously. We find the following sign pattern:

	Dividend growth	Earnings growth	Payout ratio
Dividend yield	−ve	+ve	−ve
Earnings yield	+ve	−ve	+ve

We can derive the results in the last column because the change in the payout ratio equals (logarithmic) dividend growth minus earnings growth. Hence, our results imply that higher dividend yields (higher earnings yields) strongly predict lower (higher) payout ratios tomorrow.

Can we explain the reverse patterns for the two cash-flow measures for dividend and earnings yields used jointly in predictive regressions? We can rule out a Lamont (1998) story translated to cash flows. According to Lamont, dividend yields capture price effects, whereas earnings yields capture the cyclical component in earnings and hence potentially also risk aversion. Under this scenario, we would expect the dividend yield to be negatively related to earnings growth (the usual positive cash flow prospects), but we find a positive effect.

When dividend yields are high today, we predict low dividend growth in the future because payout ratios strongly decrease. This may be the result of dividend smoothing, or it may reflect prices anticipating higher growth opportunities that decrease the payout ratio. The positive relation between current high dividend yields and future earnings growth implies that these growth opportunities do not rapidly translate into higher future earnings. The negative relation between the current earnings yield and future earnings may be consistent with either a price effect or mean reversion in earnings. The payout ratio reacts positively to an increase in the earnings yield. In the mean reversion story, this could be an artifact of dividend smoothing. In the price story, lower prices today may reflect poor future earnings and poor future growth opportunities. The poor growth opportunities may increase the payout ratio, particularly if dividends are sticky.

All in all, we find very strong evidence of dividend growth predictability, both in US and international data, by using the dividend and earnings yield jointly as predictors. We find weaker evidence in US data, but very strong international evidence, of earnings growth predictability by dividend and earnings yields. A challenge for future work is to create a present value model with sophisticated dynamics for earnings growth, payout ratios, and dividend growth to match this evidence.

7. Conclusion

The predictable components in equity returns uncovered in empirical work over the last 30 years have had a dramatic effect on finance research. Theoretical equilibrium models try to match the predictability evidence as a stylized fact. The partial equilibrium dynamic asset allocation literature investigates the impact of the predictability on hedging demands. Much of the focus has been on the predictive prowess of the dividend yield, especially at long horizons. In this article, we pose the question whether this predictability exists. After carefully accounting for small sample properties of standard tests, our answer is surprising but important. At long horizons, excess return predictability by the dividend yield is not statistically significant, not robust across countries, and not robust across different sample periods. In this sense, the predictability that has been the focus of most recent finance research is simply not there.

Nevertheless, we do find that stock returns are predictable, calling for a refocus of the predictability debate in four directions. First, our results suggest that predictability is mainly a short-horizon, not a long-horizon, phenomenon. The predictive ability of the dividend yield is best seen in a bivariate regression with short rates only at short horizons. Second, the strongest predictability comes from the short rate rather than from the dividend yield. The result that the short rate predicts equity returns goes back to at least Fama and Schwert (1977), but somehow recent research has failed to address what might account for this predictability and has mostly focused on the dividend yield as an instrument. Third, high dividend yields predict high future interest rates. Finally, dividend and earnings yields have good predictive power for future cash-flow growth rates but not future excess returns. Hence, a potentially important source of variation in price-earnings and price-dividend ratios is the predictable component in cash flows. Our results generally imply that univariate linear models of expected returns are unlikely to satisfactorily capture all the predictable components in returns.

After our results were first distributed in a working paper article (Ang and Bekaert, 2001), a number of articles have been written that confirm them. Campbell and Yogo (2006) develop a new inference methodology within the linear regression framework of Stambaugh (1999) and find that the predictive power of the dividend yield is considerably weakened but that the predictive power of the short rate is robust. Lettau and Ludvigson (2005) also find that the price-dividend ratio weakly forecasts excess returns but confirm that future dividend growth, long ignored by the literature, is predictable. Engstrom (2003) confirms our findings that univariate dividend yield regressions have difficulty capturing all the predictable components in returns, by constructing economies where the dividend yield is a noisy predictor of both excess returns and cash-flow growth.

We hope that our results will, in the short run, affect the asset allocation literature, which often has taken the predictive power of the dividend yield in a univariate regression as a stylized fact, and in the longer run will stimulate research on theoretical models that might explain the predictability patterns we demonstrate, particularly return predictability by the short rate at short horizons and the joint predictability of cash flows and excess returns. Finally, future research should also reconcile the weak out-of-sample evidence of predictability (Bossaerts and Hillion 1999, Goyal and Welch 2003, 2004) with the in-sample evidence of predictability. While such a reconciliation may require models with structural breaks or regime shifts (Timmermann and Paye 2006), we note one last interesting result derived from our US sample. Mimicking Goyal and Welch's (2004) procedure for out-of-sample forecasting over the post-1964 period, we find that the bivariate predictive regression with the short rate and dividend yield produces a lower root mean squared error than the historical mean for forecasting excess returns at a one-quarter horizon but not at long horizons. Consequently, for the bivariate predictive regression, the in-sample and out-of-sample evidence of return predictability is consistent.

Appendix A: Robust Hansen–Hodrick (1980) Standard Errors

Using GMM, the parameters $\theta = (\alpha\, \beta'_k)'$ in Equation (2) have an asymptotic distribution $\sqrt{T}(\hat{\theta} - \theta) \stackrel{a}{\sim} N(0,\Omega)$ where $\Omega = Z_0^{-1} S_0 Z_0^{-1}$, $Z_0 = \mathrm{E}(x_t x_t')$, and $x_t = (1\ z'_t)'$. We estimate S_0 by

$$\hat{S}_0 = C(0) + \sum_{j=1}^{k-1} [C(j) + C(j)'], \tag{A1}$$

where

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

and $w_{t+k} = \varepsilon_{t+k,k} x_t$. This estimator of S_0 is not guaranteed to be positive semi-definite. If it is not, we use a Newey–West (1987) estimate of S_0 with k lags. Note that for $k = 1$, the robust Hansen–Hodrick (1980) and Hodrick (1992) standard errors are identical.

Appendix B: Testing Predictability Across Horizons

To test whether the predictability coefficients are statistically significant across n horizons $k_1 \ldots k_n$, we set up the simultaneous equations:

$$\begin{aligned} \tilde{y}_{t+k_1} &= \alpha_{k_1} + \beta'_{k_1} z_t + u_{t+k_1} \\ &\vdots \\ \tilde{y}_{t+k_n} &= \alpha_{k_n} + \beta'_{k_n} z_t + u_{t+k_n}. \end{aligned} \tag{B1}$$

We produce an estimate $\hat{\beta}$ of $\beta = (\alpha_{k_1}\beta'_{k_1} \ldots \alpha_{k_n}\beta'_{k_n})'$ by performing OLS on each equation.

The moment conditions for the system in Equation (B1) are

$$\mathrm{E}(h_{t+\bar{k}}) \equiv \mathrm{E}\begin{pmatrix} h_{t+k_1} \\ \vdots \\ h_{t+k_n} \end{pmatrix} = \mathrm{E}\begin{pmatrix} u_{t+k_1} x_t \\ \vdots \\ u_{t+k_n} x_t \end{pmatrix} = \mathrm{E}(u_{t+\bar{k}} \otimes x_t) = 0, \tag{B2}$$

where $x_t = (1\ z'_t)'$, a $K \times 1$ vector, and $u_{t+\bar{k}} = (u_{t+k_1} \ldots u_{t+k_n})'$.

From standard GMM, $\sqrt{T}(\hat{\theta} - \theta) \overset{d}{\rightarrow} N(0,\Omega)$, with $\Omega = Z_0^{-1} S_0 Z_0^{-1}$, $Z_0 = (I_n \otimes \mathrm{E}(x_t x'_t))$, and

$$S_0 = \mathrm{E}(h_{t+\bar{k}} h'_{t+\bar{k}}) = \mathrm{E}\left[(u_{t+\bar{k}} u_{t+\bar{k}}) \otimes (x_t x'_t)\right]. \tag{B3}$$

The Hodrick (1992) estimate $\hat{S}_T^b$ of S_0 is given by

$$\hat{S}_T^b = \frac{1}{T} W'W, \tag{B4}$$

where W is a $T \times Kn$ matrix, $W = (W_{k_1} \ldots W_{k_n})$, and W_k, $T \times n$, is given by $W_k = (w'_{1+k}, \ldots w'_{T+k})$, and w_{t+k}, $K \times 1$, is

$$w_{t+k} = e_{t+1}\left(\sum_{i=0}^{k-1} x_{t-i}\right), \tag{B5}$$

since under the null of no predictability the one-step ahead errors $e_{t+i} = u_{t+1}$ are uncorrelated and $u_{t+k} = e_{t+1} + \ldots + e_{t+k}$. Denoting $X = (x'_1, \ldots x'_T)$, $T \times K$, an estimate of Z_0^{-1} is given by

$$\hat{Z}_T^{-1} = \frac{1}{T}[I_n \otimes (X'X)^{-1}]. \tag{B6}$$

To test the hypothesis $C\beta = 0$, we use the Newey (1985) χ^2 test:

$$(C\hat{\beta}')[C\hat{\Omega}C']^{-1} C\hat{\beta} \sim \chi^2_{\mathrm{rank}(C)}, \tag{B7}$$

with $\hat{\Omega} = \hat{Z}_T^{-1} \hat{S}_T^b \hat{Z}_T^{-1}$.

Appendix C: Testing Predictability Pooling Cross-Sectional Information

C.1 Generalizing Hodrick (1992) to Cross-sectional regressions

To pool cross-sectional country information, we estimate the system

$$\tilde{y}^i_{t+k} = \alpha_i + \beta'_i z^i_t + u^i_{t+k} \tag{C1}$$

for $i = 1 \ldots N$ countries, subject to the restriction $\beta_i = \bar{\beta} \;\; \forall i$, but impose no restrictions on α_i across countries. We take $i =$ United States, United Kingdom, France, and Germany.

Let the dimension of z_t be $(K-1)$ so that there will be a total of K regressors, including the constant terms α_i for each of N countries. In Equation (C1), we denote the free parameters $\theta = (\alpha_1 \ldots \alpha_N \bar{\beta}')'$ and the unrestricted parameters stacked by each equation $\beta = (\alpha_1 \beta'_1 \ldots \alpha_N \beta'_N)'$. We can estimate the system in Equation (C1) subject to the restriction that $C\beta = 0$, where C is a $NK \times (N-1)(K-1)$ matrix of the form:

$$C = \begin{pmatrix} \tilde{0} & I & \tilde{0} & -I & \tilde{0} & \ldots & & \\ \tilde{0} & O & \tilde{0} & I & \tilde{0} & -I & \ldots & \\ \vdots & & & & & & & \\ \tilde{0} & O & \tilde{0} & & & & \tilde{0} & -I \end{pmatrix}, \tag{C2}$$

where $\tilde{0}$ is a $(K-1) \times 1$ vector of zeros, O is a $(K-1) \times (K-1)$ matrix of zeros, and I is a $(K-1)$ rank identity matrix.

Denote

$$\begin{aligned} \tilde{y}_{t+k} &= (\tilde{y}^1_{t+k} \ldots \tilde{y}^N_{t+k})' && (N \times 1) \\ x^i_t &= (1 z^{i'}_t) && (K \times 1) \\ u_{t+k} &= (u^1_{t+k} \ldots u^N_{t+k})' && (N \times 1) \\ X_t &= \begin{pmatrix} x^1_t & & 0 \\ & \ddots & \\ 0 & & x^N_t \end{pmatrix} && (NK \times N). \end{aligned} \tag{C3}$$

Then, the system can be written as

$$\tilde{y}_{t+k} = X'_t \beta + u_{t+k}, \tag{C4}$$

subject to $C\beta = 0$. Let $Y = (\tilde{y}'_{1+k} \ldots \tilde{y}'_{T+k})'$, $X = (X'_1 \ldots X_T)'$, and $U = (u'_{1+k} \ldots u'_{T+k})'$. Then, the compact system can be written as

$$Y = X\beta + U, \quad \text{subject to } C\beta = 0. \tag{C5}$$

A consistent estimate $\hat{\beta}$ of β is given by

$$\hat{\beta} = \beta^{ols} - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C\beta^{ols}, \tag{C6}$$

with $\beta^{ols} = (X'X)^{-1}X'Y$. This gives us a consistent estimate $\hat{\theta}$ of θ.

The moment conditions of the system in Equation (C4) are

$$\mathrm{E}(h_{t+k}) = \mathrm{E}(X_t u_{t+k}) = 0.$$

By standard GMM, $\hat{\theta}$ has distribution

$$\sqrt{T}(\hat{\theta} - \theta) \overset{a}{\sim} N(0,(D'_0 S_0^{-1} D_0)^{-1}), \tag{C7}$$

with

$$D'_0 = \mathrm{E}\left(\frac{\partial h_{t+k}}{\partial \theta'}\right)$$

and

$$S_0 = \mathrm{E}(h_{t+k}h'_{t+k}).$$

The Hodrick (1992) estimate $\hat{S}_T^b$ of S_0 is given by

$$\hat{S}_T^b = \frac{1}{T}\sum_{t=k}^{T} wk_t wk'_t, \tag{C8}$$

where $wk_t(NK \times 1)$ is given by

$$wk_t = \left(\sum_{i=0}^{k-1} X_{t-i}\right) e_{t+1}.$$

Under the null hypothesis of no predictability, $u_{t+k} = e_{t+1} + \ldots e_{t+k}$, where e_{t+1} are the one-step ahead serially uncorrelated errors. This is the SUR equivalent of the Hodrick (1992) estimate for univariate OLS regressions.

An estimate $\hat{D}_T$ of D_0 is given by

$$\hat{D}'_T = \frac{1}{T}\sum_{t=0}^{T} \frac{\partial h_{t+k}}{\partial \theta'},$$

where $\theta = (\alpha_1 \ldots \alpha_N \bar{\beta}')$ and

$$-\frac{\partial h_{t+k}}{\partial \theta'} = \begin{bmatrix} 1 & z_t^{1'} & & & & 0 & \\ & & 1 & z_t^{2'} & & & \\ & & & & \ddots & & \\ & & 0 & & & 1 & z_t^{N'} \\ z_t^1 & z_t^1 z_t^{1'} & z_t^2 & z_t^2 z_t^{2'} & \ldots & z_t^N & z_t^N z_t^{N'} \end{bmatrix}.$$

The estimate $\hat{\theta}$ has the distribution

$$\sqrt{T}(\hat{\theta}-\theta) \overset{a}{\sim} N(0,[\hat{D}'_T(\hat{S}^b_T)^{-1}\hat{D}_T]^{-1}). \qquad (C9)$$

There are $(N+K-1)$ free parameters in θ with NK moment conditions. This gives $NK-(N+K-1)$ overidentifying restrictions. The Hansen (1982) χ^2 J-test of overidentifying restrictions is given by

$$J = T(\bar{h}'(\hat{S}^b_T)^{-1}\bar{h}) \sim \chi^2[NK-(N+K-1)], \qquad (C10)$$

with

$$\bar{h} = \frac{1}{T}\sum_{t=0}^{T} h_{t+k}.$$

C.2 Generalizing Cochrane-Orcutt to cross-sectional regressions

We start with the one-step ahead predictive regression in Equation (C4) with $k=1$, repeated here for convenience:

$$\tilde{y}_{t+1} = X_t'\beta + u_{t+1}, \qquad (C11)$$

subject to $C\beta = 0$, where C is given by Equation (C2). We assume that the $N \times 1$ vector or errors u_{t+1} follows the process:

$$u_{t+1} = \Phi u_t + \varepsilon_{t+1}, \qquad (C12)$$

where ε_t are IID with $\mathrm{E}(\varepsilon_t \varepsilon'_t) = \Omega$. The unconditional covariance matrix of u_t is given by:

$$\mathrm{E}(u_t u'_t) = \mathrm{devech}[(I - \Phi \otimes \Phi)^{-1}\mathrm{vec}(\Omega)].$$

We write Equation (C11) in terms of uncorrelated residuals:

$$\tilde{y}_{t+1} - \Phi\tilde{y}_{t+1} = (X_t - X_{t-1}\Phi')'\beta + \varepsilon_{t+1},$$

or as

$$y^*_{t+1} = (X^*_t)'\beta + \varepsilon_{t+1}, \qquad (C13)$$

where $y^*_{t+1} = \tilde{y}_{t+1} - \Phi\tilde{y}_t$ and $X^*_t = X_t - X_{t-1}\Phi'$. To construct y^*_{t+1} and X^*_t, we use a consistent estimate, $\hat{\Phi}$, of Φ. Using the estimate $\hat{\beta}$ of β in Equation (C6), we set $\hat{\Phi} = 1/T\sum v_t v'_{t-1}$, where v_t are the residuals $v_t = \tilde{y}_{t+1} - X'_t\hat{\beta}$ that are standardized to have unit variance.

To compute Cochrane-Orcutt standard errors for $\theta = (\alpha_1 \ldots \alpha_N \bar{\beta}')$ in the pooled system (C11) with matrix-autocorrelated residuals, we use GMM. The set of $NK \times 1$ moment conditions implied by the regression (C13) are

$$\mathrm{E}(h_{t+1}) = \mathrm{E}(X_t^* \varepsilon_{t+1}) = 0.$$

We set the estimate $\hat{S}_T$ of $S_0 = \mathrm{E}(h_{t+1}h'_{t+1})$ to be

$$\hat{S}_T = \frac{1}{T}\sum_{t=1}^{T}[X_t^*(y_{t+1}^* - X_t^{*\prime}\hat{\beta})].$$

An estimate, $\hat{D}_T$, of the derivative of the moment conditions with respect to θ can be computed by taking, and summing, appropriate elements of $X_t^* X_t^{*\prime}$.

Appendix D: Proof of Proposition

The present value relation (5) can be written as

$$\frac{P_t}{D_t} = \sum_{i=1}^{\infty} M_{t+i},$$

where

$$M_{t+i} = \mathrm{E}_t\left[\exp\left(-\sum_{j=0}^{i-1}\delta_{t+j} + \sum_{j=1}^{i} g_{t+j}^d\right)\right]. \tag{D1}$$

We show that

$$M_{t+i} = \exp(a_i + b'_i X_t + c_i \delta_t),$$

which then proves relation (9).

The initial conditions are given by

$$\begin{aligned}\mathrm{E}_t[\exp(-\delta_t + g_{t+1}^d)] &= \exp(-\delta_t)\mathrm{E}_t[\exp(e_2' X_{t+1})]\\ &= \exp\left(-\delta_t + e_2'\mu + e_2'\Phi X_t + \frac{1}{2}e_2'\Sigma e_2\right),\end{aligned} \tag{D2}$$

where $e_2 = (0, 1)'$ and equating coefficients yields Equation (11).

To prove the recursive relations (10), we use proof by induction. Suppose that $M_{t+i} = \exp(a_i + b'_i X_t + c_i \delta_t)$. Then, we can write:

$$\begin{aligned}M_{t+i+1} &= \mathrm{E}_t\left[\exp\left(-\sum_{j=0}^{i}\delta_{t+j} + \sum_{j=1}^{i+1} g_{t+j}^d\right)\right]\\ &= \mathrm{E}_t[\exp(-\delta_t + g_{t+1}^d)\exp(a_i + b'_i X_{t+1} + c_i\delta_{t+1})]\\ &= \exp\Big(a_i + c_i\alpha + (c_i\phi - 1)\delta_t + \frac{1}{2}c_i^2\sigma^2 + (e_2 + b_i + c_i\xi)'\mu\\ &+ (e_2 + b_i + c_i\xi)'\Phi X_t + \frac{1}{2}(e_2 + b_i + c_i\xi)'\Sigma(e_2 + b_i + c_i\xi)\Big).\end{aligned} \tag{D3}$$

Collecting terms, we can write:

$$M_{t+i+1} = \exp(a_{i+1} + b'_{i+1}X_t + c_{i+1}\delta_t),$$

where a_{i+1}, b_{i+1}, and c_{i+1} take the form in Equation (10). The sum of exponential affine functions of the price-dividend ratio means that this model falls under the class of affine equity pricing models developed by Ang and Liu (2001), Bekaert and Grenadier (2002), and Bakshi and Chen (2005).

Appendix E: Estimating the Present Value Model

The calibration of the present value model proceeds in two steps. First, we estimate the VAR parameters (μ, Φ, Σ) in Equation (7) of $X_t = (r_t g_t^d)'$. This estimation is complicated by the fact that in the data we observe dividends summed up over the past year, but we specify the frequency of our model to be quarterly. Since the VAR specifies the dynamics of short rates and dividend growth, we use only short rate and dividend growth data to estimate the VAR. In the second step, we hold the VAR parameters fixed and estimate the parameters of the discount rate δ_t in Equation (8). In both stages, we use simulated method of moments (Duffie and Singleton 1993). To calibrate the model, we use US data from January 1952 to December 2001.

We observe dividend growth summed up over the past 12 months, $g_t^{d,4}$, but the model requires quarterly dividend growth g_t^d. By simulating g_t^d from the VAR, we can construct $g_t^{d,4}$ using the transformation

$$g_t^{d,4} = \log\left(\frac{D_t + D_{t-1} + D_{t-2} + D_{t-3}}{D_{t-1} + D_{t-2} + D_{t-3} + D_{t-4}}\right) = g_{t-3}^d + \log\left(\frac{1 + e^{g_{t-2}^d} + e^{(g_{t-2}^d + g_{t-1}^d)} + e^{(g_{t-2}^d + g_{t-1}^d + g_t^d)}}{1 + e^{g_{t-3}^d} + e^{(g_{t-3}^d + g_{t-2}^d)} + e^{(g_{t-3}^d + g_{t-2}^d + g_{t-1}^d)}}\right). \quad \text{(E1)}$$

Equation (E1) shows that the relation between quarterly growth rates and growth rates at a quarterly frequency using dividends summed up over the past year is highly nonlinear. In particular, the summing of dividends over the past four quarters induces serial correlation up to three lags, even when g_t^d is serially uncorrelated.

To estimate the VAR on X_t, we impose a restricted companion form Φ where $\Phi_{12} = 0$, so there is no Granger-causality from dividend growth to interest rates. This assumption is motivated by an analysis of an unconstrained VAR on $(r_t g_t^{d,4})'$, where we fail to reject the null that $g_t^{d,4}$ fails to Granger-cause interest rates. Hence, we first estimate an AR(1) on quarterly short rates and then holding the parameters for r_t fixed, we estimate the remaining parameters in μ, Φ, and Σ by using first and second moments of $g_t^{d,4}$, in addition to the moments $E(r_t g_t^{d,4})$, $E(g_{t-4}^{d,4} g_t^{d,4})$, and $E(r_{t-4} g_t^{d,4})$. Hence, the system is exactly identified. The cross-moment lag length is set at four because the first three lags are affected by the autocorrelation induced by the nonlinear filter for annual dividend growth in Equation (E1). We compute the Newey–West (1987) weighting matrix with four lags using the data, so we do not need to iterate on the weighting matrix.

In the second stage, we hold the parameters of the VAR fixed at their estimates in Panel A. The parameters α, β, ϕ, and σ^2 are estimated by matching 12 moment conditions: the first and second moments of excess returns $\tilde{y}_t$ and dividend yields D_t^4/P_t in the data, with r_{t-1} and $g_{t-1}^{d,4}$ as instruments. Hence, we use the moments

$$E[q_t \otimes z_{t-1}] = 0,$$

where $q_t = [\tilde{y}_t \; D_t^4/P_t \; \tilde{y}_t^2 \; (D_t^4/P_t)^2]'$ and $z_{t-1} = (1 \; r_{t-1} \; g_{t-1}^{d,4})'$.

The relation between the closed-form quarterly dividend yields dy_t (Proposition 4.1) and the dividend yields in the data dy_t^4 (which use dividends summed over the last 4 quarters) is complex:

$$\begin{aligned}\frac{D_t^4}{P_t} &= \frac{D_t + D_{t-1} + D_{t-2} + D_{t-3}}{P_t} \\ &= \frac{D_t}{P_t} + \frac{D_{t-1}}{P_{t-1}}\frac{P_{t-1}}{P_t} + \frac{D_{t-2}}{P_{t-2}}\frac{P_{t-2}}{P_t} + \frac{D_{t-3}}{P_{t-3}}\frac{P_{t-3}}{P_t},\end{aligned} \tag{E2}$$

where the capital gain over n periods P_t/P_{t-n} can be evaluated using:

$$\frac{P_t}{P_{t-n}} = \frac{P_t/D_t}{P_{t-n}/D_{t-n}} \exp\left(\sum_{i=0}^{n-1} g_{t-i}^d\right).$$

The predictability regressions use excess log returns $\tilde{y}_{t+1} = \log(Y_{t+1}) - r_t$, where $Y_t = \log((P_{t+1} + D_{t+1})/P_t)$. Since we model $\delta_t = \log(\mathrm{E}_t(P_{t+1} + D_{t+1})/P_t))$, the expected log excess return $\mathrm{E}_t(\log(Y_{t+1}) - r_t) \neq (\log \mathrm{E}_t(Y_{t+1}) - r_t)$ is not closed form in our model but is a function of time t state variables because of the Markov structure of the model. The two quantities $\mathrm{E}_t(\log(Y_{t+1}) - r_t)$ and $(\log \mathrm{E}_t(Y_{t+1}) - r_t)$ are not equal because of the presence of state-dependent heteroskedasticity, which induces (time-varying) Jensen's inequality terms. To compute the conditional expected value of k-period excess returns, we project excess returns onto a fourth-order polynomial in (X_t,δ_t), and the fitted value is the model-implied conditional expected excess return.

References

Amihud, Y., and C. M. Hurvich, 2004, "Predictive Regressions: A Reduced-Bias Estimation Method," *Journal of Financial and Quantitative Analysis*, 39(4), 813–841.

Ang, A., and G. Bekaert, 2001, "Stock Return Predictability: Is It There?" NBER working paper 8207.

Ang, A., and G. Bekaert, 2002, "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics*, 20(2), 163–182.

Ang, A., and J. Liu, 2001, "A General Affine Earnings Valuation Model," *Review of Accounting Studies*, 6, 397–425.

Ang, A., and J. Liu, 2006, "Risk, Return and Dividends," Forthcoming in *Journal of Financial Economics*.

Bakshi, G., and Z. Chen, 2005, "Stock Valuation in Dynamic Economies," *Journal of Financial Markets*, 8(2), 115–151.

Bansal, R., and C. Lundblad, 2002, "Market Efficiency, Fundamental Values and Asset Returns in Global Equity Markets," *Journal of Econometrics*, 109, 195–237.

Bansal, R., and A. Yaron, 2004, "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Bekaert, G., and S. Grenadier, 2002, "Stock and Bond Pricing in an Affine Economy," Working Paper, Columbia Business School, New York.

Bollerslev, T., and R. J. Hodrick, 1996, "Financial Market Efficiency Tests," in M. H. Pesaran and M. R. Wickens (eds), *The Handbook of Applied Econometrics Vol 1, Macroeconomics*, Blackwell, Oxford, United kingdom, 415–458.

Bossaerts, P., and P. Hillion, 1999, "Implementing Statistical Criteria to Select Return Forecasting Models: What do we Learn?," *Review of Financial Studies*, 12, 405–428.

Boudoukh, J., and M. Richardson, 1993, "The Statistics of Long-Horizon Regressions," *Mathematical Finance*, 4(2), 103–120.

Breen, W., L. R. Glosten, and R. Jagannathan, 1989, "Economic Significance of Predictable Variations in Stock Index Returns," *Journal of Finance*, 44(5), 1177–1189.

Campbell, J. Y., 1987, "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373–399.

Campbell, J. Y., 1991, "A Variance Decomposition for Stock Returns," *Economic Journal*, 101, 157–179.

Campbell, J. Y., 2003, "Consumption-Based Asset Pricing," in G. Constantinides, M. Harris, and R. Stulz (eds), *Handbook of the Economics of Finance*, North-Holland, Amsterdam, 803–887.

Campbell, J. Y., and R. J. Shiller, 1988a, "Stock Prices, Earnings and Expected Dividends," *Journal of Finance*, 43(3), 661–676.

Campbell, J. Y., and R. J. Shiller, 1988b, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1(3), 195–228.

Campbell, J. Y., and L. M. Viceira, 1999, "Consumption and Portfolio Decisions When Expected Returns are Time Varying," *Quarterly Journal of Economics*, 114(2), 433–492.

Campbell, J. Y., and M. Yogo, 2006, "Efficient Tests of Stock Return Predictability," forthcoming in *Journal of Financial Economics*, 81(1), 27–60.

Clarida, R., J. Galí, and M. Gertler, 1999, "The Science Of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37(4), 1661–1707.

Cochrane, J. H., 1992, "Explaining the Variance of Price-Dividend Ratios," *Review of Financial Studies*, 5, 243–280.

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

Duffie, D., and K. J. Singleton, 1993, "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61(4), 929–952.

Elliot, G., and J. H. Stock, 1994, "Inference in Time Series Regression when the Order of Integration of a Regressor is Unknown," *Econometric Theory*, 10, 672–700.

Engstrom, E., 2003, "The Conditional Relationship between Stock Market Returns and the Dividend Price Ratio," Working Paper, Columbia University, New York.

Fama, E., and F. French, 1988, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–26.

Fama, E., and G. W. Schwert, 1977, "Asset Returns and Inflation," *Journal of Financial Economics*, 5, 115–146.

Ferson, W. E., S. Sarkissian, and T. T. Simin, 2003, "Spurious Regressions in Financial Economics?" *Journal of Finance*, 58(4), 1393–1414.

Goetzmann, W. N., and P. Jorion, 1993, "Testing the Predictive Power of Dividend Yields," *Journal of Finance*, 48(2), 663–679.

Goetzmann, W. N., and P. Jorion, 1995, "A Longer Look at Dividend Yields," *Journal of Business*, 68, 483–508.

Goyal, A., and I. Welch, 2003, "The Myth of Predictability: Does the Dividend Yield Forecast the Equity Premium?" *Management Science*, 49(5), 639–654.

Goyal, A., and I. Welch, 2004, "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," NBER Working Paper 10483.

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–54.

Hansen, L., and R. Hodrick, 1980, "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88, 829–853.

Hodrick, R. J., 1992, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies*, 5(3), 357–386.

Jansson, M., and M. J. Moreira, 2006, "Optimal Inference in Regression Models with Nearly Integrated Regressors," forthcoming in *Econometrica*.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin, 1992, "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure are we that Economic Time Series have a Unit Root?" *Journal of Econometrics*, 54(1–3), 159–178.

Lamont, O., 1998, "Earnings and Expected Returns," *Journal of Finance*, 53(5), 1563–1587.

Lee, B. S., 1992, "Causal Relations Among Stock Returns, Interest Rates, Real Activity and Inflation," *Journal of Finance*, 48, 1591–1603.

Lettau, M., and S. Ludvigson, 2001, "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance*, 56(3), 815–849.

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, 209–235.

LO, A W., and A. C., Mackinlay "Data-Snooping Biases in Tests of Financial Asset Pricing Models" *Review of Financial Studies*, 3(3), 431–467.

Menzly, L., J. Santos, and P. Veronesi, 2004, "Understanding Predictability," *Journal of Political Economy*, 112(1), 1–47.

Newey, W., 1985, "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29(3), 229–256.

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

Patelis, A. D., 1997, "Stock Return Predictability and the Role of Monetary Policy," *Journal of Finance*, 52(5), 1951–1972.

Phillips, P. C. B., and P. Perron, 1988, "Testing for a Unit Root in Time Series Regression," *Biometrica*, 75, 335–346.

Polk, C., S. Thompson, and T. Vuolteenaho, 2006, "New Forecasts of the Equity Premium," *Journal of Financial Economics*, 81(1), 101–141.

Richardson, M., 1993, "Temporary Components of Stock Prices: A Skeptic's View," *Journal of Business and Economic Statistics*, 11(2), 199–207.

Richardson, M., and T. Smith, 1991, "Tests of Financial Models in the Presence of Overlapping Observations," *Review of Financial Studies*, 4(2), 227–254.

Richardson, M., and J. H. Stock, 1989, "Drawing Inferences from Statistics Based on Multiyear Asset Returns," *Journal of Financial Economics*, 25, 323–348.

Shiller, R. J., and A. E. Beltratti, 1992, "Stock Prices and Bond Yields: Can Their Comovements be Explained in Terms of Present Value Models?" *Journal of Monetary Economics*, 30, 25–46.

Stambaugh, R. F., 1999, "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421.

Timmermann, A, and B. Paye, 2006, "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13, 274–315.

Torous, W., R. Valkanov, and S. Yan, 2004, "On Predicting Stock Returns with Nearly Integrated Explanatory Variables," *Journal of Business*, 77(4), 937–966.

Valkanov, R., 2003, "Long-Horizon Regressions: Theoretical Results and Applications," *Journal of Financial Economics*, 68(2), 201–232.

Wolf, M., 2000, "Stock Returns And Dividend Yields Revisited: A New Way To Look At An Old Problem," *Journal of Business and Economic Statistics*, 18(1), 18–30.